CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 3 September 2009

JASA EXPRESS LETTERS

LETTERS TO THE EDITOR

NONLINEAR ACOUSTICS [25]

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Improvement of acoustic theory of ultrasonic waves in dilute bubbly liquids

Keita Ando, Tim Colonius, and Christopher E. Brennen
Division of Engineering and Applied Science, California Institute of Technology, Pasadena, California 91125
kando@caltech.edu, colonius@caltech.edu, brennen@caltech.edu

Abstract: The theory of the acoustics of dilute bubbly liquids is reviewed, and the dispersion relation is modified by including the effect of liquid compressibility on the natural frequency of the bubbles. The modified theory is shown to more accurately predict the trend in measured attenuation of ultrasonic waves. The model limitations associated with such high-frequency waves are discussed.

PACS numbers: 43.35.Bf, 43.30.Ft, 43.20.Hq [GD]
Date Received: May 28, 2009 **Date Accepted:** June 18, 2009

1. Introduction

The acoustics of bubbly liquids have been extensively studied for many years. The traditional theory for a dilute bubbly mixture assumes that mutual interactions among the bubbles are negligible. The bubble/bubble interactions can never be ignored at resonance even in the dilute limit,[1,2] and the theory is known to overestimate attenuation under the resonant condition. The theory is also known to be inaccurate in estimating the attenuation of ultrasonic waves. So far, to the authors' knowledge, the cause of the discrepancy under the ultrasonic condition has not been revealed.

Herein, we briefly review the theory and modify the dispersion relation by including the effect of liquid compressibility on the natural frequency of the bubbles and validate this modified theory by comparing to experimental data. Finally, we discuss the model limitations associated with ultrasonic waves.

2. Review of the theory

In the classic papers of Foldy[3] and Carstensen and Foldy,[4] wave propagation in a bubbly mixture was treated as a problem of multiple scattering by randomly distributed isotropic scatterers representing the spherical bubbles, and the dispersion relation for the mixture was derived. An alternative approach is to treat the mixture as a single phase (continuum) medium. van Wijngaarden[5] defined volume-averaged quantities in order to remove the local fluctuations due to scattering and derived the averaged equations based on heuristic, physical reasoning. By linearizing van Wijngaarden's equations, Commander and Prosperetti[2] derived the dispersion relation

$$\frac{1}{c_m^2} = \frac{1}{c_l^2} + 4\pi n \int_0^\infty \frac{R_0 f(R_0) dR_0}{\omega_N^2 - \omega^2 + i2\delta\omega}, \tag{1}$$

where c_m is the complex sonic speed in the mixture, c_l is the sonic speed in the liquid alone, n is the total bubble number density, δ is the bubble-dynamic damping constant, ω is the temporal angular frequency ($\omega = 2\pi f$), ω_N is the natural frequency of the bubbles, R_0 is the equilibrium bubble radius, and $f(R_0)$ is the density function for the size distribution of the equilibrium bubble radius, which satisfies $\int_0^\infty f(R_0) dR_0 = 1$. In the derivation of the dispersion relation (1), the void fraction,

$$\alpha = \frac{4\pi}{3} n \int_0^\infty R_0^3 f(R_0) dR_0, \tag{2}$$

is assumed to be small ($\alpha \ll 1$) and relative motion between the phases is ignored. The relative motion has been shown to have minimal impact on the acoustic problem.[6]

To complete the dispersion relation (1), the bubble-dynamic constant and the natural frequency need to be specified. Commander and Prosperetti[2] used the following expressions for δ and ω_N,

$$\delta = \frac{2\mu_l}{\rho_l R_0^2} + \frac{p_{g0}}{2\rho_l \omega R_0^2} \Im\{\Upsilon\} + \frac{\omega^2 R_0}{2c_l}, \tag{3}$$

$$\omega_N^2 = \frac{p_{g0}}{\rho_l R_0^2} \left(\Re\{\Upsilon\} - \frac{2S}{p_{g0} R_0} \right). \tag{4}$$

Here, μ_l is the liquid viscosity, S is the surface tension, and p_{g0} is the internal (gas) bubble pressure (vapor pressure is typically negligible) given by $p_{g0} = p_{l0} + (2S/R_0)$, where p_{l0} is the ambient pressure. The quantity Υ is a function of the Peclet number $\mathrm{Pe} = \omega R_0^2 / \alpha_{\mathrm{th}}$, where α_{th} is the thermal diffusivity of the gas,

$$\Upsilon = \frac{3\gamma}{1 - i3(\gamma - 1)\mathrm{Pe}^{-1}(\sqrt{i\mathrm{Pe}} \coth \sqrt{i\mathrm{Pe}} - 1)}, \tag{5}$$

where γ is the ratio of specific heats. The effective polytropic index for thermal behavior of the gas is then given by $\kappa = \Re\{\Upsilon\}/3$. Since $\kappa \to 1$ as $\omega \to 0$ or $\mathrm{Pe} \to 0$, the isothermal natural frequency (defined below) is obtained in the quasistatic limit and is generally very close to the resonant frequency.

$$\omega_N^2|_{\kappa=1} = \frac{p_{g0}}{\rho_l R_0^2} \left(3 - \frac{2S}{p_{g0} R_0} \right). \tag{6}$$

It should be noted that the effect of liquid compressibility is ignored in Eq. (4) and is negligible unless the frequency is extremely high compared to the resonant frequency.[7]

We define the phase speed V and attenuation A (in decibels per unit length) as

$$V = \left[\Re\left\{ \frac{1}{c_m} \right\} \right]^{-1}, \tag{7}$$

$$A = -20(\log_{10} e)\omega \Im\left\{ \frac{1}{c_m} \right\}. \tag{8}$$

The estimated phase velocity (7) is known to yield quantitative agreement with experimental data in a wide frequency range below and above the resonance.[2,8] However, the estimated attenuation (8) under resonant and ultrasonic conditions appears to deviate from the experimental values.

Before concluding this review, we examine the model limitations. In order that the mixture be considered homogeneous and the wave structure be well resolved, we need to choose a physically appropriate averaging volume, $\Delta\mathcal{V}$, and presuppose the scale separation[9]

$$l = n^{-1/3} \ll \Delta\mathcal{V}^{1/3} \ll L, \tag{9}$$

where l is the mean bubble spacing and L is the wavelength in the mixture. Note that $R_0 \ll l$ in the dilute limit (i.e., $\alpha \to 0$). Since the wavelength of ultrasonic waves may be comparable to or shorter than the mean bubble spacing, the continuum model may be invalid. In addition, neglect of the acoustic contribution to the bubble natural frequency (4) may also give rise to a discrep-

ancy in theory and experiment between the attenuation of high-frequency waves. In Sec. 3, we discuss the effect of liquid compressibility on the attenuation of ultrasonic waves.

3. Modification to the theory

3.1 Linearized dynamics of the spherical bubbles

To obtain the formulas for the bubble-dynamic damping constant and the bubble natural frequency, we need to linearize the spherical bubble dynamics. It follows from Prosperetti[7] and Prosperetti *et al.*[10] that the linearized dynamics are described by

$$\ddot{x} + 2\,\delta\dot{x} + \omega_N^2 x = -\,\epsilon\frac{p_{l0}}{\rho_l R_0^2}e^{i\omega t}, \tag{10}$$

where ϵ is the infinitesimal amplitude of sinusoidally oscillating (farfield) liquid pressure ($|\epsilon| \ll 1$) and x is the corresponding perturbation in the bubble radius ($|x| \ll 1$):

$$p_l = p_{l0}(1 + \epsilon e^{i\omega t}), \tag{11}$$

$$R = R_0(1 + x). \tag{12}$$

Here, the damping constant and the natural frequency are

$$\delta = \frac{2\mu_l}{\rho_l R_0^2} + \frac{p_{g0}}{2\rho_l \omega R_0^2}\Im\{\Upsilon\} + \frac{\dfrac{\omega^2 R_0}{2c_l}}{1 + \left(\dfrac{\omega R_0}{c_l}\right)^2}, \tag{13}$$

$$\omega_N^2 = \frac{p_{g0}}{\rho_l R_0^2}\left(\Re\{\Upsilon\} - \frac{2S}{p_{g0}R_0}\right) + \frac{\left(\dfrac{\omega R_0}{c_l}\right)^2}{1 + \left(\dfrac{\omega R_0}{c_l}\right)^2}\omega^2, \tag{14}$$

where the last terms on the right-hand sides of the above equations represent the contributions associated with liquid compressibility. To quantify the impact of liquid compressibility on the ultrasonic waves, we compute the dispersion relation (1) based on Eqs. (13) and (14) instead of Eqs. (3) and (4) and validate the modification by comparing to experimental data below.

For future use, we develop the asymptotic limits of Eqs. (13) and (14). In the quasistatic limit (i.e., $\omega \to 0$), it follows from Prosperetti[7,11] that

$$\delta \approx \frac{2\mu_l}{\rho_l R_0^2} + \frac{\gamma - 1}{10\gamma}\frac{p_{g0}}{\rho_l \alpha_{\mathrm{th}}}, \tag{15}$$

$$\omega_N \approx \omega_N|_{\kappa=1}, \tag{16}$$

where $\omega_N|_{\kappa=1}$ is the isothermal natural frequency (6) so that liquid compressibility is unimportant. On the other hand, in the limit of $\omega \gg \omega_N|_{\kappa=1}$, it is readily shown that

$$\delta \approx \frac{c_l}{2R_0}, \tag{17}$$

$$\omega_N \approx \omega. \tag{18}$$

In this limit, the damping due to liquid compressibility dominates over the viscous and thermal contributions and the natural frequency is independent of the bubble size.

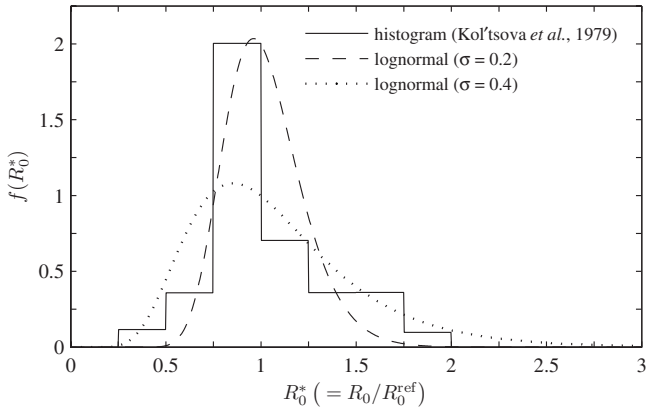


Fig. 1. Normalized histogram of the bubble size distribution of Kol'tsova *et al.* (Ref. 12) and lognormal distributions with a standard deviation σ. The probable size, R_0^{ref}, is set to be 20 μm. The measured values are based on a hydrogen/water mixture of α=0.03% at 15 °C and 1 atm.

3.2 Validation of the modification

We compare the dispersion relation (1) to the experiment of Kol'tsova *et al.*[12] who measured the attenuation in a high-frequency range up to 30 MHz. In those experiments, hydrogen bubbles were produced using electrolysis and had a size distribution with a mean radius of 15–20 μm. The histogram of the size distribution for α=0.03% (probable size, R_0^{ref}=20 μm) is plotted in Fig. 1. We assume that the size distribution for different values of α is similar to that in Fig. 1. The actual distribution may be smooth, as shown in Fig. 1, and we model it using a lognormal density function with standard deviation σ,

$$f(R_0^*) = \frac{1}{\sqrt{2\pi}\sigma R_0^*} \exp\left(-\frac{\ln^2 R_0^*}{2\sigma^2}\right), \tag{19}$$

where $R_0^*=R_0/R_0^{\mathrm{ref}}$.

Using the size distributions in Fig. 1, the phase velocity (7) and attenuation (8) are computed using Eqs. (3) and (4) or Eqs. (13) and (14) and are plotted in Fig. 2. The attenuation of Kol'tsova *et al.*[12] is also plotted (α=0.004%, $\omega_N|_{\kappa=1}/2\pi$=0.142 MHz for R_0^{ref}). It follows

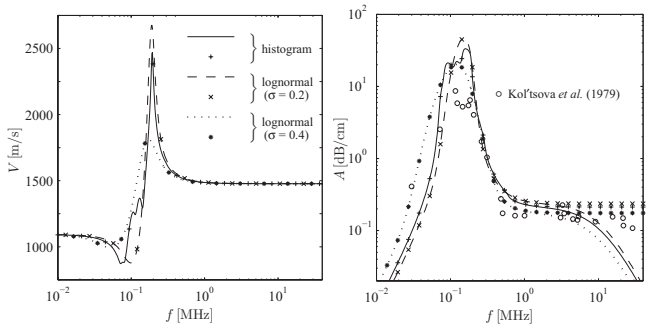


Fig. 2. Phase velocity (left) and attenuation (right) for a hydrogen/water mixture of α=0.004% and R_0^{ref}=20 μm at 15 °C and 1 atm. The lines and symbols (plus, cross, and asterisk) are computed using the dispersion relation (1) with Eqs. (3) and (4) and with Eqs. (13) and (14), respectively. The symbols (circle) denote the experimental data of Kol'tsova *et al.* (Ref. 12).

from the phase velocity plot that the present modifications to δ and ω_N have negligible impact on V. It is also found that the size distribution tends to smooth the transition in V at the resonant frequency.

However, the present modification does lead to a striking change in the attenuation for $\omega \gg \omega_N|_{\kappa=1}$. The dispersion relation (1) with the present modification predicts attenuations at high frequencies that agree with the experimental measurements. That is, liquid compressibility has major impact on the attenuation of the ultrasonic waves. As a result of Eqs. (17) and (18), the phase velocity and the attenuation for $\omega \gg \omega_N|_{\kappa=1}$ asymptote to the constant values,

$$V \approx c_l, \tag{20}$$

$$A \approx 20(\log_{10} e)\frac{3\alpha}{2R_0^{\mathrm{ref}}}C_1, \tag{21}$$

where the constant C_1 is

$$C_1 = \frac{\int_0^\infty R_0^{*2} f(R_0^*) dR_0^*}{\int_0^\infty R_0^{*3} f(R_0^*) dR_0^*}. \tag{22}$$

For the lognormal $f(R_0^*)$, $C_1 = \exp(-2.5\sigma^2)$ so that the attenuation decreases as σ increases. It should be pointed out that Kol'tsova *et al.*[12] presented the different data sets of the attenuation (with different void fractions) which remains almost constant under the ultrasonic condition. It is therefore concluded that the modified theory is superior to the previous theory when it comes to predicting this trend.

Furthermore, we notice that the size distribution increases the attenuation below the resonant frequency. From Eqs. (15) and (16), the asymptotic values at low frequency become

$$V \approx \frac{c_l}{\sqrt{1 + \dfrac{\alpha \rho_l c_l^2}{p_{l0}}}}, \tag{23}$$

$$A \approx 20(\log_{10} e)\frac{\alpha \rho_l^2 V \delta \omega^2}{3p_{l0}^2}C_2, \tag{24}$$

where the constant C_2 is

$$C_2 = \frac{\int_0^\infty R_0^{*5} f(R_0^*) dR_0^*}{\int_0^\infty R_0^{*3} f(R_0^*) dR_0^*}. \tag{25}$$

Here, we have neglected the viscous contribution in Eq. (15) since the thermal damping generally dominates over the viscous damping. In addition, it is assumed that the surface tension is negligible in Eq. (16). For the lognormal $f(R_0^*)$, $C_2 = \exp(8\sigma^2)$ so that the attenuation increases as σ increases. To interpret this tendency, consider linear bubble oscillations under a sinusoidal forcing ($p_l - p_{l0} \propto \sin(\omega t)$) of the farfield liquid pressure. The corresponding perturbation in the bubble radius oscillates with the forcing frequency ω and with a phase shift ϕ such that

$$\cos\phi = \frac{\omega_N^2(R_0) - \omega^2}{\sqrt{(\omega_N^2(R_0) - \omega^2)^2 + 4\delta^2(R_0)\omega^2}}. \tag{26}$$

Therefore, phase cancellations due to the different phases among the different-sized bubbles occur in the low-frequency regime since $\omega_N \approx \omega_N|_{\kappa=1} \neq \omega$. The phase cancellations can be regarded as apparent damping of the wave propagation in the polydisperse mixture, and the damping mechanism becomes more effective as the bubble size distribution broadens.[13,14] As a result,

the size distribution increases the attenuation, as seen in Fig. 2. However, in the ultrasonic limit, all the different-sized bubbles oscillate with the same phase due to the fact that $\omega_N \approx \omega$ (regardless of the bubble sizes); hence, in this case, the phase cancellations do not occur.

Finally, we check the model limitation (9). The mean bubble spacing in Fig. 2 is

$$l = \sqrt[3]{\frac{4\pi \int_0^\infty R_0^{*3} f(R_0^*) dR_0^*}{3\alpha}} R_0^{\mathrm{ref}} \approx 1100\ \mu\mathrm{m}, \tag{27}$$

where the histogram in Fig. 1 is used for $f(R_0^*)$. At $f=10$ MHz, the wavelength is approximated by

$$L \approx \frac{c_l}{f} \approx 150\ \mu\mathrm{m}, \tag{28}$$

which is larger than the mean bubble radius but shorter than the mean bubble spacing. Hence, the continuum assumption is invalidated, while bubble/bubble interactions may be ignored. However, despite this violation, as seen in Fig. 2, the continuum theory accurately predicts the trend in the attenuation around $f=10$ MHz. This implies that the validity of the dispersion relation (1) may extend beyond the limitation (9).

4. Conclusion

A modification to the traditional dispersion relation of linear waves in dilute bubbly liquids is made to take into account the effect of liquid compressibility (which is very important far above the resonant frequency) on linearized dynamics of spherical bubbles. The modified dispersion relation is found to accurately predict the trend in measured attenuation of ultrasonic waves. The agreement between the modified theory and experiment implies that the validity of the dispersion relation (1) may extend beyond the continuum model limitation (9).

Acknowledgments

The authors gratefully acknowledge the support by ONR Grant No. N00014-06-1-0730 and by NIH Grant No. PO1 DK043881.

References and links

[1]P. C. Waterman and R. Truell, "Multiple scattering of waves," J. Math. Phys. **2**, 512–537 (1961).
[2]K. W. Commander and A. Prosperetti, "Linear pressure waves in bubbly liquids: Comparison between theory and experiments," J. Acoust. Soc. Am. **85**, 732–746 (1989).
[3]L. L. Foldy, "The multiple scattering of waves," Phys. Rev. **67**, 107–119 (1945).
[4]E. L. Carstensen and L. L. Foldy, "Propagation of sound through a liquid containing bubbles," J. Acoust. Soc. Am. **19**, 481–501 (1947).
[5]L. van Wijngaarden, "One-dimensional flow of liquids containing small gas bubbles," Annu. Rev. Fluid Mech. **4**, 369–396 (1972).
[6]L. d'Agostino, C. E. Brennen, and A. J. Acosta, "Linearized dynamics of two-dimensional bubbly and cavitating flows over slender surfaces," J. Fluid Mech. **199**, 485–509 (1988).
[7]A. Prosperetti, "Thermal effects and damping mechanisms in the forced radial oscillations of gas bubbles in liquids," J. Acoust. Soc. Am. **61**, 17–27 (1977).
[8]S. A. Cheyne, C. T. Stebbings, and R. A. Roy, "Phase velocity measurements in bubbly liquids using a fiber optic laser interferometer," J. Acoust. Soc. Am. **97**, 1621–1624 (1995).
[9]A. Prosperetti, "Fundamental acoustic properties of bubbly liquids," in *Handbook of Elastic Properties of Solid, Liquids, and Gases*, edited by M. Levy, H. Bass, and R. Stern (Academic, New York, 2001), Vol. **4**, pp. 183–205.
[10]A. Prosperetti, L. A. Crum, and K. W. Commander, "Nonlinear bubble dynamics," J. Acoust. Soc. Am. **83**, 502–514 (1988).
[11]A. Prosperetti, "The thermal behaviour of oscillating gas bubbles," J. Fluid Mech. **222**, 587–616 (1991).
[12]I. S. Kol'tsova, L. O. Krynskii, I. G. Mikhailov, and I. E. Pokrovskaya, "Attenuation of ultrasonic waves in low-viscosity liquids containing gas bubbles," Sov. Phys. Acoust. **25**, 409–413 (1979).
[13]P. Smereka, "A Vlasov equation for pressure wave propagation in bubbly fluids," J. Fluid Mech. **454**, 287–325 (2002).
[14]T. Colonius, R. Hagmeijer, K. Ando, and C. E. Brennen, "Statistical equilibrium of bubble oscillations in dilute bubbly flows," Phys. Fluids **20**, 040902 (2008).

Robust acoustic particle manipulation: A thin-reflector design for moving particles to a surface

P. Glynne-Jones, R. J. Boltryk, and M. Hill
School of Engineering Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom
p.glynne-jones@soton.ac.uk; r.j.boltryk@soton.ac.uk; m.hill@soton.ac.uk

N. R. Harris
Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom
nrh@ecs.soton.ac.uk

P. Baclet
ENSIAME, Université de Valenciennes et du Hainaut Cambrésis, F-59313 Valenciennes, France
bacletp@orange.fr

Abstract: Existing ultrasonic manipulation devices capable of pushing particles to a surface ("quarter-wave" devices) have significant potential in sensor applications. A configuration for achieving this that uses the first thickness resonance of a layered structure with both a thin reflector layer and thin-fluid layer is described here. Crucially, this mode is efficient with lossy reflector materials such as polymers, produces a more uniform acoustic radiation force at the reflector, and is less sensitive to geometric variations than previously described quarter-wave devices. This design is thus expected to be suitable for mass produced, disposable devices.

PACS numbers: 43.25.Qp, 43.20.Ks [AN]
Date Received: May 1, 2009 **Date Accepted:** July 1, 2009

1. Introduction

Previous literature has described two major classes of planar acoustic particle manipulation devices.

(a) Those where the dominant resonance is in the fluid layer, leading to agglomeration at one or more pressure nodes within the fluid layer (Hawkes and Coakley, 2001).
(b) Those where a resonant reflector layer provides a pressure release boundary condition, causing the agglomeration position to occur at a pressure node close to the fluid/reflector interface (Hill, 2003; Hawkes *et al.*, 2004). These devices typically have fluid-layer thickness close to $\lambda/4$ and reflector thicknesses $n\lambda/2$, and are often referred to as "quarter-wave devices." Quarter-wave devices with no reflector are postulated by Hawkes *et al.* (2002). In their notation, the (non-quarter-wave) device described here is close to "000," i.e., the carrier, fluid, and reflector layers are all vanishingly thin.

Quarter-wave devices have potential application in sensor technology when combined with surface immunoassays, as the acoustic radiation forces can be used to drive particles of interest, such as bacterial spores (Hawkes *et al.*, 2004; Martin *et al.*, 2005), to an antibody coated sensor surface. Quarter-wave designs have several limitations when used in these applications:

In a conventional design, the radiation force at the fluid/reflector boundary tends to zero. This constraint can be avoided by modifying layer thicknesses (Glynne-Jones *et al.*, 2009); however, it has been observed both in models, and experimentally, that the proximity of the pressure node to the interface can lead to a significantly non-uniform force at the interface. This

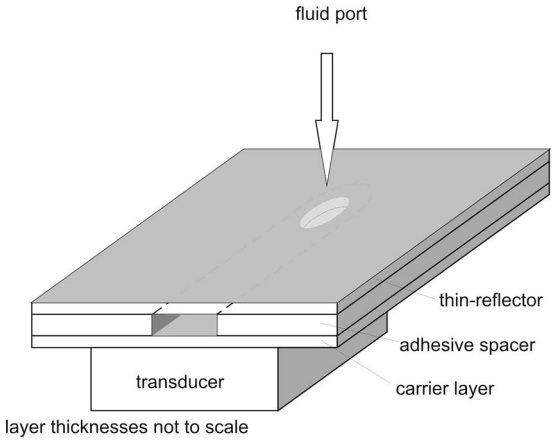


Fig. 1. Thin-reflector device design, showing a cross section across the fluid channel. Particles are driven toward the boundary between the reflector layer and fluid layer.

occurs because lateral variations in the acoustic resonance can distort the expected position of the pressure node sufficiently to cause the node position to be in the reflector in some places and in the fluid in others; thus a particle at the reflector may experience forces either toward or away from the reflector in a single device. Also, this trade-off can cause a pressure anti-node to occur in the channel near the carrier layer, which means that particles near the carrier layer will be driven away from the reflector. In applications where it is required to drive a particle positively onto the reflector surface this limits the effectiveness of such a device.

The resonant nature of the reflector layer in these designs means that a material with a high Q-factor, such as glass, is usually employed [though designs with polymer layers have been previously reported (Gonzalez *et al.*, 2008)]. In cell-based applications this can lead to significant adhesion if the surface is left untreated. The devices are also sensitive to variations in the layer thicknesses (Townsend *et al.*, 2008).

The authors describe here a new thin-reflector (and thin-fluid) arrangement, shown in Fig. 1, which overcomes the above constraints. It operates at the first thickness resonance of a composite structure consisting of the following layers: transducer (typically lead zirconate titanate, PZT), an optional carrier layer that serves to isolate the transducer from the fluid layer, a fluid layer, and a reflector layer. This leads to pressure nodes at only the air boundaries of the device. By making both the reflector layer and fluid layers much less than a wavelength in thickness, it is found that particles in all parts of the fluid channel are attracted toward the fluid/reflector layer boundary. This configuration has not, to the authors' knowledge, been previously described in a planar manipulation device.

This arrangement also has the advantage that the reflector layer can be thin enough to be compatible with high numerical aperture microscope objectives: important in applications such as bio-sensing and cell handling.

In the modeling below the acoustic radiation force on a particle at points within the fluid is calculated from the following equations derived by Gor'kov (Gor'kov, 1962).

The acoustic radiation force (a time averaged quantity) is given by

$$\langle F(r)\rangle = -\nabla\langle\phi(r)\rangle, \tag{1}$$

where the force potential $\langle\phi(r)\rangle$ is given by

$$\langle\phi(r)\rangle = -V\left[\frac{3(\lambda-1)}{2\lambda+1}\langle\bar{E}_{\mathrm{kin}}(r)\rangle - \left(1-\frac{1}{\sigma^2\lambda}\right)\langle\bar{E}_{\mathrm{pot}}(r)\rangle\right]. \tag{2}$$

Table 1. Thin-reflector device dimensions and material properties used in finite element modeling.

Layer	Thickness (μm)	Thickness as a fraction of acoustic wavelength at $f=1.20$ MHz	Material	Model parameters: Density (kg/m^3)	Speed of sound (m/s^2)	Young's modulus (GPa)	Poisson's ratio	Q -factor
Transducer	1000	0.27	PZT-26 (Ferroperm)	Properties as per manufacturer datasheet				100
Carrier layer	104	0.06	Cellulose acetate	1435	2000	4.38	0.29	30
Fluid	120	0.10	Water	1000	1480	···	···	30
Reflector	104	0.06	Cellulose acetate	As above				

Here λ is the ratio of particle density to fluid density, σ is the ratio of speed of sound in the particle to that in the fluid, V is the particle volume, and $\langle \bar{E}_{\text{kin}}(r) \rangle$ and $\langle \bar{E}_{\text{pot}}(r) \rangle$ are the time averaged kinetic and potential energy densities of the sound wave in the fluid.

2. Design, modeling, and results

Whereas the final reflector layer in conventional manipulation devices must have a well controlled thickness to create the necessary boundary conditions for the fluid layer, the final layer in this design can be vanishingly small, and as such the overall acoustic response is not as sensitive to variations in its thickness (a future paper will discuss this in more detail). For clarity, although its function is no longer primarily as a reflector, this layer will continue to be described as "the reflector layer" in this letter.

Table 1 lists the key layer thicknesses, along with the layer thickness as a proportion of the acoustic wavelength in that material at resonance—it is interesting to note how small this proportion is for the fluid and reflector layers. The channel outline was formed by cutting a slot of width 3 mm and length 18 mm into a strip of adhesive transfer tape (3M, 926ATG). This was sandwiched between two squares of cellulose acetate, with fluidic ports at the end of the channel formed by holes in the cellulose acetate. The transducer was a 1 mm thick rectangle (8×6 mm^2) of PZT26 (Ferroperm piezoceramics), and was coupled to the channel with a water soluble ultrasonic coupling gel (Tensive, Parker Laboratories).

The modeled acoustic pressure is distributed across the device as shown in Fig. 2(a). The graph is for an excitation of 10 V_{pp} at the model's resonance frequency of 1.20 MHz. The pressure distribution is calculated using the ANSYS finite element package: The model is essentially one-dimensional, and consists of a strip of elements with boundary conditions to mimic an infinite plain. Table 1 includes the materials data used in the model. Elements capable of modeling both the piezoelectric response to an applied harmonic voltage, and the acoustic/structural interactions are used. The model includes a 10 μm thick gel layer (modeled with the properties of water) between the transducer and carrier layer.

The modeled acoustic radiation force profile on a 10 μm diameter polystyrene bead is shown in Fig. 2(b). It can be seen that at all positions across the channel, there is a force toward the reflector. The magnitude of this force is subject to a large modeling error, as it depends on the damping in the device (results here are for Q-factors of 100 for the transducer and 30 for subsequent layers), which have been only roughly estimated, and must include corrections to allow for non-parallelisms in the device (Gröschl, 1998).

The electrical input impedance of the assembled device is shown in Fig. 3, where it is compared to the modeled impedance. It can be seen that the actual device exhibits a number of

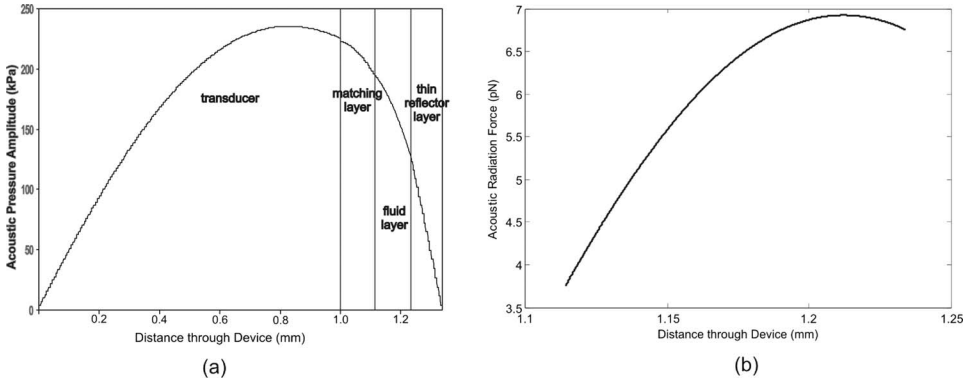


Fig. 2. (a) Acoustic pressure amplitude through the thickness of the device modeled using finite element analysis and (b) acoustic radiation force in the fluid layer on a 10 μm bead polystyrene bead with a drive voltage of 10 V_{pp}. The positive force indicates a force toward the reflector.

resonances not modeled by the one-dimensional model; it is anticipated that these correspond to lateral acoustic resonances in the transducer. The mode of interest in the device has an impedance minimum at 1.179 MHz. Given the estimated nature of the speed of sound in the reflector and carrier layers (2000 m s^{-1}) and thickness of the gel layer, this is reasonably close to the modeled minimum at 1.20 MHz.

To test the device, a dilute solution (7.5×10^5 beads/ml) of polystyrene microspheres of diameter 10 μm (Polysciences Inc., Fluoresbrite microspheres No. 19096) was flowed through the device driven by a syringe micropump at an average velocity of 1 mm s^{-1}. The transducer was driven directly from a signal generator (TTi TG1304) with a sine-wave of frequency 1.179 MHz and amplitude 10 V_{pp}. Under these conditions all the beads passing through the device came into contact with the reflector, many of them sticking to it (this was verified visually with a microscope; all free beads could be observed within a single focal plane, and seen moving between and colliding with beads adhered to the reflector surface thus confirming their close proximity to the surface).

To assess the efficiency of the device the acoustic radiation force was balanced against the buoyant weight of the bead (Martin *et al.*, 2005). It was found that the region showing the strongest forces (lateral variations in acoustic force were apparent) would balance a bead at a voltage of 1.79 V. Since the force on a bead is proportional to the square of applied voltage, it can be deduced that at 10 V_{pp} drive there is a maximum force of 8.8 pN on the bead. This is close to the value predicted by the model [see Fig. 2(b)].

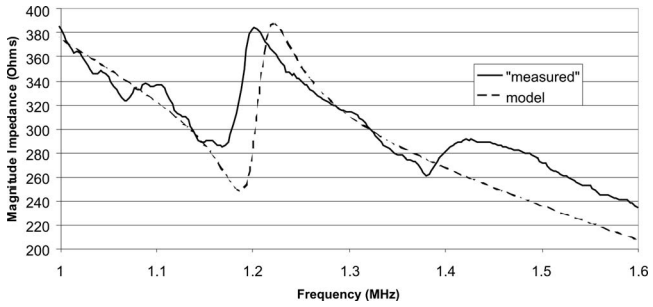


Fig. 3. Device electrical input impedance as measured and modeled. The mode of interest, with a minimum at 1.179 MHz in the model, is seen.

3. Conclusions

The thin-reflector design successfully manipulates particles into contact with the reflector. In contrast to a quarter-wave design the force at the surface is reliably positive, and less sensitive to small variations in reflector- and fluid-layer thicknesses. The device can be constructed from lossy materials such as plastics, potentially useful for disposable devices, and those where the surface chemistry of glass is unsuitable.

Acknowledgment

The authors gratefully acknowledge the support of the EPSRC for the project "Ultrasonic manipulation & transport of DNA molecules in evanescent light fields," Research Grant No. EP/D03454X/1.

References and links

Glynne-Jones, P., Boltryk, R. J., Hill, M., Zhang, F., Dong, L. Q., Wilkinson, J. S., Melvin, T., Harris, N. R., and Brown, T. (**2009**). "Flexible acoustic particle manipulation device with integrated optical waveguide for enhanced microbead assays," Anal. Sci. **25**, 285–291.

Gonzalez, I., Gomez, T., and Fernandez, L. (**2008**). "Experimental study of particle motion within a microchannel narrower than half a wavelength," in *The Sixth USWNet Conference* (ETH, Zurich, Switzerland).

Gor'kov, L. P. (**1962**). "On the forces acting on a small particle in an acoustical field in an ideal fluid," Sov. Phys. Dokl. **6**, 773–775.

Gröschl, M. (**1998**). "Ultrasonic separation of suspended particles—Part I: Fundamentals," Acustica **84**, 432–447.

Hawkes, J. J., Gröschl, M., Benes, E., Nowotny, H., and Coakley, W. T. (**2002**). "Positioning particles within liquids using ultrasound force fields," Revista de Acustica **33**.

Hawkes, J. J., and Coakley, W. T. (**2001**). "Force field particle filter, combining ultrasound standing waves and laminar flow," Sens. Actuators B **75**, 213–222.

Hawkes, J. J., Long, M. J., Coakley, W. T., and McDonnell, M. B. (**2004**). "Ultrasonic deposition of cells on a surface," Biosens. Bioelectron. **19**, 1021–1028.

Hill, M. (**2003**). "The selection of layer thicknesses to control acoustic radiation force profiles in layered resonators," J. Acoust. Soc. Am. **114**, 2654–2661.

Martin, S. P., Townsend, R. J., Kuznetsova, L. A., Borthwick, K. A. J., Hill, M., McDonnell, M. B., and Coakley, W. T. (**2005**). "Spore and micro-particle capture on an immunosensor surface in an ultrasound standing wave system," Biosens. Bioelectron. **21**, 758–767.

Townsend, R. J., Hill, M., Harris, N. R., and McDonnell, M. B. (**2008**). "Performance of a quarter-wavelength particle concentrator," Ultrasonics **48**, 515–520.

Pulse-echo interaction in free-flying horseshoe bats, *Rhinolophus ferrumequinum nippon*

Yu Shiori
Faculty of Engineering, Doshisha University, Kyotanabe 610-0321, Japan
yu.shiori@jp.sony.com

Shizuko Hiryu[a)]
Faculty of Life and Medical Sciences, and Bio-navigation Research Center, Doshisha University, Kyotanabe 610-0321, Japan
shiryu@mail.doshisha.ac.jp

Yu Watanabe
Faculty of Life and Medical Sciences, Doshisha University, Kyotanabe 610-0321, Japan
dmi0126@mail4.doshisha.ac.jp

Hiroshi Riquimaroux and Yoshiaki Watanabe
Faculty of Life and Medical Sciences, and Bio-navigation Research Center, Doshisha University, Kyotanabe 610-0321, Japan
hrikimar@mail.doshisha.ac.jp, kwatanab@mail.doshisha.ac.jp

Abstract: Because horseshoe bats emit a long-duration pulse, the returning echo temporally overlaps with the emitted pulse during echolocation. Here, the pulse-echo interaction that horseshoe bats actually experience during flight was examined using onboard telemetry sound recordings. Doppler-shifted returning echoes produced beats in the amplitude patterns of constant-frequency components. Bats shortened the pulse duration with target distance, but the overlap duration was at least 8 ms within the approach phase. The computations suggest that the phase difference in slowly amplitude-modulated sound (the beat signal) provides a useful cue for target localization.

PACS numbers: 43.80.Ka [CM]
Date Received: March 23, 2009 **Date Accepted:** June 29, 2009

1. Introduction

The greater horseshoe bat (*Rhinolophus ferrumequinum*) and the mustached bat (*Pteronotus parnellii*) emit compound pulses consisting of constant-frequency (CF) and frequency-modulated (FM) components. These bat species, termed CF-FM bats, emit pulses of relatively long duration, ranging from 10 to 50 ms in *R. ferrumequinum* (Tian and Schnitzler, 1997) and from 8 to 32 ms in *P. parnellii* (Henson *et al.*, 1987). CF-FM species change the frequency of the CF component of the emitted pulse to keep the echo CF constant [Doppler-shift compensation (DSC), Schnitzler, 1968]. If a target is close, within a certain target range, temporal overlap between the emitted pulse and the returning echo could occur at the ear. Horseshoe bats only perform DSC when the pulse and the returning echo overlap in time (Schuller, 1974, 1977); therefore, it is believed that the pulse-echo overlap is required for the induction of DSC in *R. ferrumequinum*. In contrast, *Hipposideros terasensis*, another CF-FM species, does not experience pulse-echo overlaps because of its short call duration, but individuals perform DSC (Hiryu *et al.*, 2005).

[a)]Author to whom correspondence should be addressed.

Henson *et al.* (1987) first demonstrated the interaction pattern between the pulse and the echo that actually reaches the bat's ear during flight using a radio transmitter attached to the mustached bat (*P. parnellii*). Once an outgoing pulse overlaps with a returning echo, the pulse-echo interaction produces *beats*, amplitude modulations in the envelope that depend on the frequency difference between the emitted pulse and the Doppler-shifted echo. It was hypothesized that the beats produced by overlaps between pulses and echoes in CF components (pulse-echo overlap) could provide information regarding the frequency difference between the pulse and the Doppler-shifted echo (Suga *et al.*, 1975), or target movement and angle (Grinnell, 1970). However, little information is currently available regarding the beats that bats actually experience during flight.

In this study, the authors recorded emitted pulses and returning echoes in *R. ferrumequinum nippon* during stereotyped landing flights using onboard telemetry sound recordings. The pulse-echo interaction was quantitatively investigated, and they propose a potential model of sound localization derived from the pulse-echo interaction of CF sounds.

2. Materials and methods

Three male *R. ferrumequinum nippon* were used in this study. The specimens were captured from a natural cave in Hyogo Prefecture, Japan, under a license and in compliance with current Japanese laws. The animals were housed in a temperature- and humidity-controlled facility at Doshisha University, Japan, and were allowed free access to food and water. The day and night cycle of the room was controlled at 12 h of darkness and 12 h of light. The experiments complied with the *Principles of Animal Care,* publication No. 86-23, revised in 1985, of the National Institutes of Health and with current Japanese laws.

The experiments were conducted in a radio-wave shielded flight chamber [8 m(L) $\times$3 m(W) $\times$2 m(H)] under long-wavelength lighting with red filters (>650 nm) to prevent bats from using vision. The bats were released at one end of the chamber and allowed to fly freely to the opposite end of the chamber, where a landing net [0.9 m(W) $\times$ 1.1 m(H)] was set on the wall 1.5 m above the floor. Echolocation signals were measured using a lightweight, custom-made on-board wireless ultrasonic microphone, Telemike, placed above the bat's head, and the recording procedure was the same as previously reported (Hiryu *et al.*, 2008). Measured signals were high-pass filtered at 20 kHz (NF, model 3625, Yokohama, Japan) and digitized by a 16-bit, 384-kHz digital audio tape (DAT) recorder (SONY, SIR-1000 W, Tokyo, Japan). Simultaneously, the flight behavior of each bat was recorded with a dual high-speed video camera system at 125 frames/s. The transistor-transistor logic (TTL) signal triggering the video cameras was digitally recorded using the DAT recorder so that flight coordinates could be synchronized with sound data. The three-dimensional coordinates of a flying bat were reconstructed from video images using a commercial motion analysis program (DITECT, DIPP-MOTION 2D ver.2.1) so that the distance between the bat and the target wall (target distance) and the flight speed of the bat were obtained.

Acoustic parameters were analyzed using a custom program in MATLAB on a personal computer. Pulses and echoes were extracted from the displayed spectrogram and a short-time fast Fourier transformation (Hanning window; frequency resolution, 46.9 Hz) was performed on these pulses and echoes. In this study, the authors focused on CF_2 components for analyzing beats because pulse-echo beats in the CF_2 components were distinguished in the Telemike recording data. Theoretically, beats should also occur in the fundamental and other harmonics. However, they are too weak to be measured by Telemike. By comparing the echo delays shown in the sound data to the expected delays computed from the target distance using the formula $2x/c$, where x is the target distance and c is the sound velocity in air [344 m/s; Fig. 1(C)], the echoes from the target wall of the flight chamber could be identified in the spectrogram. Only echoes returning from the target wall were used in the analysis of the CF_2 frequency.

The periods on the beats were analyzed by extracting an enlarged amplitude pattern of each pulse-echo overlap using routines written in MATLAB. Beat frequency was determined from the reciprocal of the modulation period shown in the displayed amplitude pattern [Fig.

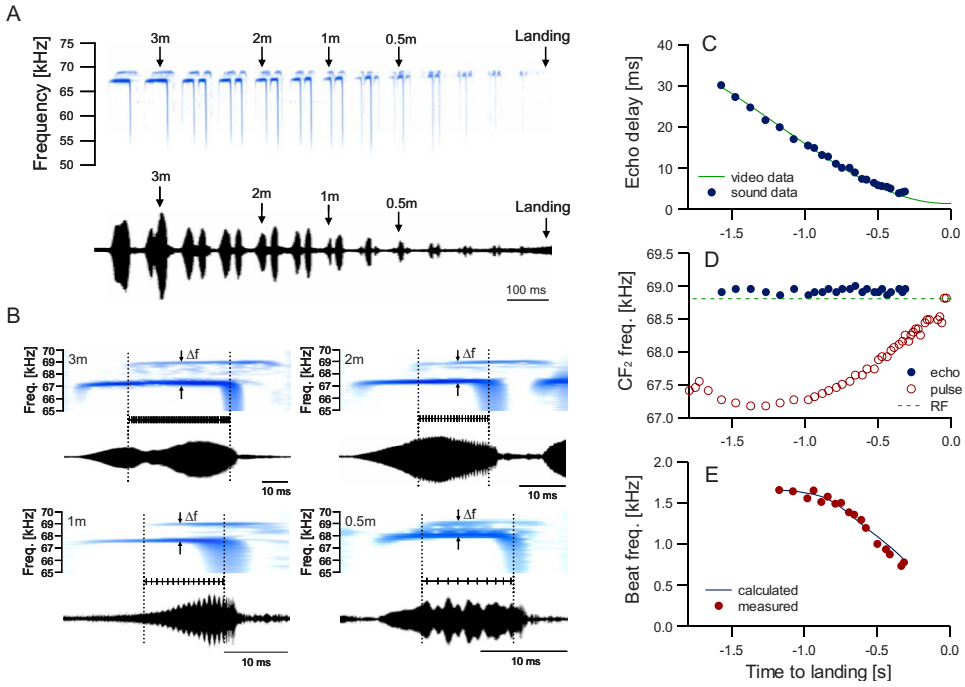


Fig. 1. (Color online) Pulse-echo pairs recorded with a Telemike while *R. ferrumequinum nippon* approached the target wall. (A) Spectrogram and temporal amplitude pattern. The arrows show sounds emitted by the bat at target distances of 3, 2, 1, and 0.5 m. (B) Enlarged spectrograms and temporal amplitude patterns of the sounds marked with arrows in A. The Δf value indicates the frequency difference in the CF_2 component between the pulse and the echo. Pairs of vertical dotted lines show the duration of overlap of CF_2 components between pulse and echo, which was determined by displaying the spectrogram with a high time resolution. Horizontal bars with tick marks illustrate the timing of the peak of the amplitude modulation caused by the beat in the overlapped portion. (C) Changes in the echo delay from the target wall determined from three-dimensional coordinate data of the bat's location (line) and sound data (solid circle). (D) CF_2 frequencies of pulse-echo pairs as a function of time to landing. Open and solid circles indicate the CF_2 frequency of the emitted pulse and the returning echo from the target wall, respectively. The broken line indicates the RF (68.81 kHz). (E) Beat frequency determined from relative flight speed of the bat to the target wall (line) and from the reciprocal of the modulation period in the observed amplitude pattern (solid circle).

1(B)]. To confirm whether the beat frequency of the pulse-echo overlap corresponded with the difference in CF_2 frequencies between an emitted pulse and the Doppler-shifted returning echo, the beat frequency f_{beat} was estimated from the relative flight speed of the bat toward the target wall using the following expression:

$$f_{\text{beat}} = \frac{2v}{c - v} f_{\text{pulse}}, \tag{1}$$

where v is the relative flight speed of the bat toward the target wall and f_{pulse} is the CF_2 frequency of the emitted pulse. The duration of the pulse was determined from the spectrogram with a high time resolution at −25 dB relative to the peak intensity of the pulse. The overlap duration in the CF component of the pulse and echo could be obtained by subtracting the echo delay from the duration of the pulse. Experimental data were taken from a total of 12 flight sessions by three bats.

3. Results and discussion

Figure 1(A) shows a typical spectrogram and a temporal amplitude pattern for echolocation pulses and echoes of *R. ferrumequinum nippon* during a landing flight, recorded by a Telemike mounted on the animal. Doppler-shifted returning echoes overlapped with outgoing pulses and

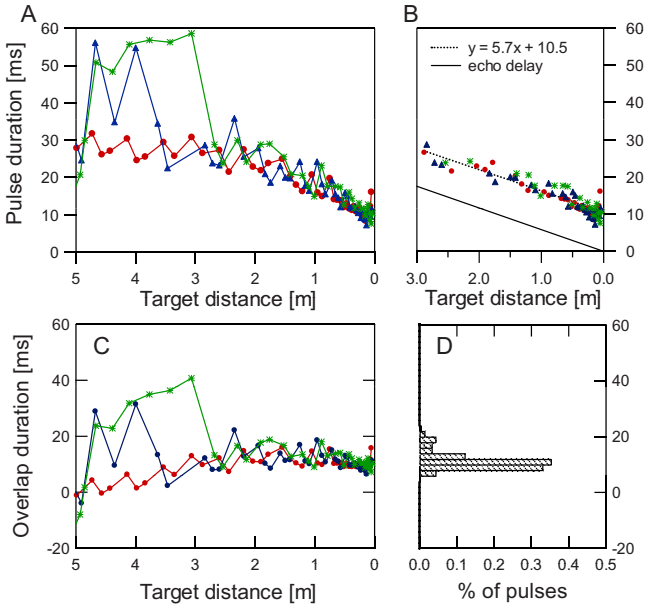


Fig. 2. (Color online) Duration of the pulse-echo overlap during landing flights. (A) Changes in the duration of the echolocation pulse as a function of target distance for three bats; different symbols indicate different bats. (B) Changes in the durations of shorter pulses (pulses other than the longest-duration pulse within one periodic alteration between short and long pulses). The solid line indicates the echo delay. The dashed line shows the linear regression fit of data combined from three flights. (C) Overlap durations between pulse and echo for the three flights. (D) Proportion representing the distribution of the overlap duration of shorter pulses when within 3 m of the target wall ($n=307$) for all flights.

caused beats in the amplitude pattern of the CF components [Fig. 1(B)]. By comparing echo delays measured from Telemike sounds to three-dimensional coordinate data, these overlapping echoes were identified as echoes from the target wall of the flight chamber [Fig. 1(C)]. During stereotyped landing flights, the bats changed the CF of their emitted pulses depending on changes in flight speed relative to the target wall (i.e., they demonstrated DSC). Bat flight speeds reached a maximum of 3–4 m/s in the chamber, and the bats decreased the CF_2 (the second harmonic of the CF component) frequency of the emitted pulse by 1.5–1.7 kHz from the resting frequency (RF) [Fig. 1(D)]. The CF_2 frequency of the echoes from the target wall (68.92 ± 0.06 kHz) was 110 Hz higher than the RF (68.81 ± 0.04 kHz) but was stable throughout a flight, indicating that bats focused on the target wall as their destination during the landing flight. Measured beat frequencies had maximum values of 1.5–1.7 kHz and decreased as a bat approached the target wall [Figs. 1(B) and 1(E)]. The authors confirmed that the measured beat frequency corresponded to the estimated value derived from the relative flight speed of the bat toward the target wall [Fig. 1(E)].

Bats shortened the pulse duration as they approached the target wall by alternately producing short- and long-duration pulses [Fig. 2(A)]. To demonstrate the presence of a pulse-echo overlap by shortening pulse duration with shorter target distance, the authors described the changes in the duration of shorter pulses (i.e., pulses other than the longest-duration pulses recorded within one alteration between short and long durations). Fluctuations in pulse duration [Fig. 2(A)] and the pulse-echo overlap [Fig. 2(C)] were obvious when the individual was more than 3 m from the target, but the duration of shorter pulses proportionally decreased with the echo delay when within 2–3 m [Fig. 2(B)]. The slope of the decrease in the durations of shorter pulses was 5.7 ms/m [dashed line in Fig. 2(B); linear regression fit, $d=5.7x+10.5$], which almost corresponded with the slope of the echo delay ($2/c=5.8$ ms/m where the sound velocity c in air is 344 m/s). As a result, the duration of the pulse-echo overlap $d_{p\text{-}e}$ was relatively stable for shorter-duration pulses, regardless of the target distance x,

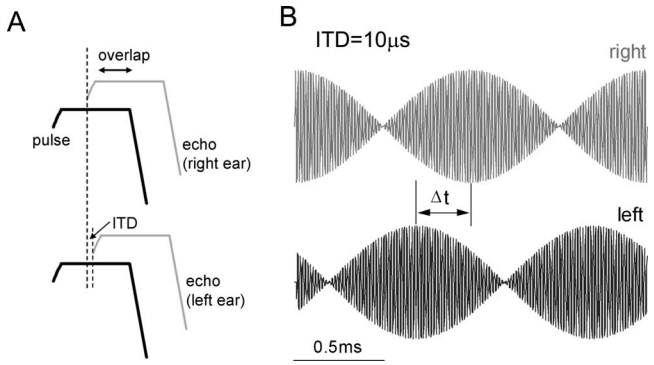


Fig. 3. (A) Diagram showing a pulse-echo interaction assumed to be returning from a point off to the side of the bat's flight path. (B) Estimated beat signals at right (top) and left (bottom) ears with an ITD of 10 μs. For the calculation, the CF_2 frequency of the pulse was 68 kHz, and the magnitude of the Doppler-shift was 1 kHz. The arrow indicates the phase difference between the peaks in the beats between the right and left ears.

$$d_{p\text{-}e} = d - \frac{2}{c}x \approx 10.5. \tag{2}$$

The observed duration of the pulse-echo overlap had a median value of 10.5 ms for shorter-duration pulses within 3 m from the target wall [$n=307$ from 12 flight sessions, Fig. 2(D)], and the authors confirmed that bats shortened pulse duration with target distance. However, individuals showed the presence of a pulse-echo overlap for at least approximately 8 ms within the approach phase (11.3 ± 3.1 ms for shorter pulses, $n=307$).

Rhinolophus ferrumequinum nippon shortened the pulse duration as target distance shortened. This suggests that after the approach phase starts, bats might adjust the pulse duration according to the echo delay from a large consistent target (i.e., the target wall). As a consequence of shortening pulse duration with shorter target distance, the duration of the pulse-echo overlap was relatively stable for a certain period of time. Although the authors cannot provide a definite reason for this species to show the existence of a relatively stable pulse-echo overlap for shorter-duration pulses, bats may obtain target information from the beats arising from the pulse-echo overlap during flight.

The envelope pattern of the beat signal depended on the pulse-echo interference; that is, beat frequency changed according to the frequency difference between a pulse and its echo in CF components [Fig. 1(B)]. On the other hand, the phase of the amplitude modulation in the beat signal changed according to the *onset* phase difference between a pulse and its echo. Figure 3 shows an example of the estimation of a pulse-echo overlapped signal at the right and left ears of a bat during flight, where the Doppler-shifted echo was assumed to be returning from an angle slightly off axis from the bat's flight path. The position of the envelope peak of the beat signal differed between the right and left ears [see Δt in Fig. 3(B)], indicating that the phase difference in the amplitude modulation occurred because of an interaural time difference (ITD). Henson *et al.* (1987) originally suggested that directional information from a target might be reflected in such phase difference in the amplitude modulation of the beat signal. When a bat receives an echo returning from a point off to side of the flight path, the position of the envelope peak of the beat signal varies between the right and left ears according to the ITD. In Fig. 3(B), the estimations demonstrate that the phase difference Δt in the envelope pattern of the beat signals between right and left ears was calculated to be 309 μs, which was almost 30 times larger than the ITD (for the calculation, the authors assumed that the CF_2 frequency of the pulse was 68 kHz, and the magnitude of the Doppler-shift was 1 kHz with an ITD of 10 μs). Because Δt varies with the ITD (when a target is located in front of a bat, Δt corresponds to 0), directional information regarding the target can be expressed in the phase difference of the slowly amplitude-modulated sound arising from the beat.

Because bats have small heads and use high frequency sounds, it is unlikely that ITD would be used as a critical cue for horizontal sound localization (Covey and Casseday, 1995). Instead of the ITD, periodic amplitude envelope change (easily detected peripheral change) caused by pulse-echo overlap might provide a useful cue for target localization. However, further investigation is required to demonstrate the kinds of information that bats can actually obtain from differences in the beat signals between their two ears. In this study, the authors introduced a hypothesis for acoustical localization using beat signals based on their Telemike recording data. Their computations show one possibility, namely, that pulse-echo interactions cause phase differences of the beat signal at the two ears, which could serve as a potential localization cue. In this context, beat signals caused by pulse-echo interactions in CF sounds could also provide an engineering perspective for extracting target information, such as sound localization.

Acknowledgments

This work was partly supported by a grant to the Research Center for Advanced Science and Technology at Doshisha University from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, Special Research Grants for the Development of Characteristic Education from the Promotion and Mutual Aid Corporation for Private Schools of Japan, and the Innovative Cluster Creation Project.

References and links

Covey, E., and Casseday, J. H. (**1995**). "The lower brainstem auditory pathways," in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 235–295.

Grinnell, A. D. (**1970**). "Comparative auditory neurophysiology of neotropical bats employing different echolocation signals," Zeitschrift fur Vergleichende Physiologie **68**, 117–153.

Henson, O. W., Jr., Bishop, A. L., Keating, A. W., Kobler, J. B., Henson, M. M., Wilson, B. S., and Hansen, R. (**1987**). "Biosonar imaging of insects by *Pteronotus p. parnellii*, the mustached bat," Nat. Geor. Res. **3**, 82–101.

Hiryu, S., Katsura, K., Lin, L. K., Riquimaroux, H., and Watanabe, Y. (**2005**). "Doppler-shift compensation in the Taiwanese leaf-nosed bat (*Hipposideros terasensis*) recorded with a telemetry microphone system during flight," J. Acoust. Soc. Am. **118**, 3927–3933.

Hiryu, S., Shiori, Y., Hosokawa, T., Riquimaroux, H., and Watanabe, Y. (**2008**). "On-board telemetry of emitted sounds from free-flying bats: Compensation for velocity and distance stabilizes echo frequency and amplitude," J. Comp. Physiol. [A] **194**, 841–851.

Schnitzler, H. U. (**1968**). "Die ultraschallortungslaute der hufeisen-fledermäuse (Chiroptera-Rhinolophidae) in verschiedenen orientierungssituationen [The ultrasonic sounds of horseshoe bats (Chiroptera-Rhinolophidae) in different orientation situations]," Zeitschrift fur Vergleichende Physiologie **57**, 376–408.

Schuller, G. (**1974**). "The role of overlap of echo with outgoing echolocation sound in the bat *Rhinolophus ferrumequinum*," Naturwiss. **61**, 171–172.

Schuller, G. (**1977**). "Echo delay and overlap with emitted orientation sounds and Doppler-shift compensation in the bat, *Rhinolophus ferrumequinum*," J. Comp. Physiol. [A] **114**, 103–114.

Suga, N., Simmons, J. A., and Jen, P. H. (**1975**). "Peripheral specialization for fine analysis of Doppler-shifted echoes in the auditory system of the 'CF-FM' bat Pteronotus parnellii," J. Exp. Biol. **63**, 161–192.

Tian, B., and Schnitzler, H. U. (**1997**). "Echolocation signals of the greater horseshoe bat (*Rhinolophus ferrumequinum*) in transfer flight and during landing," J. Acoust. Soc. Am. **101**, 2347–2364.

Determining material damping type by comparing modal frequency estimators

D. K. Anthony and F. Simón
Instituto de Acústica (Consejo Superior de Investigaciones Científicas), C/Serrano 144, 28006 Madrid, Spain
iaca344@ia.cetef.csic.es, psimon@ia.cetef.csic.es

Jesús Juan
Laboratorio de Estadística, Universidad Politécnica de Madrid, 28006 Madrid, Spain
jjuan@etsii.upm.es

Abstract: The accuracy of modal frequency and damping estimators for non-lightly damped single degree of freedom systems depend on the response parameter used as well as the damping mechanism. Therefore, in order to make accurate modal parameter measurements, the damping mechanism at play must be known to be either viscous or hysteretic *a priori*. Here, comparisons between the evaluated frequency values are used to glean this information. The damping mechanism of an experimental system (consisting of resilient layer and mass plate) is then determined using two simple modal parameter estimators and applying statistical methods.

PACS numbers: 43.40.At, 43.40.Yq, 43.55.Vj, 43.40.Tm [JM]
Date Received: March 9, 2009 **Date Accepted:** June 24, 2009

1. Introduction

There are many reported techniques that measure a modal frequency parameter or the modal damping of a single degree of freedom (SDOF) system. The frequency parameter of most interest is the natural frequency, f_n, which is the oscillation frequency without energy loss. However, there are also the damped frequency ($f_d = f_n\sqrt{1-\zeta^2}$) and resonance frequency ($f_r = f_n\sqrt{1-2\zeta^2}$) that are related to f_n by the system damping, which is expressed as the critical damping ratio (ζ) and defined in the text.

When using measurements to estimate the natural frequency and the modal damping of a system, the accuracy of the technique depends on the response parameter evaluated (e.g., velocity or a spectrum based on this) and also the system damping mechanism, especially more so for non-lightly damped systems. But for lightly damped systems, the assumption of a viscous damping model is usually sufficient as it is only necessary to account for energy loss in the system; the damping mechanism is not a significant factor on model performance, and it is seen that there is little difference in-between the three modal frequency parameters (indeed if the differences are small enough, the measurement resolution may not allow their distinction). However, viscous damping is strictly due to energy dissipation through fluid flow, and in many systems no such damping mechanism exists but instead the damping is provided through the deformation of solid materials (termed *hysteretic* damping).

Thus, in order to accurately determine modal parameters for non-lightly damped systems, it is important to use frequency and damping estimators that are accurate for the actual damping mechanism, and therefore this needs to be known *a priori*. However, in some cases the damping mechanism (or the dominant damping mechanism) at play may not be apparent. Here, the accuracy of two techniques is studied against the response parameter and the system damping type. Patterns of estimation errors for viscous and hysteretic damping can be used to glean the damping mechanism at play, and thus the results from appropriate modal parameter estimators are taken. This is then demonstrated experimentally. In this letter, the circular frequency, f, and the angular frequency, $\omega(=2\pi f)$, are used interchangeably, and are understood to refer to the same frequency with common subscripts.

2. Dual damping type SDOF system model

The description of a SDOF system either by its equation of motion or response in the frequency domain is available in many texts, for example, see Refs. 1–3. These deal with systems with both viscous and hysteretic damping. Based on these, a system model that allows for both types of damping has the equation of motion in displacement, x, with applied harmonic force, F_o, as

$$m\ddot{x} + \left(cK_v + \frac{h}{\omega}\bar{K}_v\right)\dot{x} + kx = F_o e^{j\omega t}. \tag{1}$$

m is the rigid system mass, c and h are the viscous and hysteretic damping factors, respectively, and k is the spring constant. K_v is a Boolean variable and $\bar{K}_v$ its logical ones-complement, and they determine the damping mechanism (K_v=1, viscous damping; K_v=0, hysteretic damping). The corresponding frequency response function (FRF) in terms of displacement, G_x, is found to be

$$G_x(\omega) = \frac{1}{m}\frac{1}{\omega_n^2 - \omega^2 + 2j\zeta\omega_n\omega K_v + j\eta\omega_n^2\bar{K}_v}, \tag{2a}$$

where

$$\omega_n = \sqrt{\frac{k}{m}} \tag{2b}$$

and

$$\zeta = \frac{c}{2\omega_n m}. \tag{2c}$$

ω_n is the system natural frequency and ζ is the *viscous damping ratio*. The hysteretic damping is represented by the *structural damping factor*, η. In the cited texts, (i) it can be seen that it is related to h and c, and (ii) its equivalent value of $\zeta(\zeta_{\text{eq}})$ at resonance is given. These relations are

$$\eta = h/k, \tag{3a}$$

$$\zeta_{\text{eq}} = \eta/2. \tag{3b}$$

Here, for convenience and generality, a *generalized damping factor*, α, is used that may be directly replaced by either ζ or ζ_{eq}, as pertinent to the system studied.

The FRF expressed using the velocity or acceleration responses can be derived from G_x using the generalized derivative property of the Fourier transform,[4] however, here a quadrature-corrected FRF, Q, is used which allows measurement methods to be applied to different response parameters while maintaining the similar characteristic responses in the real and imaginary spectral parts as for G_x. Such spectra have been used before in Refs. 5 and 6. Using the subscript p to represent the response parameter, where $p \in \{x, v, a\}$ (displacement, velocity, or acceleration, respectively), Q is defined as

$$Q_{\langle p\rangle}(\omega) = \frac{A_{\langle p\rangle}}{\omega_n^2 - \omega^2 + 2j\alpha\omega_n(K_v\omega + \bar{K}_v\omega_n)} = \frac{A_{\langle p\rangle}}{D} = \frac{A_{\langle p\rangle}[(\omega_n^2 - \omega^2) - 2j\alpha\omega_n(K_v\omega + \bar{K}_v\omega_n)]}{D^*D}, \tag{4}$$

where $*$ is the complex conjugate operator. The numerator part $A_{\langle p\rangle}$ is defined as

$$A_x = 1, \tag{5a}$$

$$A_v = \omega, \tag{5b}$$

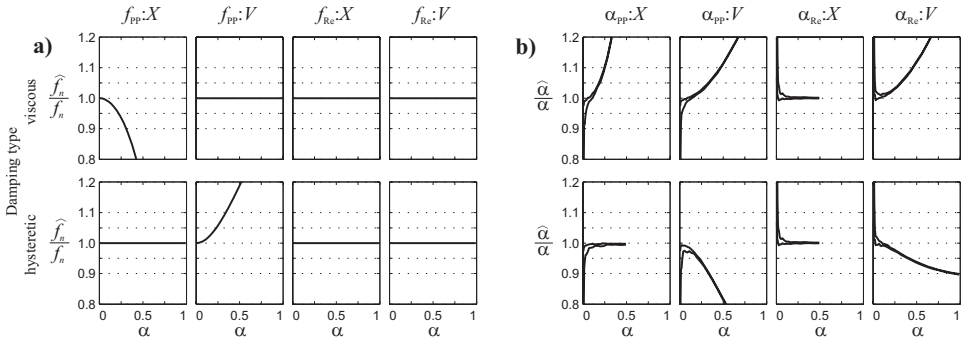


Fig. 1. The accuracy of (a) the evaluated frequencies (normalized by f_n) and (b) the evaluated values of damping (normalized by α) using $|Q|$ and $\mathrm{Re}\{Q\}$ for $p \in \{x, v\}$, against the generalized damping factor, α, for systems with viscous or hysteretic damping.

$$A_a = \omega^2, \tag{5c}$$

and thus $G_x \equiv Q_x$.

3. Modal parameter estimation

Here, modal parameters are estimated using techniques applied to the spectral magnitude [$|Q|$ or $\mathrm{abs}(Q)$] or the real part ($\mathrm{Re}\{Q\}$), and a notation is used whereby the frequency and damping evaluated have a subscript indicating the spectrum type used. Additionally, and where pertinent, the response parameter is indicated afterwards, separated by a colon. For example, f_{abs}:X represents the modal frequency measured from $|Q|$ applied to the system response G_x. Here the damping is calculated using an expression that is apt for non-lightly damped systems, and not one often found in texts that rely on simplifications applicable only to lightly damped systems.[2]

3.1 Modal natural frequency estimation

The determination of the modal frequency from $\mathrm{abs}(Q)$ is defined as the frequency where $\max(|Q_{\langle p \rangle}|)$ occurs,[1–3] and whose absolute value is found to be $A/\sqrt{D^* D}$, where A and D are defined above. Setting its derivative with respect to ω to zero, in order to find the maximum value, results in

$$D^* D(\alpha_v + 2\alpha_a \omega) = A\omega[K_v 4\zeta^2 \omega_n^2 \omega - 2\omega(\omega_n^2 - \omega^2)]. \tag{6}$$

From this the value of ω_{abs} evaluated in Eq. (6) for each K_v and $\langle p \rangle$ is

$$K_v = 1, \langle p \rangle = x \quad \Rightarrow \quad \omega_{\mathrm{abs}} = \omega_n \sqrt{1 - 2\zeta^2}, \tag{7a}$$

$$K_v = 1, \langle p \rangle = v \quad \Rightarrow \quad \omega_{\mathrm{abs}} = \omega_n, \tag{7b}$$

$$K_v = 0, \langle p \rangle = x \quad \Rightarrow \quad \omega_{\mathrm{abs}} = \omega_n, \tag{7c}$$

$$K_v = 0, \langle p \rangle = v \quad \Rightarrow \quad \omega_{\mathrm{abs}} = \omega_n \sqrt[4]{1 + 4\zeta^2}. \tag{7d}$$

Thus, for a system with hysteretic damping ω_n can be correctly determined by identifying $\max(|G_x|)$, and likewise for a viscously damped system by $\max(|Q_v|)$, as seen in Fig. 1(a). Otherwise ω_n is under- or over-estimated. In the former case, Eq. (7a), the frequency determined is the resonance frequency, ω_r, as expected.[1–3] Since the frequency measured depends on the damping mechanism, it must be known *a priori* in order to measure ω_n accurately using $\mathrm{abs}(Q)$.

Table 1. The equality and inequality patterns between the frequencies evaluated from the frequency estimators considered, for both damping types.

Damping type	K_v	Frequency estimators			
		$f_{abs}:X$	$f_{abs}:V$	$f_{Re}:X$	$f_{Re}:V$
Viscous	1	$<f_n$	$=f_n$	$=f_n$	$=f_n$
Hysteretic	0	$=f_n$	$>f_n$	$=f_n$	$=f_n$

The modal frequency defined using $\mathrm{Re}\{Q\}$ is the frequency where $\mathrm{Re}\{|Q_{\langle p\rangle}|\}=0$. From Eq. (4) it is easily seen that $\omega_{Re}=\omega_n$ independent of both the response parameter and the damping mechanism,[5] where the trivial solution $\omega=0$ is ignored. The use of acceleration ($\langle p\rangle$=a) is not considered further here.

3.2 Modal damping estimation

The damping can be estimated from both $\mathrm{abs}(Q_{\langle p\rangle})$ and $\mathrm{Re}\{Q_{\langle p\rangle}\}$ using two additional frequency points determined from each spectrum. This leads to the two independent damping measures:[2]

$$\alpha_{abs}=\frac{f_u^2-f_l^2}{4(f_{abs})^2}, \tag{8a}$$

$$\alpha_{Re}=\frac{f_b^2-f_a^2}{4(f_{Re})^2}, \tag{8b}$$

where f_l and f_u are the lower and upper half-power points on $|Q_{\langle p\rangle}|$, and f_b and f_a are the frequencies of the maximum and minimum values of $\mathrm{Re}\{|Q_{\langle p\rangle}|\}$, which are illustrated in Refs. 1 and 2. The upper and lower frequency points for each estimator are not necessarily the same.

Since the analysis of the damping requires two additional measurements theoretical analysis is complex and non-productive, and the analysis proceeds using a numerical model based on Eqs. (4) and (5). Spectra were generated over the full range of α ($0<\alpha<1$), and for various combinations of $\Delta f/f_n$ (where Δf is the frequency resolution). Then the damping was evaluated using each estimator ($\hat{\alpha}$) and the response parameters $\{x,v\}$, and is shown in Fig. 1(b) normalized by α for system with each damping type. It is noted that for smaller values of damping, there is a variation in the measured values that depends on $\Delta f/f_n$. This is due to the fortuitous alignment of the discrete points on the frequency axis with the true position of the characteristics of the spectrum.[5] However, here the focus is on non-lightly damped systems.

It is seen that $\alpha_{Re}:X$ determines the damping correctly independent of damping type. The measurement is limited to a value of 0.5 as the lower frequency point, f_a, moves toward dc and is no longer defined. It is also seen that $\alpha_{abs}:X$ also provides an accurate determination if the damping type is hysteretic. All other estimations provide increasingly significant errors with increasing damping, especially $\alpha_{PP}:X$ for viscous damping which is a commonly cited estimation technique.

4. Evaluation of damping type from parameter measurement patterns

For a non-lightly damped system, it can be seen from Fig. 1(a) that there are significant differences between the frequencies evaluated by different estimators, and that these depend on the damping mechanism. These can be reduced to equality and inequality patterns, as shown in Table 1. Thus, if one of these patterns can be identified, then the damping mechanism can be gleaned from this. In practice, a distribution of values would be available from a number of experimental measurements of frequency for each estimator, and thus the pattern identification proceeds on a statistical basis. It is noted that using $|Q|$ alone is not sufficient as for both damping types $f_{abs}:V$ is always greater than $f_{abs}:X$. In theory, it would be possible to also perform

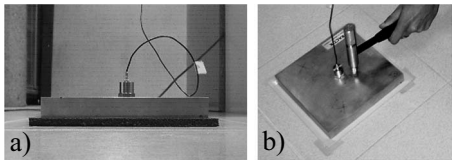


Fig. 2. The experimental system (a) showing the mass plate on top of the (black) polyurethane foam used here and (b) delivering an impact (using a different foam).

damping type identification from the relations between the values of the damping estimators. However, as the damping measurements depend on three frequencies, both the numerical variation (due to resolution) and statistical variation (due to repeated measurement in an experimental system) for each frequency would be compounded.

5. Application to experimental system to determine damping type

Measurements were taken from a system based on the characterization of materials used as resilient layers with application in the sound reduction of impact transmission within buildings.[7,8] An 8 kg solid mass plate was placed on the material under test (MUT) that is placed on a rigid massive base. The MUT was a 12-mm-thick closed-cell polyurethane foam with a density of $0.032\ \mathrm{g/cm^3}$. An impact hammer was used to excite the plate vertically, whose oscillatory motion was measured by an accelerometer, see Fig. 2. Eight impact measurements were performed daily for six MUT samples (labeled A–F), over 6 days (288 records in total). The records consisted of a 3 s measurement sampled at 16 393 samples/s, and each was analyzed using $|Q_{\langle p\rangle}|$ and $\mathrm{Re}\{Q_{\langle p\rangle}\}$ for $p \in \{x, v\}$. The acceleration time history is converted to the frequency domain by the Fourier transform and is normalized by the impact force spectrum to form G_a. Using the relations described in Sec. 2, the spectra G_x and Q_v are found, noting that $Q_a = -G_a$ (the accelerance FRF). Except for the removal of any dc offset on the acceleration signal, as the first resonance of the solid mass occurs near 3.3 kHz and can be ignored, the only further signal processing required is spectral smoothing using a Hamming window of length 11 data points.

Two typical velocity responses are shown in Fig. 3 where it is seen that the first part of the responses do not demonstrate regular exponential decay. For this reason only the latter part of the responses are used by truncating the velocity responses at the third zero-crossing point. These signals are treated as surrogate displacement responses and are also used to generate the corresponding velocity responses.[6]

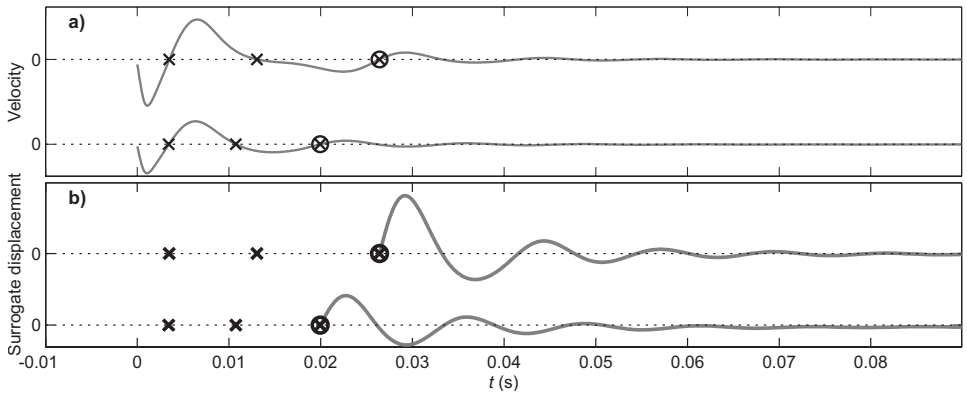


Fig. 3. (a) Two typical impact velocity responses, showing the first three zero-crossing points (×). (b) The responses truncated at the third crossing point (⊗) and shown to an expanded amplitude scale, used as surrogate displacement responses in order to construct the velocity in the latter part of the impact responses.

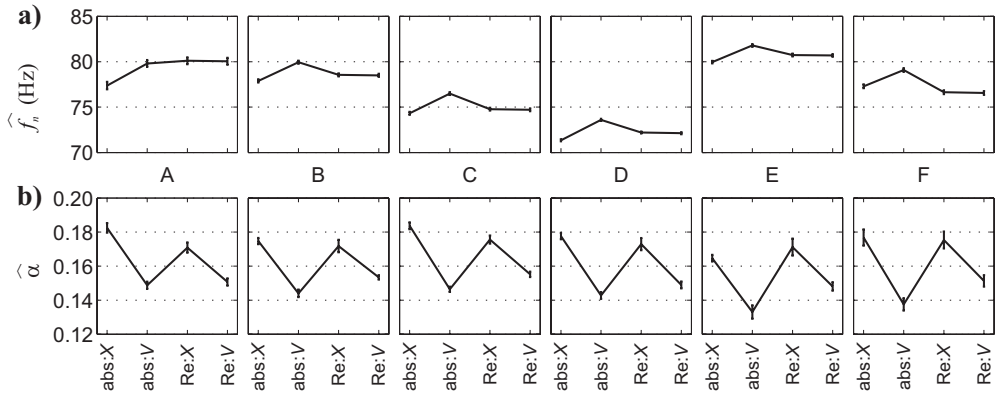


Fig. 4. The average values of (a) the evaluated frequency and (b) the damping for each estimator, for the six material samples (A–F). Error bars indicate the 95% confidence limits.

The evaluated frequency and damping values were analyzed using a randomized-block analysis of variance.[9] Figure 4 shows the average values for both the evaluated frequency ($\hat{f}_n$) and $\hat{\alpha}$ along with the individual 95% confidence limits. The variability between measurements for each sample is very small, as can be appreciated from the confidence limits. However, a larger variation exists between the corresponding data due to the accuracy of the estimators illustrating the heterogeneity of the MUT. With reference to Table 1, it is seen that the relations between $\hat{f}_n$ for MUT samples B–F suggest that a hysteretic damping mechanism is at play. The relations between the values for $\hat{\alpha}$ also support this. The concordance for sample C in $\hat{\alpha}$, however, is not as good. The values of $\hat{f}_n$ for sample A appear to indicate viscous damping, while the values for $\hat{\alpha}$ do not; it appears that sample A is acting differently. So, the values of f_n and α for the samples B–F can be correctly estimated using the estimators accurate for hysteretic damping. The error in using an inappropriate estimator is seen and is especially significant for $\hat{\alpha}$, where the error approaches 20% in the worst case.

6. Summary and conclusions

The accuracy of techniques used to measure modal parameters often depends on the response parameter used and the damping mechanism of the system, which must therefore be known *a priori*. Comparisons between different estimators have been used to allow the damping mechanism to be gleaned. The natural frequency and damping can also be accurately determined from the real spectral part independent of damping type. Unfortunately, this method is sensitive to phase errors.[5] It has been assumed here that one damping type is dominant; in some applications both types may be active. Finally, it is seen to be valuable to use statistical analysis to allow conclusions regarding the damping type to be drawn, due to variations in repeated experimental measurements.

Acknowledgments

The first author performed this work while on an invited stay at the named institution. This work has been partly supported by a research grant from the Spanish *Ministerio de Fomento* (C5/2006).

References and links

[1] *Fundamentals of Noise and Vibration*, edited by F. J. Fahy and J. G. Walker (E & FN Spon, Great Britain, 1998).
[2] D. J. Ewins, *Modal Testing: Theory, Practice and Application*, (Research Studies, Great Britain, 2000).
[3] *Vibration Damping, Control and Design*, edited by C. W. de Silva (CRC, Canada, 2007).
[4] A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems* (Prentice-Hall, Englewood Cliffs, NJ, 1983).

[5]D. K. Anthony, "Practical improvements to real and imaginary spectral based modal parameter measurements of SDOF systems," Appl. Acoust. **70**, 1219–1225 (2009).
[6]D. K. Anthony and F. Simón, "Generating 'idealised' impulse response functions to improve or repair single degree of freedom system measurements," Appl. Acoust. **70**, 531–539 (2009).
[7]EN 29052-1:1992, "Acoustics—Determination of dynamic stiffness—Part 1: Materials used under floating floors in dwellings" (1992).
[8]F. Simón and D. K. Anthony, "Comparison between different methods of characterizing elastic layers," in The 13 International Congress on Sound and Vibration (ICSV13), Vienna, Austria (2006), Paper No. 915.
[9]D. C. Montgomery *Design and Analysis of Experiments* (Wiley, New York, 1996).

Experimental investigation on pore size effect on the linear viscoelastic properties of acoustic foams

Mickaël Deverge
LAUM, CNRS, Université du Maine, Avenue O. Messiaen, 72085 Le Mans, France
mickael.deverge@univ-lemans.fr

Lazhar Benyahia
PCI CNRS, Université du Maine, Avenue. O. Messiaen, 72085 Le Mans, France
lazhar.benyahia@univ-lemans.fr

Sohbi Sahraoui
LAUM, CNRS, Université du Maine, Avenue. O. Messiaen, 72085 Le Mans, France
sohbi.sahraoui@univ-lemans.fr

Abstract: This paper presents linear viscoelastic measurement on a large frequency range (10^{-2}–10^{8} Hz) for cross-linked polymer open-cell foams of same density and different pore sizes. This large extension of frequency range is obtained by the validation of a frequency-temperature superposition principle, commonly used with polymers. At higher frequencies, the shear moduli are independent of the pore size. In acoustical insulation range (1 Hz–16 kHz), the shear moduli decreases with the foams' pore size.

PACS numbers: 43.50.Gf, 43.35.Mr, 43.20.Ye [MS]
Date Received: April 27, 2009 **Date Accepted:** June 30, 2009

1. Introduction

The use of cellular materials based on polymer foams has shown an increase interest these past decades, particularly in automotive industry for acoustic absorption.[1] Several methods for mechanical characterization in the linear domain were developed. Three principal measurement techniques can be currently distinguished.

- Quasi-static methods, limited to low frequencies; they offer an excellent frequency resolution.[2]
- Dynamic methods, based on sample resonance; they make it possible to cover an important frequency range but with a poor frequency resolution; moreover fluid-structure coupling effects cannot be neglected when measurements are not carried out in vacuum.[3]
- Acoustic methods, based on acoustical excitation and Rayleigh wave velocity measurements.[4]

An alternative to extend the frequency range for measurements with an enhanced resolution is the use of frequency-temperature superposition (FTS) principle, largely validated on polymers.[5,6] Indeed, Leaderman[7] noted that the behavior of a viscoelastic material at high temperature and short time is equivalent to its behavior at lower temperature and longer time. Thus, increasing temperature and reducing frequency (or inversely) are similar operations for these materials. In other words, this principle is based on the equivalence of temperature and frequency variations of some physical variables, such as shear moduli or viscosity. The application of FTS leads to a master curve, given for a reference temperature T_{ref}. Thus, the frequency is considerably extended by multiplying the frequency with a shift factor a_T depending on temperature. However, the FTS principle cannot be applied to all materials. It is not possible to determine *a priori* if a material satisfies this principle before performing the test. Usually, FTS

principle is not satisfied when two relaxation processes of different nature occur on the studied frequency range.[5] More details on this approach and on the master curve computing can be found in previous papers.[5,7,8]

Recently, the FTS method has been partially validated, on a reduced frequency range, by Sfaoui,[9] then Etchessahar[8] on polymer foams. In the present work, the authors extend this method for open-cell foams to a wider frequency range (from 10^{-2} to 10^{8} Hz). Actually, the authors analyzed the effect of pore size on the linear viscoelastic properties of open-cell foams having the same density and based on the same cross-linked polymer.

2. Material and method

Full open-cell foams are provided by Recticel® company. The authors selected four references from Bulpren® S family based on cross-linked polyurethane, which are used as heat or acoustic insulators. The samples have the same chemical composition, the same porosity (0.98), and the same mass density (29–31 kg/m^3). They only differ by their pore sizes: S20 (1.01–1.69 mm), S30 (0.72–1.01 mm), S60 (0.39–0.5 mm), and S90 (0.27–0.32 mm). The pore size was given by Recticel® in terms of minimum and maximum pore diameters from micrographs. For rheological measurements, cylindrical samples (diameter of 45 mm and height of 10 mm) were carefully cut and prepared.

Dynamic measurements of complex shear modulus G^* were carried out with a commercial apparatus from Rheometrics Scientific (RDA2) equipped with a heating and cooling oven. The choice of a non-resonant torsion technique ensures an excellent frequency resolution and allows simplified assumptions: Non-coupling fluid-structure effects, Poisson's ratio ν is real and independent of frequency.[10] To avoid any slip phenomenon, polyurethane foam samples were stuck by a two-sided adhesive tape between two parallel aluminum plates. The upper side of samples remains fixed and is connected to a torque sensor. The lower side of the samples was harmonically excited in torsion with a constant and controlled angular frequency ω: $\gamma(t) = \gamma_0 \sin(\omega t)$. $\gamma(t)$ was the imposed time dependent strain and γ_0 its amplitude. From the torque the authors deduced a stress $\sigma(t) = \sigma_0 \sin(\omega t + \delta)$ where σ_0 was the stress amplitude and δ was the phase angle. The stress equation can be written as $\sigma(t) = G'(\omega)\gamma_0 \sin(\omega t) + G''(\omega)\gamma_0 \cos(\omega t)$ where $G'(\omega)$ and $G''(\omega)$ represented, respectively, the elastic and viscous shear moduli. Thus at a given temperature these moduli depend only on frequency and are directly connected to the amplitudes σ_0 and γ_0 as well as to the phase angle δ by the following equation:

$$G^*(\omega) = G'(\omega) + iG''(\omega) \quad \text{with} \quad G'(\omega) = \frac{\sigma_0}{\gamma_0}\cos(\delta) \quad \text{and} \quad G''(\omega) = \frac{\sigma_0}{\gamma_0}\sin(\delta), \quad (1)$$

where i is the complex number ($i^2 = -1$).

Measurements were repeated at various temperatures T ranging from 20 to −25 °C. A latency of 10 min was necessary before each temperature measurement in order to guarantee temperature homogeneity in the sample. Indeed, while the temperature was fixed, the shear moduli were measured versus time at a frequency of 1 Hz in the linear regime. The authors found that after 10 min, a stationary regime was reached and the moduli remained constant. At each temperature, the frequency sweep was fixed between 0.016 and 16 Hz in the linear regime[5] where material response remained independent of the applied shear strain γ_0. Master curves of shear moduli were obtained by using time-temperature equivalence principle at $T_{ref} = 20$ °C.

3. Results and discussion

The shear moduli were plotted versus frequency in Fig. 1 for the sample S20 at different temperatures above the glass transition of the polyurethane ($T_g = -25$ °C). Either the elastic (G') or the viscous (G'') modulus showed a small dependency on the explored frequency range. The moduli were increased by a maximum factor around 5 between 10^{-2} and 20 Hz.

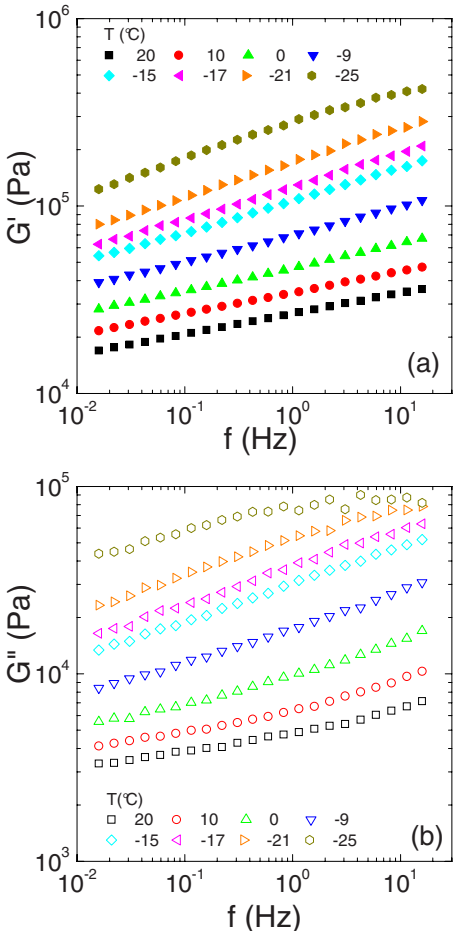


Fig. 1. (Color online) (a) Real part G' and (b) imaginary part G'' of shear modulus of Bulpren S20 foam, subjected to a strain of 0.05%, at various temperatures as indicated in the figure.

Figure 2 presented the master curves at a reference temperature T_{ref}=20 °C for all Bulpren S foams tested in this study. The determination of master curves was obtained by multiplying the frequency by a factor a_T and the moduli by a factor b_T. The factors a_T (not shown) were close to the one obtained in a previous study[8] and are independent of pore size. b_T are close to unity and reflect the mass density change with temperature.

All foams showed a similar dependence on the range of explored frequency. In particular, at high frequencies, the moduli were identical and independent of the pore size. Since the densities of all foams tested here were identical, the authors may assume that the moduli at high frequency depended on the volume fraction of the polymer. There was no effect of the microscopic structure (various pore sizes for various foams) but only response of polymer, which composes skeleton and was identical for the foams

At low frequencies (1 Hz–16 kHz), the shear moduli decreased slightly when increasing the foams' pore size. The dependence of the elastic modulus G' on the pore size, at lower frequency, was reported in the insert of Fig. 2. Similar observations were reported in literature but the opinions still diverge on how the elastic modulus of foams should depend on the pore size (see Refs. 11 and 12 for example, and references therein). Since the explored range of the pore size was limited, any previous empirical model was better to fit the experimental results.

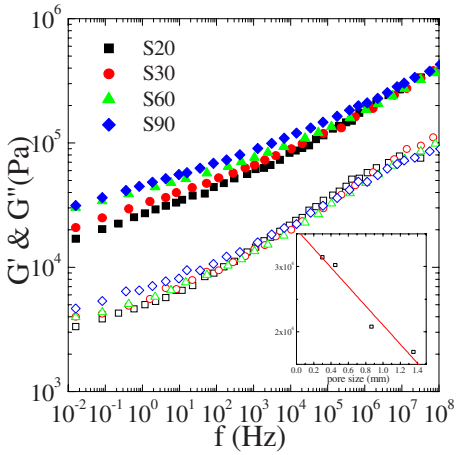


Fig. 2. (Color online) Master curves for whole Bulpren S foams: Real G' (closed symbols) and imaginary G'' parts (open symbols) of shear moduli, at reference temperature $T_{ref}=20$ °C. The insert represents the elastic shear modulus G' (at 10^{-2} Hz) versus the mean of pore size. The solid line is a guide for eyes.

4. Conclusion

In this work, complex shear modulus G^* was determined on a large frequency range (from 10^{-2} to 10^8 Hz) by using a quasi-static method coupled to a time-temperature superposition principle. The FTS principle was generalized here to cross-linked polymer open-cell foams and to a larger frequency range. The authors showed that FTS is useful to extend the frequency range in the polymer foams' study. The construction of shear moduli master curves revealed that (a) at high frequency, G' and G'' were independent of the pore size having identical mass density, since the moduli depended on the polymer, which composes the skeleton of the foams at high frequency, and that (b) at low frequencies (1 Hz–16 kHz), G' and G'' decreased when decreasing frequency, this trend being more pronounced as the pore size increased.

References and links

[1]R. Deng, P. Davies, and A. K. Bajaj, "Flexible polyurethane foam modelling and identification of viscoelastic parameters for automotive seating applications," J. Sound Vib. **262**, 391–417 (2003).
[2]S. Sahraoui, E. Mariez, and M. Etchessahar, "Mechanical testing of polymeric foams at low frequency," Polym. Test. **20**, 93–96 (2000).
[3]T. Pritz, "Dynamic Young modulus and loss factor of plastic foams for impact sound isolation," J. Sound Vib. **178**, 315–322 (1994).
[4]J. F. Allard, M. Henry, L. Boeckx, P. Leclaire, and W. Lauriks, "Acoustical measurement of the shear modulus for thin porous layers," J. Acoust. Soc. Am. **117**, 1737–1743 (2005).
[5]J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1961).
[6]M. L. Williams, R. F. Landel, and J. D. Ferry, "The temperature dependence of relaxation mechanisms in amorphous polymers and other glass-forming liquids," J. Am. Chem. Soc. **77**, 3701–3707 (1955).
[7]H. Leaderman, "Textile materials and the time factor: I. Mechanical behavior of textile fibers and plastics," Text. Res. J. **11**, 171–193 (1941).
[8]M. Etchessahar, S. Sahraoui, L. Benyahia, and J. F. Tassin, "Frequency dependence of elastic properties of acoustic foams," J. Acoust. Soc. Am. **117**, 1114–1121 (2005).
[9]A. Sfaoui, "On the viscoelasticity of the polyurethane foam," J. Acoust. Soc. Am. **97**, 1046–1052 (1995).
[10]T. Pritz, "Measurement methods of complex Poisson's ratio of viscoelastic materials," Appl. Acoust. **60**, 279–292 (2000).
[11]K. Li, X. L. Gao, and A. K. Roy, "Micromechanics model for three-dimensional open-cell foams using a tetrakaidecahedral unit cell and Castigliano's second theorem," Compos. Sci. Technol. **63**, 1769–1781 (2003).
[12]E. Olevsky and A. Molinari, "Kinetics and stability in compressive and tensile loading of porous bodies," Mech. Mater. **38**, 340–366 (2006).

Comparison of vu-meter-based and rms-based calibration of speech levels

Mead C. Killion
Etymotic Research, Inc., 61 Martin Lane, Elk Grove Village, Illinois 60007
m_killion@etymotic.com

Abstract: A difference of approximately 5 dB exists between the level of spoken English determined using the ANSI standard vu-meter method compared to the common root-mean-square (rms) method. If the rms method is substituted for the present ANSI standard method for calibrating a speech audiometer, for example, the reported speech reception thresholds will improve 5 dB: Speech levels read approximately 5 dB less using rms. Similarly, the reported signal-to-noise ratio required to understand speech in a speech-spectrum noise will be 5 dB better using rms. A simple method for obtaining a close approximation to traditional calibrations using a modified rms method is given.

PACS numbers: 43.71.An, 43.66.Yw [QJF]
Date Received: March 5, 2009 **Date Accepted:** July 9, 2009

1. The historical method

In 1940, Chinn *et al.* 1940 published the paper describing the vu meter. The vu meter was fairly soon adopted for speech research and audiology, and has been the basis for the standard method for setting the calibration tone on a speech recording, which is that "…the rms sound pressure level of a 1000 Hz signal [is] adjusted so that the vu meter deflection produced by the 1000 Hz signal is equal to the average peak vu meter deflection produced by the speech signal." These instructions go back to Z24.13-1953 (American National Standard Specifications for Audiometers) and continue unchanged through ANSI Standard S3.6-1969 (1969) to the present ANSI Standard S3.6-2004 (2004). The above quote is found in Sec. 6.2.11 of the latter standard. The standard calibration method was employed in the classic NU-4 and NU-6 speech tests (Tillman *et al.* 1963; Tillman and Carhart, 1966) and the more recent MIT female-talker recordings (Rabinowitz *et al.*, 1992) used in the QuickSIN test (Killion *et al.*, 2004).

As taught by Tillman, the most accurate readings require two readings of the vu meter for a given segment of recorded material. In the first pass, the reader notes about where on the meter each peak occurs. On the second pass (or third, as needed), the reader's eyes are fixed on the approximate location of a given peak. Using this method, the exact reading for each peak can be obtained within 0.1–0.2 dB. The author has personally trained skeptical colleagues and found that their "average of frequent peaks" results agree with the author's within 0.2 dB, even though no prescription of which peaks to use was given other than "typically two to three peaks per sentence."

2. Experiment 1

Ludvigsen (1992) recently compared various measures of speech level. The vu-meter method was not included, leading the present author to exchange DAT recordings of running speech for measurement and correlation with Ludvigsen (1992). Ludvigsen (1992) later simulated in software the vu-meter ballistics and 1.4 power characteristic of the copper-oxide rectifier used in the standard vu meter.

Figure 1 shows the output of Ludvigsen's software-simulated vu meter for 30 s of running speech. The 0 dB reference value on the ordinate is the normalized rms value for this 30 s of running speech after pauses and silent periods had been removed.

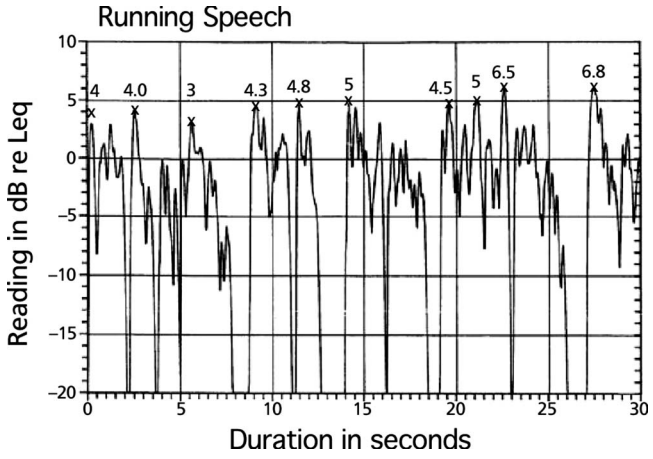


Fig. 1. Simulated (×) and actual vu-meter readings (numbers) for running speech with rms value of 0 dB.

After measuring the rms value of the same speech sample in the same way, the author made vu-meter readings as described above on a 25-year-old laboratory-reference vu meter belonging to the author. For this purpose, the analog output of the DAT recorder was adjusted so that the maximum peak stayed on the vu-meter scale, after which the individual vu-meter readings were normalized by the rms level. A 1 kHz sine wave was used to transfer the rms level measured on COOLEDIT to the vu meter.

Ludvigsen's vu-meter simulation and the author's direct reading of a vu meter with standard ballistics showed the average of the frequent peaks of the vu meter as 4.8 dB above the rms reading of the speech.

Because of the possibility of background noise in the speech sample, the removal of pauses and silent periods was done in this experiment using both eye and ear: The waveform cuts were determined from the appearance of the waveform on the computer screen, combined with listening for an error in the selection. If the background noise level is low enough and the talker does not make extraneous noises, automatic methods may be employed.

3. Experiment 2

Sixteen of the sentences recorded by Rabinowitz *et al.*, (1992) were chosen for a second comparison of the vu and rms methods. The original recordings of IEEE sentences of Rabinowitz *et al.* 1992 were made during their study of the contribution of visual cues to speech intelligibility. The audio recording of one of their female talkers was used, with permission, for the target speech in the "QuickSIN" speech-in-noise test later described by Killion *et al.* (2004).

Out of the 96 sentences on the QuickSIN CD, the 16 sentences with a 25 dB signal-to-noise ratio (SNR) (or more precisely, talker-to-babble ratio) were selected for analysis. The peak vu-meter reading and rms value of each of those sentences was determined as described above. The presence of the four-talker babble on the QuickSIN sentences with 25 dB SNR was estimated to have a negligible effect (less than 0.02 dB) on the composite level, compared to the original recordings of Rabinowitz *et al.* (1992).

The average of the peak vu-meter readings across sentences was 0.2 dB above the vu-meter reading of the calibration tone, consistent with the ANSI standard vu-meter method Rabinowitz *et al.* (1992) used for calibrating the speech levels on their recordings.

More importantly, the rms level of the 16 sentences (pauses deleted) was 4.6 dB below the calibration tone.

Taking into account the +0.2 dB difference in vu-meter-to-cal-tone reading, the difference between rms and vu-meter reading found in Experiment 2 was exactly 4.8 dB, the same as in Experiment 1.

Thus the same 4.8 dB difference obtained earlier by Ludvigsen (1992) and the author using a single sentence and a male talker was obtained again with 16 sentences and a female talker. The standard deviation of the rms and vu methods, incidentally, was nearly identical across the individual sentences.

It is worth noting that a total of 30 s of speech was used in Experiment 1, and approximately 60 s of speech was used in Experiment 2 for the 16 sentences.

4. A simple COOLEDIT method for approximating vu-meter readings

In a search for a simple COOLEDIT-based method of obtaining speech levels equivalent to vu-meter readings, the author highlighted, on the COOLEDIT waveform display, 50 ms segments of the two or three highest-amplitude portions of the speech waveform in each sentence. Using the *Analyse, Statistics* option available in COOLEDIT, the *Total RMS Power* in dB re full scale was obtained for each chosen segment. The average of those 2 or 3 dB values was examined as an approximation to a vu-meter reading by looking at the rms-to-vu-meter difference for that sentence. That method, using the sentence "A rod is used to catch pink salmon," gave an rms-to-simulated-vu difference of 4.7 dB, indicating a good approximation was obtained.

5. Discussion

Both the rms and vu-meter methods provide accurate measures of speech levels. It should be possible to use one and add or subtract 4.8 dB as a good estimate of the other. Given the difference, however, it is important to report the method used to adjust the calibration tone on any recording.

It is even more important to provide information as to the method used in an experiment, to avoid continuing confusion with reports of surprisingly good SNR performance from digital signal processing of speech in broadband noise when the apparent "improvement" has been augmented by use of a rms rather than a vu measure of speech levels. To explain, both the rms and vu-meter methods will yield essentially identical values for broadband random noise, but approximately a 5 dB difference for actual speech.

Acknowledgments

Carl Ludvigsen wrote the software to simulate the ANSI standard vu-meter characteristics in 1993, and provided the graphical comparison between rms and vu readings shown in Fig. 1. He gave his permission to use Fig. 1, but reported that his vu-meter code was written in Borland Pascal and is now buried in backup tapes. Fortunately, Lobdell and Allen (2007) recently developed a new vu-meter-simulation code that meets the ANSI standard and, more importantly, is readily available. In their paper, they illustrate the practical use of this code in obtaining new information regarding the statistics of speech.

References and links

ANSI Standard S3.6-1969 (**1969**). "Specification for audiometers."
ANSI Standard S3.6-2004 (**2004**). "Specification for audiometers."
Chinn, H. A., Gannett, D. K., and Morris, R. M. (**1940**). "A new standard volume indicator and reference level," Proc. IRE **28**, 1–8.
Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., and Banerjee, S. (**2004**). "Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **116**, 2395–2405.
Lobdell, B. E., and Allen, J. B. (**2007**). "A model of the VU (volume-unit) meter, with speech applications," J. Acoust. Soc. Am. **121**, 279–285.
Ludvigsen, C. (**1992**). "Comparison of certain measures of speech and noise level," Scand. Audiol. **21**, 23–29.
Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). "Relations among different measures of speech reception in subjects using a cochlear implant," J. Acoust. Soc. Am. **92**, 1869–1881.
Tillman, T. W., Carhart, R., and Wilber, L. (**1963**)."A test for speech discrimination composed of CNC monosyllabic words (N. U. auditory test No. 4)," Report No. SAM-TDR-62, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.
Tillman, T. W., and Carhart, R. (**1966**). "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University auditory test no. 6," Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Aerospace Medical Division (AFSC), Brooks Air Force Base, Texas.
Z24.13-1953 (**1953**). "American National Standard, Specifications for Audiometers."

Phonetically optimized speaker modeling for robust speaker recognition

Bong-Jin Lee, Jeung-Yoon Choi, and Hong-Goo Kang
Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-dong Seodaemoon-gu Seoul 120-749, Korea
lbjcom@dsp.yonsei.ac.kr, jychoi@yonsei.ac.kr, hgkang@yonsei.ac.kr

Abstract: This paper proposes an efficient method to improve speaker recognition performance by dynamically controlling the ratio of phoneme class information. It utilizes the fact that each phoneme contains different amounts of speaker discriminative information that can be measured by mutual information. After classifying phonemes into five classes, the optimal ratio of each class in both training and testing processes is adjusted using a non-linear optimization technique, i.e., the Nelder–Mead method. Speaker identification results verify that the proposed method achieves 18% improvement in terms of error rate compared to a baseline system.

PACS numbers: 43.72.Pf, 43.72.Fx, 43.72.Kb [DO]
Date Received: April 23, 2009 **Date Accepted:** July 21, 2009

1. Introduction

Automatic speaker recognition, the task of verifying a speaker's identity using his/her voice, has been greatly extended in recent years.[1] It often uses spectral features such as Mel-frequency cepstral coefficients (MFCCs) but characteristics of spectral features vary along with the phonetic contents of input speech. Since speaker recognition systems are designed for text-independent application for flexibility, the spectral variation in the input signal is always very high. Due to limitations on the number of input features, however, it is not easy to include all the characteristics of a speaker in a stochastic model such as Gaussian mixture model (GMM).[2,3] In other words, to improve speaker recognition performance, it is very important to build a GMM model that represents speaker characteristics in realistic conditions. Various approaches have been proposed to overcome this limitation by utilizing phoneme information.[4–7]

In this paper, we also propose a method that utilizes phoneme information to improve speaker recognition performance. While previous studies focus on using separate models for each phoneme and combining scores,[4,6,7] we focus on finding an optimal phoneme class ratio, the portion of each phoneme class in an utterance, that maximizes speaker recognition performance based on mutual information. In speaker recognition, some researchers use this measurement to measure or improve speaker recognition accuracy.[8,9] In this paper, we experimentally re-evaluate the speaker discriminative power of each phoneme class using mutual information and then find the optimal phoneme class ratio. We adopt the Nelder–Mead method, which is widely used for nonlinear optimization of multi-dimensional data.[10] Experimental results show that the optimal phoneme class ratio we find is somewhat different from that occurring in normal speech: the portion of consonants is increased, but that of vowels is reduced. This approach can be applied to both training and testing, but the improvement is more significant when it is used for testing. Speaker identification results show that the proposed system that uses the optimal phoneme class ratio has around 18% better performance than a conventional system.

The rest of this paper is organized as follows. First, we review the concept of mutual information, which measures the speaker discriminative power of given speech signals, and suggest how to find the optimal phoneme class ratio based on aspects of information theory in Sec. 2. Section 3 shows the experimental setup and results, which verify the usefulness of the proposed algorithm. The conclusions and future work are given in Sec. 4.

2. Optimization of Phoneme Class Ratio

2.1 Problem formulation

Mutual information represents the amount of information shared by two given random variables. Equation (1) represents the definition of mutual information.

$$I(C;X) = H(C) - H(C|X). \tag{1}$$

In the equation, $H(C)$ denotes the entropy of a specific speaker presence, and $H(C|X)$ denotes the entropy of specific speaker presence when the feature set X is given. Eriksson *et al.*[8] showed that the error rate of a speaker recognition system decreases as mutual information increases. In Eq. (1), $H(C)$ can be simplified by the following equation assuming that the distribution of the speaker presence probability is uniform.

$$H(C) = -\sum_{s=1}^{S} P(s)\log P(s) = \log S, \tag{2}$$

where S denotes the number of speakers and $P(s)$ is the probability of each speaker's presence. Since $H(C)$ only depends on the number of speakers, $H(C|X)$ is the only term to affect $I(C;X)$. It is evident that minimizing $H(C|X)$ achieves maximization of $I(C;X)$. Thus, our goal is to minimize $H(C|X)$ using the following definition:

$$H(C|X) = \int_{\mathbf{x}\in X} p(\mathbf{x})H(C|\mathbf{x})d\mathbf{x} = -\int_{\mathbf{x}\in X} p(\mathbf{x})\sum_{s=1}^{S} P(s|\mathbf{x})\log_2 P(s|X=\mathbf{x})d\mathbf{x}. \tag{3}$$

Equation (3) can be approximated by the law of large numbers

$$H(C|X) \approx -\frac{1}{N}\sum_{n=1}^{N}\sum_{s=1}^{S} P(s|\mathbf{x}_n)\log_2 P(s|\mathbf{x}_n), \tag{4}$$

where $\mathbf{x}_n$ is the nth feature and N is the total number of features. We can classify $\mathbf{x}_n$ into K classes using pre-defined class information.

$$H(C|X) \approx -\frac{1}{N}\sum_{k=1}^{K}\sum_{n_k=1}^{N_k}\sum_{s=1}^{S} P(s|\mathbf{x}_{n_k,k})\log_2 P(s|\mathbf{x}_{n_k,k}), \tag{5}$$

where N_k is the number of features in the kth class and $x_{n_k,k}$ is the nth feature in the kth class. Trivially, $\Sigma_{k=1}^{K} N_k = N$. The portion of each class's entropy can be represented by

$$H_k(\mathbf{p}) = -\frac{1}{N_k}\sum_{n_k=1}^{N_k}\sum_{s=1}^{S} P(s|\mathbf{x}_{n_k,k})\log_2 P(s|\mathbf{x}_{n_k,k}), \tag{6}$$

where $\mathbf{p}$ is the vector whose element contains the ratio of each class to all features, i.e., $p_k = N_k/N$. Thus, it also satisfies the following constraint:

$$\sum_{k=1}^{K}\frac{N_k}{N} = \sum_{k=1}^{K} p_k = 1. \tag{7}$$

$P(s|\mathbf{x}_{n,k})$ is defined as follows:

$$P(s|\mathbf{x}_{n_k,k}) = \frac{p(\mathbf{x}_{n_k,k}|\lambda_{s,\mathbf{p}})}{\sum_{s'=1}^{S} p(\mathbf{x}_{n_k,k}|\lambda_{s',\mathbf{p}})}, \tag{8}$$

where $\lambda_{s,\mathbf{p}}$ is the GMM of speaker s trained by features with the class ratio of $\mathbf{p}$, and $p(\mathbf{x}|\lambda_{s,\mathbf{p}})$ is the likelihood of feature $\mathbf{x}_{n_k,k}$ given $\lambda_{s,\mathbf{p}}$. Thus, we can rewrite Eq. (5) using $H_k(\mathbf{p})$:

$$H(C|X) \approx \sum_{k=1}^{K} p_k H_k(\mathbf{p}). \tag{9}$$

Similarly, if we define $I_k(\mathbf{p})$ to represent the portion of class k in $I(C|X)$, $I(C|X)$ can be rewritten as follows:

$$I(C|X) \approx \sum_{k=1}^{K} p_k I_k(\mathbf{p}), \tag{10}$$

where $I_k(\mathbf{p})$ is defined as follows:

$$I_k(\mathbf{p}) = H(C) - H_k(\mathbf{p}). \tag{11}$$

Now we have to be concerned about the redundancy of each class. Redundancy of a class k usually increases as the class ratio p_k increases. Increasing redundancy caused by including unnecessary data actually degrades speaker recognition performance. For example, in our preliminary experiments on varying the ratio of vowels and consonants, we find that speaker identification performance is better when the ratio of vowels is 80% compared to 90%, even though it is known that vowels have more speaker discriminative information than consonants. Thus, we also need to consider the redundancy of the data while maximizing mutual information by controlling the phoneme class ratio. One simple solution is removing the p_k term from Eq. (10) because it directly relates to redundancy. Equation (12) shows the modified equation:

$$I(\mathbf{p}) \equiv I(C|X) \approx \frac{1}{K}\sum_{k=1}^{K} I_k(\mathbf{p}). \tag{12}$$

The equation denotes the average $I_k(\mathbf{p})$ for all classes. Therefore, the objective of the proposed algorithm is finding the optimal ratio of phoneme classes that maximizes $I(\mathbf{p})$:

$$\mathbf{p}_{\text{opt}} = \arg\max_{\mathbf{p}} I(\mathbf{p}). \tag{13}$$

There is one more issue about the relation between mutual information and speaker recognition accuracy. According to Eriksson *et al.*,[8] the relation between mutual information and speaker recognition accuracy becomes exact as the speaker recognition accuracy increases. If the portion of any class is very small, we cannot say that the class has a large amount of speaker discriminative information even though mutual information of that class is large. In this case, we may assume that the $I_k(\mathbf{p})$ of the class is meaningless. Thus, we force $I_k(\mathbf{p})$ to zero when p_k is smaller than a certain threshold θ, smaller than a minimum probability, and $I_k(\mathbf{p})$ is larger than a minimum $I_k(\mathbf{p}_{\min})$. Equation (14) shows the modification rule:

$$I_k(\mathbf{p}) = \begin{cases} 0 & \text{when } \begin{cases} p_k < \theta, \text{ and} \\ p_k < p_{k,\min}, \text{ and} \\ I_k(\mathbf{p}) > I_k(\mathbf{p}_{\min}) \end{cases} \\ I_k(\mathbf{p}) & \text{otherwise}. \end{cases} \tag{14}$$

In the equation above, θ denotes the threshold of p_k. If p_k is larger than θ, it means that we can use $I_k(\mathbf{p})$ and if p_k is smaller than θ, we do not use $I_k(\mathbf{p})$ because it is unlikely that the value is

meaningful. $p_{k,\min}$ and $I_k(\mathbf{p}_{\min})$ denote the minimum $I_k(\mathbf{p})$ and corresponding p_k. In this paper, we experimentally find $p_{k,\min}$ and $I_k(\mathbf{p}_{\min})$ during the optimization process, as it is hard to find a global minimum value theoretically.

2.2 Optimization method to find the optimal ratio

Since the speaker model $\lambda_{s',\mathbf{p}}$, which is given in Eq. (8), should be retrained by the expectation-maximization algorithm whenever $\mathbf{p}$ varies, we cannot directly find a $\mathbf{p}$ that maximizes Eq. (12). In this case, the Nelder–Mead method popularly used for nonlinear optimization in many-dimensional data is suitable.[10] Generally, the Nelder–Mead method is an unconstrained optimization method but our application has two constraints. One is given in Eq. (7), and the other is as follows:

$$0 \leq p_k \leq 1 \quad \text{for all } k. \tag{15}$$

Inclusion of these constraints does not affect the applicability of the method. The first constraint, Eq. (7), does not affect the result if we choose the initial vertices to satisfy the constraint. To satisfy the second constraint, we adjust the coefficients that are used to find the reflection and new vertex to ensure the vertices are located in a suitable range.

3. Experiments and Results

3.1 Experimental setup

We perform experiments to verify the feasibility of the proposed algorithm with the TIMIT corpus, which contains phoneme information for all sentences.[11] From the TIMIT corpus, each phoneme is classfied into one of seven classes: *stops, affricates, fricatives, nasals, glides* and *semivowels, vowels*, and *others*. Among these classes, affricates take up a very small portion of the TIMIT corpus and some speakers do not have this class; others comprise labels which are not speech segments. Thus, we only use five classes, disregarding affricates and others. MFCCs up to order 20 are extracted using a 20 ms windowed speech signal and the analysis frame is shifted every 10 ms in the baseline system. The boundary regions of each phoneme are omitted to remove the effect of transition regions. In training the speaker model using extracted MFCCs, we construct 16-mixture GMMs using five sentences for each speaker. After training GMMs for all speakers, we evaluate $I(\mathbf{p})$ using two sentences that are not used in training. When we evaluate the performance of the system, we use the remaining three sentences.

In the experiments, we have to extract features that are adjusted in the given phoneme class ratio in every iteration of the algorithm because the speaker model $\lambda_{s,\mathbf{p}}$ is retrained during the algorithm according to the variation in the phoneme class ratio $\mathbf{p}$. To adjust the phoneme class ratio, we analyze the speech signal every 0.25 ms and sample features from each class to satisfy the given class ratio p_k. Therefore, the number of features N_k that belongs to class k becomes as follows:

$$N_k = Np_k, \tag{16}$$

and we set N as the number of features when the analysis interval is 10 ms. This adjusting method is also used in the testing procedure. In the Nelder–Mead method, we initialize the vertices as suitable values around conventional phoneme class ratios and repeat the algorithm 100 times to converge the vertices.

3.2 Results and analysis

Before performing experiments, we evaluate the phoneme class ratio $\mathbf{p}$, the modified mutual information $I(\mathbf{p})$, and $I_k(\mathbf{p})$ of each class from the TIMIT corpus. Table 1 shows the results. As the table shows, $I_k(\mathbf{p})$ of vowels is larger than any other class. It means that vowels have more speaker discriminative information. Moreover, $I_k(\mathbf{p})$ of nasals is quite large even though the nasals ratio is just 5.78%, which confirms the results presented by Eatock and Mason.[5]

Table 1. Class ratio of baseline system and the proposed system when θ=0.08 and $I_k(\mathbf{p})$ of each class in TIMIT corpus.

	Ratio (%)			
Class	Baseline	Proposed	Difference	$I_k(\mathbf{p})$
Stops	6.94	7.34	+0.40	3.66
Fricatives	18.01	18.40	+0.39	3.20
Nasals	5.78	13.40	+7.62	5.89
Semivowels	12.83	8.40	−4.43	5.83
Vowels	56.44	52.46	−3.98	6.52
All	100	100	…	5.59

Next, we estimate the optimal class ratio using the proposed method and compare the speaker recognition error rate of the proposed system with the baseline system. The optimal threshold θ, which is the boundary of the compensation of $I_k(\mathbf{p})$, is found experimentally by varying from 0.0 to 0.5. Tables 1 and 2 show the proposed class ratio and the speaker identification results of the proposed system when the threshold θ=0.08. Table 1 shows the proposed class ratio and the baseline class ratio for comparison. As the table shows, the ratio of consonants is increased and that of vowels is decreased. The ratio of stops and fricatives is increased about 0.4% and that of nasals is increased more than 7%. On the other hand, the ratios of semivowels and vowels are decreased 4.43% and 3.98%. From this result, we can see that nasals are important to improve the speaker recognition performance. Also, semivowels and vowels have more redundant information than other classes even though they greatly contribute to the performance of the speaker recognition system. Table 2 shows the results of speaker identification tests of the proposed system and the baseline system. The table shows the average error rate, 95% confidence interval, and the minimum and the maximum of the confidence interval. The error rate of the proposed system is 18.33% lower than that of the baseline system on average. From these results, we can confirm that the performance of the proposed algorithm is superior to conventional algorithms.

We next apply the proposed method of adjusting the class ratio to the training and testing procedures to investigate the effect of each procedure. Figure 1 shows the result. The *y*-axis denotes the speaker recognition error rate with 95% confidence interval, and the *x*-axis denotes the type of system. *Baseline* and *proposed* denote the systems that use the phoneme class ratio of the baseline or the proposed system for both training and testing. $P+B$ means that the proposed class ratio is used for the training procedure and the baseline class ratio is used for the testing procedure. $B+P$ is the opposite of $P+B$. As the figure shows, the class ratio in the testing procedure influences the result more than that in the training procedure. In other words, the selection of test segments is more important than more accurate modeling of the distribu-

Table 2. Speaker identification error rate of baseline system and the proposed system θ=0.08.

	Error rate (%)	
	Baseline	Proposed
Average	4.571	3.734
95% conf.	0.118	0.104
Min	4.463	3.629
Max	4.690	3.838
Improvement	…	18.33

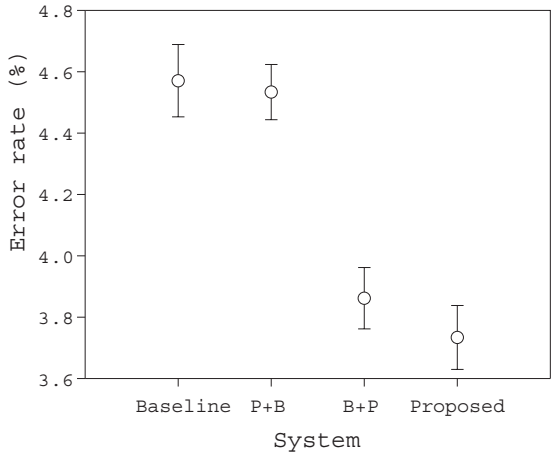


Fig. 1. Speaker identification error rate of the proposed system. P+B: training-proposed/testing-baseline; B+P: training-baseline/testing-proposed.

tion. In addition, we can say that speaker recognition performance can be improved by adjusting the phoneme class ratio of test data even if the speaker model has already been trained by conventional methods. Of course, the best performance can be achieved when both training and test methods adopt the proposed class ratios.

4. Conclusions and Future Work

In this paper, we proposed a method for finding an optimal phoneme class ratio that utilizes mutual information to improve speaker recognition performance. First, we defined $I_k(\mathbf{p})$, the portion of a class k in mutual information $I(C|X)$, and proposed a method for finding an optimal phoneme class ratio by maximizing the average $I_k(\mathbf{p})$. From the results of speaker identification tests using optimal phoneme class ratios, we verified that the proposed system improves speaker identification performance about 18% compared to a conventional system. We also found that using the proposed phoneme class ratio is still applicable to test processes even if the speaker model has been trained with data having conventional phoneme class ratios.

Future work will be finding an optimal method of dividing classes. In this paper, we used the phoneme class label of TIMIT to simplify the problem. However, phonemes have different characteristics even though they are in the same phoneme class. Therefore, we need to examine optimal classification methods for phonemes for better performance. In addition, the accuracy of such classification methods needs to be considered in practical applications.

References and links

[1]S. Furui, "Fifty years of progress in speech and speaker recognition," J. Acoust. Soc. Am. **116**, 2497–2498 (2004).
[2]B. S. Atal, "Text-independent speaker recognition," J. Acoust. Soc. Am. **52**, 181 (1972).
[3]D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. Speech Audio Process. **3**, 72–83 (1995).
[4]R. Auckenthaler, E. Parris, and M. Carey, "Improving a gmm speaker verification system by phonetic weighting," in IEEE International Conference on Acoustics, Speech, and Signal Processing (1999), Vol. **1**, pp. 313–316.
[5]J. Eatock and J. Mason, "A quantitative assessment of the relative speaker discriminating properties of phonemes," in IEEE International Conference on Acoustics, Speech, and Signal Processing (1994), Vol. **i**, pp. 133–136.
[6]D. Gutman and Y. Bistritz, "Speaker verification using phoneme-adapted Gaussian mixture models," in EUSIPCO-2002 the XI European Signal Processing Conference (2002), Vol. **3**, pp. 85–88.
[7]L. Rodriguez-Linares and C. Garcia-Mateo, "Phonetically trained models for speaker recognition," J. Acoust. Soc. Am. **109**, 385–389 (2001).

[8]T. Eriksson, S. Kim, H.-G. Kang, and C. Lee, "An information-theoretic perspective on feature selection in speaker recognition," IEEE Signal Process. Lett. **12**, 500–503 (2005).
[9]M. K. Omar, J. Navratil, and G. N. Ramaswamy, "Maximum conditional mutual information modeling for speaker verification," in EUROSPEECH (2005).
[10]J. A. Nelder and R. Mead, "A simplex method for function minimization," Comput. J. **7**, 308–313 (1965).
[11]J. S. Garofalo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "The DARPA TIMIT acoustic-phonetic continuous speech corpus CDROM," Linguistic Data Consortium (1993).

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.

Temporal weighting in loudness of broadband and narrowband signals (L)

Jan Rennies[a] and Jesko L. Verhey
AG Neuroakustik, Institut für Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 26 August 2008; revised 6 July 2009; accepted 7 July 2009)

Temporal weights used by listeners when judging the overall loudness of a stimulus were measured for a 1-s-long noise centered around 2 kHz, whose level was randomly perturbed every 100 ms. The bandwidth was either 6400 Hz (broadband condition) or 400 Hz (narrowband condition). The first 100 ms contributed significantly more than later segments to overall loudness perception in the broadband condition. The effect was significantly reduced in the narrowband condition which is in line with the hypothesis that a greater spectral loudness summation at stimulus onset might be the mechanism behind the onset accentuation.
 [DOI: 10.1121/1.3192348]

PACS number(s): 43.66.Cb, 43.66.Fe, 43.66.Ba, 43.66.Mk [BCM] Pages: 951–954

I. INTRODUCTION

Zwicker (1977) found for non-stationary sounds that the peak values of the loudness determine overall loudness. This is a special case of the percentile loudness, i.e., the loudness which is reached or exceeded by a certain percentage of loudness values over time (e.g., Stemplinger and Fastl, 1997; Fastl and Zwicker, 2007). Some studies use different parameters to describe different aspects of loudness. For example, Glasberg and Moore (2002) used peak values of loudness to describe temporal integration and a mean value of loudness to predict the overall loudness of modulated signals. All the above concepts are based on the assumption that the temporal positions of the loud portions of the sound are not important and no accentuation of the beginning or the end is made. This assumption was recently challenged by studies on temporal weights in loudness judgement by Ellermeier and colleagues. They investigated the importance of different temporal segments for global loudness judgments based on stimuli with only small, random, level fluctuations (Ellermeier and Schrödl, 2000; Pedersen and Ellermeier, 2008). For broadband-noise stimuli of 1 s duration, they found that the beginning (and to a lesser extent the end) of their stimuli were of particular importance for the overall loudness. The present study investigates whether the dominance of stimulus onset for global loudness judgments, as measured by Ellermeier and Schrödl (2000) and Pedersen and Ellermeier (2008), is affected by the bandwidth of the stimulus. Stimuli similar to the ones used by Ellermeier and Schrödl (2000) were used to confirm that the onset plays a dominant role in determining the loudness of broadband stimuli. Conditional-on-single-stimulus (COSS) analysis (Berg, 1989) was applied to estimate weights assigned to temporal segments of the stimuli. Subsequently, the same group of subjects was tested using narrowband noise.

II. METHODS

The stimuli in the broadband condition of the experiment were similar to those used by Ellermeier and Schrödl (2000). They used white noise of 1 s duration and varied the level every 100 ms. The levels of each of the ten segments were randomly chosen from a normal distribution with a mean value of either 61 ("signal") or 60 dB sound pressure level ("noise"). The standard deviation was 2 dB in both cases. The same parameters of level distributions, segments, and duration were used for the two condition of the present study. In the broadband condition, the bandwidth and geometric center frequency of the stimulus were set to 6400 and 2000 Hz, respectively. In the narrowband condition, a stimulus with the same center frequency and a bandwidth of 400 Hz was tested. For the smaller bandwidth of the noise, slow intrinsic envelope fluctuations could have disturbed the random levels assigned to the individual temporal segments. In order to reduce the influence of the intrinsic fluctuations, double-iterated low-noise was used which was generated on the basis of a method proposed by Kohlrausch *et al.* (1997).[1]

[a]Author to whom correspondence should be addressed. Electronic mail: jan.rennies@uni-oldenburg.de. Present address: Fraunhofer Institute for Digital Media Technology, Oldenburg, Germany.

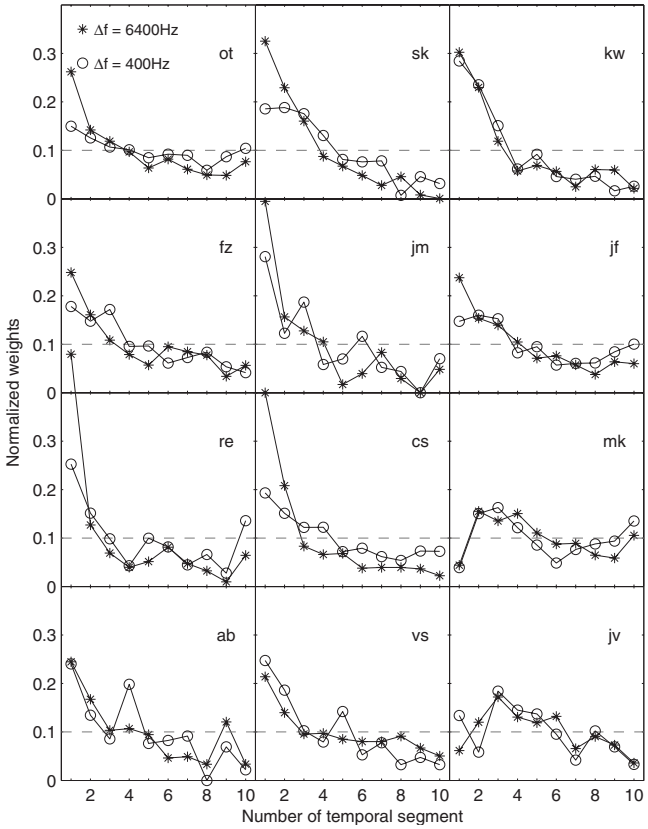


FIG. 1. Individual data for all 12 subjects. Normalized weights are shown for the ten temporal segments. Asterisks and circles represent weights for bandwidths of 6400 and 400 Hz, respectively. Dashed lines indicate equal weights for all segments.

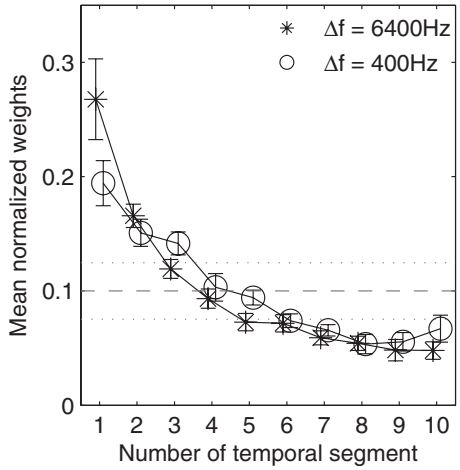


FIG. 2. Mean normalized weights across all subjects for the ten temporal segments. Asterisks and circles represent weights for 6400 and 400 Hz bandwidths, respectively. Error bars represent ±1 standard error. The dashed line indicates equal weights for all segments. Dotted lines represent 99% confidence limits for equal weights obtained from Monte-Carlo simulations.

Each level step in the stimulus was gated with a 2.5 ms raised-cosine ramp, as were stimulus onset and offset.

All stimuli were generated digitally using MATLAB at a sampling rate of 44.1 kHz. A personal computer controlled stimuli generation and presentation. The stimuli were D/A converted (RME ADI-8 PRO), amplified (Tucker-Davis HB7), and presented diotically to the subjects via Sennheiser HD650 headphones.

Twelve normally hearing subjects participated in both parts of the experiment. All had hearing thresholds ≤15 dB hearing level at standard audiometric frequencies between 125 and 8000 Hz. The subjects were between 22 and 26 years of age and were paid for their participation.

A two-alternative forced-choice procedure was used. In each trial, the subjects heard two sounds, signal and noise, in random order, separated by 500 ms of silence. Their task was to indicate which of the two sounded louder by pressing the corresponding button on the keyboard. The intervals were highlighted on the screen during stimulus presentation. New noise samples and envelopes were generated for each trial; signal and noise were generated independently from each other. No feedback was provided. In the broadband condition, each subject made 3000 comparisons between signal and noise for stimuli with a bandwidth of 6400 Hz, which were divided into blocks of 100 comparisons. Several such blocks were presented in one session. Three to four sessions were needed for each subject to complete the measurements. The same procedure was used in the narrowband condition with 400 Hz wide stimuli. Altogether, approximately 10 h of measurement time were needed for each subject.

The COSS analysis introduced by Berg (1989) was applied to estimate weights for each temporal segment and for each subject. The rationale behind this concept is that listeners base their decision on a weighted sum of several observations, in this case the levels in the individual segments of the stimuli. For the two-interval paradigm used in the present study, the subject will perceive interval 1 to contain the louder stimulus if

$$\sum_{i=1}^{10} w_i x_{1i} > \sum_{i=1}^{10} w_i x_{2i}, \tag{1}$$

where w_i is the weight given to the ith segment and x_{1i} and x_{2i} indicate the levels of the ith segment in the first and second intervals, respectively (for details, see Berg, 1989). Confidence intervals for the null hypothesis are estimated by means of Monte Carlo simulations (for a motivation see Berg, 1989).[2]

III. RESULTS

Figure 1 shows individual data for stimulus bandwidths of 6400 Hz (asterisks) and 400 Hz (circles). Normalized weights ($\Sigma_{i=1}^{10} w_i = 1$) are shown as a function of temporal segment for each subject. For comparison, dashed lines indicate equal weights for all segments. For broadband noise, all but two subjects (*jv*, *mk*), had the highest weight at the first temporal segment. For later temporal segments, weights decreased. The difference between the highest and lowest weights varied substantially across subjects, from 0.11 (*mk*) to 0.47 (*re*). Some subjects tended to have higher weights for the last segment than for segments in the temporal center.

The two subjects with a lower weight for the first segment for the broadband noise showed a similar result for the narrowband noise. All other subjects showed a higher weighting of the onset for both narrowband and broadband

noise. Three of the subjects showed a weight for the first segment which was similar for the narrowband and the broadband conditions. For 8 of the 12 subjects, the magnitude of the first weight was smaller for the narrowband condition than for the broadband stimulus. The differences between the highest and lowest weights for each subject were smaller for the narrow than for the larger bandwidth, ranging from 0.09 (*ot*) to 0.28 (*jm*). In general, weights were similar for the two bandwidths except for the first temporal segment.

The main effects were also reflected in the mean data over all subjects, as shown in Fig. 2. Error bars indicate ± 1 standard error of the mean. The highest weights were assigned to the first temporal segment for both bandwidths, and this effect was more pronounced for the bandwidth of 6400 Hz. Monte-Carlo simulations showed that, for both bandwidths, the mean weights and standard error ranges of the first segment were clearly outside the 99% confidence limits, such that the hypothesis of uniform weights was rejected for both bandwidths. A paired, two-sided t-test indicated that the weighting of the first temporal segment was lower for the narrowband noise than for the broadband noise $[t(11)=2.75,\ p<0.02]$.

IV. DISCUSSION

An onset accentuation for broadband signals similar to the one found in the present study was observed by Ellermeier and Schrödl (2000) and Pedersen and Ellermeier (2008).[3] The temporal weight of the first temporal segment in the present study (0.23) was slightly higher than found by Ellermeier and Schrödl (2000) (0.15). A possible source of this deviation is the feedback they provided to the subjects. Pedersen and Ellermeier (2008) showed that trial-by-trial feedback indicating whether the subjects had indicated the signal interval as the louder one modified the weights toward a more uniform curve; i.e., the first segment was weighted less than in the no-feedback condition. The mean weight for the first temporal segment in the condition without feedback obtained by Pedersen and Ellermeier (2008) was about 0.25, which is in quantitative agreement with the result for the broadband stimuli in the present study. The average increase in the weights for the last temporal segment found by Ellermeier and Schrödl (2000) and Pedersen and Ellermeier (2008), which resulted in a bowl-shaped distribution of weights over time, was not observed in the mean data of the present study, although some subjects of the present study showed a similar tendency. Considering the inter-subject variability, the different sets of subjects most likely contributed to the differences between the studies. Altogether, the results of the present study for broadband stimuli are in good agreement with data from previous studies.

No underlying mechanism for the accentuation of the stimulus onset was proposed by Ellermeier and Schrödl (2000). Plank (2005) discussed the possibility of a loss of certain items in memory as an explanation. If a limited capacity of memory is assumed, then listeners could tend to mainly store information from the first temporal segment and consequently base their judgment on the first segments. However, Plank (2005) argued that this could not explain data from discrimination tasks of her study, in which subjects had to detect specific temporal profiles. Pedersen and Ellermeier (2008) suggested a "multiple look" strategy, in which listeners can weigh looks differently depending on their task. So far, however, the mechanism underlying the primacy effect remains unclear.

The results for the narrowband condition in the present study indicate that the bandwidth of the stimuli plays an important role. Decreasing the bandwidth to 400 Hz significantly reduced the weight for the first segment, while weights were similar for later segments. Under the assumption that a larger weight corresponds to an increased loudness, this is in line with recent studies showing that spectral loudness summation depends on stimulus duration (Verhey and Kollmeier, 2002; Fruhmann *et al.*, 2003; Anweiler and Verhey, 2006; Verhey and Uhlemann, 2008).

They showed that spectral loudness summation (quantified as the level difference between equally loud narrowband and broadband stimuli) depends on the duration of the signal. While there was some variability in the degree of loudness summation, these studies found that the level difference between narrowband noises and equally loud broadband noises was about 4–9 dB larger for short stimuli (typically 10 ms) than for long stimuli (typically 1 s).[4] Verhey and Kollmeier (2002) also found a significantly greater spectral loudness summation for 100 ms stimuli (the duration of one segment in the present study) than for 1000 ms stimuli.

Thus, it is possible that the higher weight assigned to the first temporal segment in the present study is due to an increased spectral loudness summation at the beginning of the stimuli. In the light of this hypothesis, the significantly higher weight for the first segment of the narrowband stimulus can be explained by assuming that, even for the small bandwidth, more than one critical band is excited and that the spectral loudness summation is still different for the first segment than for the later segments. In line with this hypothesis, Verhey and Kollmeier (2002) obtained a 1 dB greater spectral loudness summation between the bandwidths 200 and 400 Hz for 100 ms signals than for 1000 ms signals. Thus, a model simulating the duration effect in spectral loudness summation (currently not implemented in loudness models, see, e.g., Verhey and Uhlemann, 2008) may account for the increased temporal weight of the first segment found in the present study without changing the stage determining overall loudness (e.g., the peak value).

The dependence of spectral loudness summation on duration is not observed for every listener. For example, the level difference between equally loud narrowband and broadband noise was more than 3 dB larger for short than for long stimuli at the two largest bandwidths for 7 out of 12 test subjects in Verhey and Uhlemann (2008). For the remaining five subjects, the effect was reduced or absent. A similar individual variation in the results was found in the present study for the temporal weights: for the first segment, 8 of the 12 subjects showed higher weights for broadband noise than for narrowband noise. Four subjects did not show this bandwidth effect. In the light of this similarity of the individual variations in the data of the present study and the data on duration effects in spectral loudness summation, the absence

of a higher weight for the first segment for broadband noise than for narrowband noise in some subjects might be explained by an absent duration dependence of spectral loudness summation in these subjects.

An alternative explanation is that only part of the primacy effect is due to a greater spectral loudness summation at stimulus onset and the rest is due to higher-order processes that emphasize the onset of the signal irrespective of its spectrum. Further studies are necessary to explore the dependence of weights at stimulus onset on overall stimulus duration and number of temporal segments.

ACKNOWLEDGMENTS

We thank Bruce Berg for his support related to the COSS analysis. This work was partly supported by the Deutsche Forschungsgemeinschaft (SFB/TRR 31).

[1]Note that the processing of the auditory periphery changes the phase and amplitude characteristics of the noise, i.e., low-noise noise may no longer be low-noise noise at the output of the auditory filters (Kohlrausch *et al.*, 1997). However, especially for the broadband stimulus, the overall level in each segment is not dominated by the fluctuations of the noise, which are random in each auditory channel. Thus, these fluctuations after auditory filtering should not influence the pattern of temporal weights assigned to segments of the stimuli.

[2]The underlying simple model for the simulations assumes the same weight for each segment, such that the decision rule becomes as follows: respond "interval 1 is louder" if $\Sigma_{i=1}^{10} x_{1i} > \Sigma_{i=1}^{10} x_{2i}$). Fifty simulations of 3000 trials were made and estimates of the weights were calculated for each simulation. The standard deviation of the estimated weight of an arbitrary temporal position was used to obtain the 99% confidence interval.

[3]While COSS analysis was used in the present study and by Ellermeier and Schrödl (2000), Pedersen and Ellermeier (2008) applied logistic regression to estimate weights in a similar task on loudness perception. However, Plank (2005) showed that the two methods yielded very similar results and, thus, results are comparable between the studies.

[4]Some of the early studies (Port, 1963; Zwicker, 1965) found the same spectral loudness summation for long and short signals when compared at the same reference loudness for the two durations. For high loudness values, a similar trend is also observed in the data from the loudness scaling experiment in Anweiler and Verhey (2006). However, for the present study, it is more appropriate to compare spectral loudness summation for different durations at the same reference level (as done in the more recent studies) since the average level of each segment is equal to the level of the whole signal. For a comparison at the same reference level, all recent studies show a greater spectral loudness summation for short signals than for long signals.

Anweiler, A., and Verhey, J. (**2006**). "Spectral loudness summation for short and long signals as a function of level," J. Acoust. Soc. Am. **119**, 2919–2928.

Berg, B. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Ellermeier, W., and Schrödl, S. (**2000**). "Temporal weights in loudness summation," in *Fechner Day 2000, Proceedings of the 16th Annual Meeting of the International Society for Psychophysics*, edited by C. Bonnet (Université Louis Pasteur, Strasbourg), pp. 169–173.

Fastl, H., and Zwicker, E. (**2007**). *Psychoacoustics: Facts and Models*, 3rd ed. (Springer, Berlin).

Fruhmann, M., Chalupper, J., and Fastl, H. (**2003**). "Zum einfluss von innenohrschwerhörigkeit auf die lautheitssummation (Influence of cochlear hearing impairment on spectral loudness summation)," in *DAGA 2003—Fortschritte der Akustik, Proceedings of the 29th Annual Meeting of the Deutsche Gesellschaft für Akustik e.V.*, pp. 253–254.

Glasberg, B. R., and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds," J. Audio Eng. Soc. **50**, 331–341.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence of the role of envelope fluctuations," Acust. Acta Acust. **83**, 659–669.

Pedersen, B., and Ellermeier, W. (**2008**). "Temporal weights in level discrimination of time-varying sounds," J. Acoust. Soc. Am. **123**, 963–972.

Plank, T. (**2005**). "Auditive Unterscheidung von zeitlichen Lautheitsprofilen (Auditory discrimination of temporal loudness profiles)," Ph.D. thesis, Universität Regensburg, Regensburg, Germany.

Port, E. (**1963**). "Über die Lautstärke kurzer Schallimpulse (Loudness of short sound pulses)," Acustica **13**, 212–223.

Stemplinger, I., and Fastl, H. (**1997**). "Accuracy of loudness percentile versus measurement time," in Proceedings of Inter-Noise '97, Vol. **III**, pp. 1347–1350.

Verhey, J., and Kollmeier, B. (**2002**). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am. **111**, 1349–1358.

Verhey, J., and Uhlemann, M. (**2008**). "Spectral loudness summation for sequences of short noise bursts," J. Acoust. Soc. Am. **123**, 925–934.

Zwicker, E. (**1965**). "Temporal effects in simultaneous masking and loudness," J. Acoust. Soc. Am. **38**, 132–141.

Zwicker, E. (**1977**). "Procedure for calculating loudness of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.

Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners (L)[a)]

Aniket A. Saoji
Auditory Research and Development, Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California 91342

Leonid Litvak
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California 91342

Anthony J. Spahr
Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287

David A. Eddins[b)]
Department of Otolaryngology, University of Rochester, 2365 South Clinton Avenue, Suite 200, Rochester, New York 14618 and International Center for Hearing and Speech Research, Rochester Institute of Technology, Rochester, New York 14623

(Received 7 January 2009; revised 12 June 2009; accepted 19 June 2009)

Speech understanding by cochlear implant listeners may be limited by their ability to perceive complex spectral envelopes. Here, spectral envelope perception was characterized by spectral modulation transfer functions in which modulation detection thresholds became poorer with increasing spectral modulation frequency (SMF). Thresholds at low SMFs, less likely to be influenced by spectral resolution, were correlated with vowel and consonant identifications [Litvak, L. M. *et al.* (2008). J. Acoust. Soc. Am. **122**, 982–991] for the same listeners; while thresholds at higher SMFs, more likely to be affected by spectral resolution, were not. Results indicate that the perception of broadly spaced spectral features is important for speech perception.
 [DOI: 10.1121/1.3179670]

PACS number(s): 43.71.Ky, 43.71.Es, 43.66.Ts, 43.66.Fe [RYL] Pages: 955–958

I. INTRODUCTION

Cochlear implant (CI) listeners have markedly reduced spectral resolution relative to normal-hearing listeners. The resulting spectral smoothing has greater impact on the perception of fine spectral details characteristic of high spectral modulation frequencies than broadly spaced spectral details characteristic of lower spectral modulation frequencies (Henry and Turner, 2003; Henry *et al.*, 2005; Litvak *et al.*, 2007; Won *et al.*, 2007). If phoneme recognition in CI listeners is limited by poorer than normal spectral resolution, then one would predict stronger correlations between phoneme recognition and the perception of higher than lower spectral modulation frequencies. Here the authors evaluate this hypothesis by measuring spectral modulation detection thresholds over a wide range of modulation frequencies from low to high, resulting in individual spectral modulation transfer functions (SMTFs) for each of 25 CI users.

The relation between spectral resolution and phoneme perception was highlighted by Henry *et al.* (2005) who used a modified and abbreviated version of the "ripple phase-reversal task" described by Supin *et al.* (1994) as an index of spectral resolution. The spectral modulation depth was fixed at 30 dB and the highest spectral modulation rate at which listeners can discriminate a 180° shift in spectral modulation phase was determined. Average phase-reversal threshold for normal-hearing (NH) listeners was 4.8 cycles/octave as compared to a threshold of 0.6 cycles/octave for CI listeners (Henry *et al.*, 2005). This index of spectral resolution was correlated with vowel (r=0.80) and consonant (r=0.81) identifications in their CI listeners. Similarly, Litvak *et al.* (2007) reported detection thresholds in CI listeners for relatively low spectral modulation frequencies (0.25 and 0.5 cycles/octave). The ability to detect the broad spectral peaks characteristic of low spectral modulation frequencies differed substantially across CI listeners and was correlated with vowel (r=−0.84) and consonant (r=−0.82) identifications. Assuming that the ability to detect low spectral modulation frequencies was affected by poorer than normal spectral resolution in CI users, Litvak *et al.* (2007) used a vocoder simulation to model this relationship in NH listeners, effectively smearing the spectrum. The simulation results were similar to those obtained for CI listeners.

In theory, the modulation detection thresholds of Litvak *et al.* (2007) and the single phase-reversal threshold reported by Henry *et al.* (2005) represent three points on the SMTF. In listeners with acoustic hearing, 0.25 and 8.0 cycle/octave represent points near the extremes of the SMTF (e.g., Bernstein and Green, 1987; Summers and Leek, 1994; Eddins and Bero, 2007). Detection thresholds at the upper extreme are severely limited by spectral resolution, while detection thresholds at the lower extreme are not influenced by the limits of spectral resolution (Eddins and Bero, 2007; Saoji and Eddins 2007; Summers and Leek, 1994) and instead may

[a)]Portions of these data were presented at the 2005 Conference on Implantable Auditory Prosthesis at Asilomar, Pacific Grove, CA.
[b)]Author to whom correspondence should be addressed. Electronic mail: david_eddins@urmc.rochester.edu

reflect the ability to analyze broad spectral features, analogous to the across-channel intensity comparisons of the profile analysis task (e.g., Green, 1988).

In light of the data from acoustic hearing, the data of Litvak *et al.* (2007) raise two interesting possibilities. First, if the primary limitation for spectral modulation perception is spectral resolution, then individual differences among listeners should result in SMTFs that differ by a single spectral resolution factor and phoneme perception should be correlated equally as well with detection at each modulation frequency. Second, the SMTF may be limited by spectral resolution, particularly at high modulation frequencies, and other factor(s) at low modulation frequencies, leading to complex individual differences in the shape of the SMTF. In this case, correlations between phoneme perception should be different for low and high spectral modulation frequencies, depending on the factors contributing to individual differences.

The present study explores these possibilities and expands on the report by Litvak *et al.* (2007) by reporting spectral modulation detection thresholds at higher spectral modulation frequencies (1.0 and 2.0 cycles/octave) in combination with thresholds reported by Litvak *et al.* (2007) at low spectral modulation frequencies (0.25 and 0.5 cycles/octave) for the same CI listeners. Auditory abilities that may underlie individual differences in the form of the SMTF among CI listeners are considered in terms of changes in fitting parameters associated with the SMTFs and the potential contributions of loudness cues.

II. METHODS

CI subjects included 25 post-lingually deafened adults (38–65 years old) using either the Advanced Bionics CII or HiRes/90 000 CI. Table I includes relevant details about each CI subject. These subjects also participated in the study reported by Litvak *et al.* (2007).

Digital stimuli were routed via soundcard (M-Audio, Audiophile 2496) to a body worn platinum series processor through the Advanced Bionics direct-connect system. Soundcard output was attenuated such that the electric input to the DirectConnect® system was equivalent to a 60 dB sound pressure level (SPL) acoustic input to the speech processor microphone. Listeners were seated in a quiet room and adjusted the PSP volume to a "comfortable" level at the beginning of the first testing session.

Stimulus generation (MATLAB®) involved applying the desired spectral shape to a noise carrier in the frequency domain. Inverse Fourier transform on the complex buffer pair resulted in the desired noise band (for details, see Litvak *et al.*, 2007). Spectral modulation detection thresholds were obtained using a cued, three-interval, two-alternative, forced-choice procedure with feedback combined with a three-down, one-up adaptive tracking rule that estimates 79.4 percent correct detection. For each 60-trial adaptive track, modulation depth was varied in a step size of 2 dB for the first three reversals and 0.5 dB for the remainder of the track. Final thresholds were computed as the average modulation depth corresponding to the last even number of reversals, excluding the first three. Thresholds were measured at 0.25, 0.5, 1.0, and 2.0 cycles/octave in random order in brief sessions spanning 1–2 days.

TABLE I. The table shows listener-specific information characterizing the 25 CI listeners. The implant was either a CII or HiRes 90 000 with a HiRes sound coding strategy. *S*—subject number, *A*—age of the listener (years), *E*—experience with their implant (months), and EL array—electrode array. [The demographic data in Table I are accurate, and reflect correction of several errors from the paper of Litvak *et al.* (2007).]

S	*A*	*E*	EL array
1	53	6	1J
2	52	30	HiFocus I
3	60	28	HiFocus I
4	68	21	HiFocus I
5	70	12	Helix
6	61	18	HiFocus I
7	52	14	1J
8	37	13	Helix
9	62	55	HiFocus II
10	52	33	HiFocus I
11	49	32	HiFocus I
12	46	19	Helix
13	39	27	1J
14	35	22	1J
15	63	25	HiFocus I
16	54	16	HiFocus I
17	32	37	HiFocus I
18	54	47	HiFocus II
19	62	48	HiFocus I
20	74	29	1J
21	38	20	1J
22	49	14	Helix
23	42	20	1J
24	56	12	Helix
25	35	42	HiFocus I

Unmodulated and modulated stimuli were scaled to equivalent levels; however, the CI processing algorithm and/or differences in growth of loudness across electrodes could potentially introduce loudness differences between the two stimuli. A subset of nine CI listeners completed a loudness matching task to evaluate the availability of an overall loudness cue. A two-interval, two-alternative, forced-choice method was used. The modulated signal was presented in one interval and the unmodulated standard in the other interval in random order. The modulation depth was 5 dB above the listener's highest spectral modulation detection threshold for a given modulation frequency (i.e., 5 dB re maximum). The level of the unmodulated standard was randomly chosen from 55 to 65 dB SPL in 1 dB steps. The 11 comparisons were repeated 10 times for a total of 110 trials per subject at 0.25, 0.5, and 1 cycles/octave.

III. RESULTS AND DISCUSSION

The SMTFs obtained for the 25 CI listeners are shown in Fig. 1 (left panel). For four CI listeners, modulation detection thresholds could not be obtained at 2.0 cycles/octave given a maximum possible modulation depth of 60 dB. The overall pattern of SMTFs obtained for CI listeners indicates that modulation detection thresholds are generally lowest at

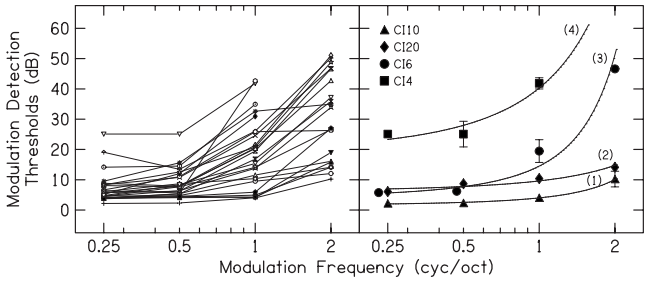


FIG. 1. Left panel shows the SMTFs for the 25 CI listeners. The ordinate represents spectral modulation detection thresholds (dB) and the abscissa shows the modulation frequency (cycles/octave). Right panel shows the four illustrated SMTF patterns (1, 2, 3, and 4 for subjects CI10, C20, CI6, and CI4, respectively) and corresponding exponential fits (solid lines). For clarity, thresholds for some modulation frequencies have been shifted to the left/right about the nominal modulation frequency.

0.25 cycles/octave and increase progressively with increasing modulation frequency. Analysis of the data at 0.25, 0.5, and 1.0 cycles/octave from all 25 CI listeners using Friedman's test indicated a significant effect of modulation frequency (chi value=34.62, $p<0.0001$). Likewise, analysis of the data from the 21 listeners for whom thresholds could be measured at all modulation frequencies using Friedman's test indicated a significant effect of modulation frequency (chi =55.32, $p<0.0001$). Tukey–Cramer test revealed a statistically significant ($p<0.05$) difference among the thresholds at each of the four modulation frequencies.

As reported by Litvak *et al.* (2007) (their Figs. 1 and 3), there was considerable variability in the vowel and consonant identification scores across the same 25 CI listeners, with performance ranging from 16% to 96% for vowel identification and from 28% to 92% for consonant identification. To determine whether or not a systematic relationship exists between spectral modulation detection thresholds (from 0.25 to 2.0 cycles/octave) and vowel or consonant identification, single- and multiple-factor linear regression analyses were undertaken. Following Bonferroni correction applied for $\alpha=0.05$, all single factor correlations were statistically significant. Correlations were higher for consonant than vowel identification and were highest for 0.5 cycles/octave, where $r=-0.75$ for vowel identification and $r=-0.82$ for consonant identification. Thus, spectral modulation detection thresholds at 0.5 cycles/octave account for approximately 57% of the variance associated with vowel identification and 67% of the variance associated with consonant identification in CI listeners. To better establish the relation between combinations modulation detection thresholds and vowel and consonant identification, multi-factor forward stepwise linear regression analyses were performed. A Bonferroni correction was applied for $\alpha=0.05$. For vowel identification, threshold at 0.25 cycles/octave was the strongest predictor ($r=-0.77$). For consonant identification, the modulation frequency of 0.5 cycles/octave was sole predictor present in the final regression equation, with $r=-0.69$.

This is interesting considering that differences in spectral resolution affect spectral modulation detection at higher (e.g., 1 and 2 cycles/octave) spectral modulation frequencies more than at lower (e.g., 0.25 and 0.5 cycles/octave) spectral modulation frequencies. If differences in spectral resolution among CI users are correlated with their phoneme perception (Henry *et al.*, 2005; Litvak *et al.*, 2007), then one would expect the modulation detection thresholds measured at 1 and/or 2 cycles/octave to be the strongest predictor of phoneme recognition scores in CI users. It is possible that CI users rely more on the broadly spaced spectral maxima and minima in the spectral envelope for phoneme perception.

To better understand the variability in the SMTFs and the relationship with phoneme perception in CI listeners, threshold functions were fitted (in the least-squares sense) with an exponential function with two parameters [the rate of the exponent (b) and the y intercept (A)] as shown in

$$f(x) = Ae^{bx}, \tag{1}$$

where $f(x)$ represents modulation depth at threshold corresponding the modulation frequency x. Changes in the exponent b result in contraction or dilation along the x axis, altering the steepness of the function. Changes in the scalar A result in contraction or dilation along the y axis, changing the y intercept. Exponential fits to the SMTFs for the 25 CI listeners accounted for an average of 93% of the threshold variance. Four illustrative patterns that capture the variability among CI listeners are shown in Fig. 1 (right panel). Filled symbols represent modulation detection thresholds and solid lines represent fitted functions. Using pattern 1 (shown for subject CI10; triangles) as a reference, the shape, y-intercept value, and exponent value can be compared to the other three patterns. Relative to pattern 1, pattern 2 (e.g., subject CI 20; diamonds) reveals a modest increase in both the y-intercept and exponent, consistent with a complex change in the SMTF rather than a change in a single spectral resolution factor. Pattern 3 (e.g., subject CI6; circles) shows a modest change in the y-intercept and a large change in the exponent, reflecting a much larger change in spectral resolution than for pattern 2. This clearly indicates that two CI users with different spectral resolutions can have similar modulation detection thresholds at low modulation frequencies. Likewise, two CI listeners with same spectral resolution can have different abilities to detect low spectral modulation frequencies (e.g., subjects C6 and C20, not shown). Pattern 4 (e.g., subject CI4; squares) includes large changes in both the exponent and y-intercept parameters.

Correlations between the exponent b and speech perception measures obtained for the 25 listeners were $r=0.12$ for vowel identification and $r=0.06$ for consonant identification, neither of which was significant ($p>0.05$). Similar analyses for the scalar A resulted in correlations of $r=-0.75$ for vowel identification and $r=-0.80$ for consonant identification, both of which were significant at the 0.01 level. A non-significant ($p>0.05$) correlation of $r=-0.34$ was obtained between the exponent and scalar values, b and A. The correlation between the ability to detect the low modulation frequencies and vowel and consonant identifications may be attributed to the differences in the ability of the CI listeners to compare spectral maxima and minima in the spectral envelope that span a broad audio-frequency range. A general reduction in intensity resolution in CI listeners would also impact their ability to encode and detect spectral modulation independent of spectral modulation frequency. Combined reductions in spec-

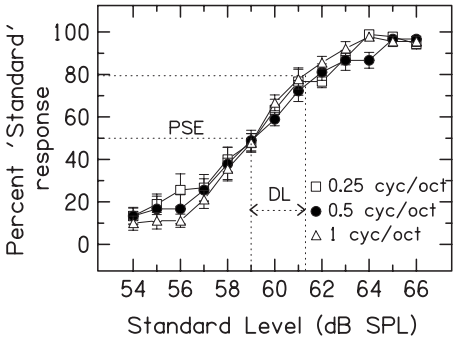


FIG. 2. Performance on the loudness matching task with percent standard responses as a function of the standard level in dB SPL. The three psychometric functions represent the average percent correct across the nine CI listeners for each of three spectral modulation frequencies (0.25, 0.5, and 1.0 cycles/octave). Dotted lines indicate the PSE. Error bars about each symbol represent the standard error of the mean and some are obscured by the symbol.

tral and intensity resolutions could lead to an overall elevation and an increased slope in the SMTF (e.g., Fig. 1, pattern 4).

The relation between spectral modulation detection and speech perception reported here is similar to the results of CI simulations reported by Litvak *et al.* (2007) in which performance of individual CI listeners was matched by manipulating the parameters of their vocoder simulation. While performance on all perceptual tasks was dependent on the vocoder parameters for NH listeners, NH and CI listeners matched for their spectral modulation detection thresholds had similar vowel and consonant identification scores. This close correspondence between performance for NH listeners using their CI simulation and actual CI listeners, matched according to sensitivity to spectral modulation, is consistent with the notion that, in the absence of fine spectral details, CI listeners may be relying on their ability to identify broad spectral patterns for speech identification.

To better determine whether or not an overall loudness cue may have been used by the CI listeners, 9 (filled symbols, Fig. 1, upper panel) of the original 25 CI listeners performed the loudness comparison task. Psychometric functions averaged across the nine listeners are shown in Fig. 2 with percent "standard" responses as a function of the standard level. For each individual psychometric function at each modulation frequency, the best fitting logistic function (in the least-squares sense) was computed. The point corresponding to 50% correct averaged across all (27 functions) was 59.0 dB SPL. One possibility is that this 1.0 dB shift, relative to the nominal level of 60 dB SPL, is due to a small but consistent loudness cue associated with spectral modulation. The 50% point, or the point of subjective equality (PSE), is shown by the dashed lines in the lower left corner. Since the spectral modulation detection thresholds corresponded to 79.4% correct, this point will be taken as the "threshold" for "louder" judgments, and the difference between the 50% and 79% points on the psychometric function may be taken as the difference limen (DL) for the attribute under study (e.g., loudness or intensity). The loudness DL corresponds to 2.3 dB (61.3−59.0=2.3 dB) averaged across conditions and modulation frequency. Given that the horizontal shift in the psychometric function is less than half the value of the loudness DL, it is unlikely that that shift reflects a usable loudness cue. The PSE, threshold, and resulting DL values are shown graphically in Fig. 2 by the dotted lines. Furthermore, derived slopes and intercepts were not significantly correlated with modulation detection thresholds or vowel or consonant identification ($p>0.05$).

IV. CONCLUSIONS

Individual differences in spectral modulation perception and phoneme recognition (e.g., Litvak *et al.*, 2007; Henry *et al.*, 2005) do not depend on a single spectral resolution factor. Instead, individual SMTFs differ widely in their shape, reflecting multiple underlying factors. Significant correlations were observed among phoneme recognition scores and spectral modulation detection thresholds at low but not high spectral modulation frequencies. Reduced spectral resolution in CI listeners effectively eliminates perception of the fine spectral details of speech that correspond to high spectral modulation frequencies. As a result, they rely on the broad spectral envelope features characteristic of low modulation frequencies, and detection thresholds for those low modulation frequencies account for a significant proportion of the variance in phoneme recognition among CI listeners. These results also indicate that a one-point measure of spectral modulation detection may adequately characterize spectral resolution but is insufficient to describe individual differences in spectral envelope perception among CI listeners.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gulam Emadi for technical assistance, the anonymous reviewers, and the associate editor. The work was supported by Advanced Bionics LLC.

Bernstein, L. R., and Green, D. M. (**1987**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.
Eddins, D. A., and Bero, E. M. (**2007**). "Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region," J. Acoust. Soc. Am. **121**, 363–372.
Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, Oxford).
Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.
Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," J. Acoust. Soc. Am. **118**, 1111–1121.
Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners," J. Acoust. Soc. Am. **122**, 982–991.
Saoji, A. A., and Eddins, D. A. (**2007**). "Spectral modulation masking patterns reveal tuning to spectral envelope frequency," J. Acoust. Soc. Am. **122**, 1004–1013.
Summers, V., and Leek, M. R. (**1994**). "The internal representation of spectral contrast in hearing-impaired listeners," J. Acoust. Soc. Am. **95**, 3518–3528.
Supin, A., Popov, V. V., Milekhina, O. N., and Tarakasanov, M. B. (**1994**). "Frequency resolving power measured by rippled noise," Hear. Res. **78**, 31–40.
Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2007**). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," J. Assoc. Res. Otolaryngol. **8**, 384–392.

47-channel burst-mode recording hydrophone system enabling measurements of the dynamic echolocation behavior of free-swimming dolphins (L)

Josefin Starkhammar[a)]
Electrical Measurements, Faculty of Engineering LTH, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden

Mats Amundin
Kolmarden Wildlife Park, SE-618 92 Kolmarden, Sweden and Department for Physics, Chemistry and Biology (IFM), Linkoping University, SE-581 83 Linkoping, Sweden

Johan Nilsson and Tomas Jansson
Electrical Measurements, Faculty of Engineering LTH, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden

Stan A. Kuczaj
Department of Psychology, The University of Southern Mississippi, P.O. Box 5025, Hattiesburg, Mississippi 39406-5025

Monica Almqvist and Hans W. Persson
Electrical Measurements, Faculty of Engineering LTH, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden

(Received 9 February 2009; revised 30 June 2009; accepted 30 June 2009)

Detailed echolocation behavior studies on free-swimming dolphins require a measurement system that incorporates multiple hydrophones (often >16). However, the high data flow rate of previous systems has limited their usefulness since only minute long recordings have been manageable. To address this problem, this report describes a 47-channel burst-mode recording hydrophone system that enables highly resolved full beamwidth measurements on multiple free-swimming dolphins during prolonged recording periods. The system facilitates a wide range of biosonar studies since it eliminates the need to restrict the movement of animals in order to study the fine details of their sonar beams. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3184536]

PACS number(s): 43.80.Ev, 43.80.Ka, 43.60.Qv [WWA] Pages: 959–962

I. INTRODUCTION

The echolocation of dolphins and other odontocetes has been extensively studied, a primary focus being the beam axis (Nachtigall and Moore, 1988; Thomas and Kastelein, 1990; Au, 1993; Villadsgaard *et al.*, 2007; Kyhn *et al.*, 2009). Recording scenarios that focus on the beam axis typically provide accurate and effective measurements. However, since these situations often require that the echolocating animal be kept stationary, it is likely that the full dynamics of the sonar beam has not yet been described. In addition, these static test conditions are by definition impossible to use in other important contexts, such as in studies of the spontaneous use of echolocation by free-swimming dolphins, object investigation behavior in groups of dolphins, and calf mimicry of their mother's echolocation clicks. Although detailed sonar studies have been conducted with free-swimming dolphins (Sigurdson, 1996; Martin *et al.*, 2005, among others), these studies used relatively few hydrophones and consequently had limited sonar beam coverage. As a result, much is known about the beam axis, and little is known about the rest of the beam.

Recording dolphin sonar in dynamic test conditions requires a system able to deal with varying measurement parameters, including the animal's relative distance to the receivers, the number of animals present, the orientation of the beam relative to the receivers, and the required time for the animal to respond to an echolocation task. Consequently, a measurement system capable of long recording periods, large beamwidth coverage, and high spatial and temporal resolutions is needed in order to localize the beam axis and to measure the rest of the beam with high accuracy from free-swimming dolphins.

Multi-channel sonar recording systems have previously been reported by Miller and Tyack (1998), Ball and Buck (2005), Starkhammar *et al.* (2007), Amundin *et al.* (2008), and Moore *et al.* (2008), among others. The most extensive system developed thus far was created by Moore *et al.* (2008), who employed 24 hydrophones in an array. It was designed for high spatial resolution across the array area since it was used in a study of beamwidth control in a bottlenose dolphin (*Tursiops truncatus*) and so required measure-

[a)]Author to whom correspondence should be addressed. Electronic mail: josefin.starkhammar@elmat.lth.se

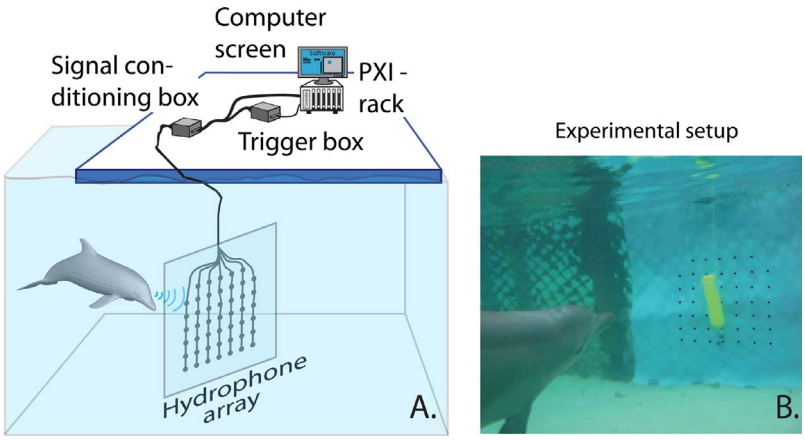


FIG. 1. (Color online) (A) A schematic drawing of the system setup in the field test. (B) An underwater photograph of the experimental setup.

ments of small spatial alterations of the amplitude distribution across the whole beam cross-section. The system recorded data during 5 s intervals with a sample rate of 312.5 kS/s, resulting in a data flow rate of 16 Mbytes/s. Although this system worked well for its purpose, the extremely high data flow rate makes it unsuitable for measuring echolocation behavior in free-swimming dolphins. Using this system, a single minute of recordings would result in a 1 Gbyte large file.

A measurement system optimized for dynamic test conditions requires large beamwidth coverage and higher spatial and temporal resolutions. The system must also be able to record for longer time periods in order to be useful in the field. This requires an increase in the number of array elements, the physical size of the array, and, preferably, also the sample rate. In addition, the problems associated with extremely high data flow rates must be solved.

This report describes a system with a measurement approach optimized for studies of dolphin sonar under dynamic test conditions. The system uses a larger number of hydrophones (47 channels), allows an increased sample rate (1 MHz), and acquires data with lower data flow rates than previously reported multi-hydrophone systems. This facilitates full waveform recordings of the echolocation activity of dolphins during prolonged time periods and comprehensive beamwidth coverage, provided that the dolphins are within reasonable distance from the screen. In addition, the approach enables real-time analysis and real-time visualization of data during recordings. This measurement approach and recording system enables researchers to investigate dolphin sonar use in a wider range of contexts than has previously been made possible. In the following sections, the authors describe this system and provide examples of its potential uses with free-swimming dolphins.

II. MATERIALS AND METHODS

A new 47-channel dolphin echolocation measurement system was developed and tested with a group of 19 Atlantic bottlenose dolphins, housed together in a large open sea pen at Roatán Institute of Marine Science, Roatán, Honduras. All dolphins were allowed to swim freely and to explore at will objects suspended in front of or behind the recording hydrophone array. The size of array was 0.75×0.75 m^2. Figure 1(A) shows a schematic representation of the experimental setup, and Fig. 1(B) shows a photograph of the setup during one trial.

Typically, multi-hydrophone arrays used in biosonar applications produce a considerable amount of data due to the relatively high sample rate required to reconstruct the full waveform accurately in post-analysis [approximately ten times the maximum frequency for accurate visualization (Buchla and McLachlan, 1992)]. The long recording times required in test conditions with spontaneously echolocating dolphins using a 47-element array system would result in unmanageably large data files after only a few seconds, using the data acquisition approaches in previous systems. Therefore, an alternative approach to continuous sampling of all parallel channels was needed for the 47-element hydrophone array system.

In order to facilitate longer recordings and to keep the data flow rate manageable with a high sample rate, the system was designed to be triggered by one echolocation click at a time and to not sample data during the silent periods that occur between clicks in click trains. Figure 2 describes the basic data acquisition method. Basically, the system only acquires a small pre-set number of samples, containing only the actual click, when a hydrophone output exceeds the chosen trig level. This is referred to here as burst-mode sampling (as opposed to continuous sampling). The time stamps (T1 and T2 in Fig. 2) correspond to the start time of each sample burst and are stored in association with each click.

Successful burst-mode sampling required data acquisition hardware capable of extremely fast re-triggering of measurements even after the particularly short inter-click intervals (<1 ms) that may occur in click train "buzzes" (see

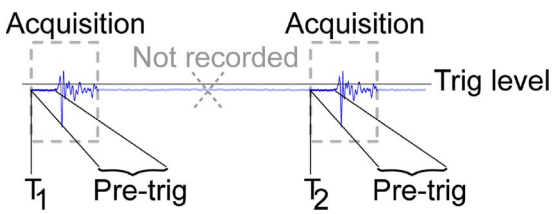


FIG. 2. (Color online) Basic data acquisition method.

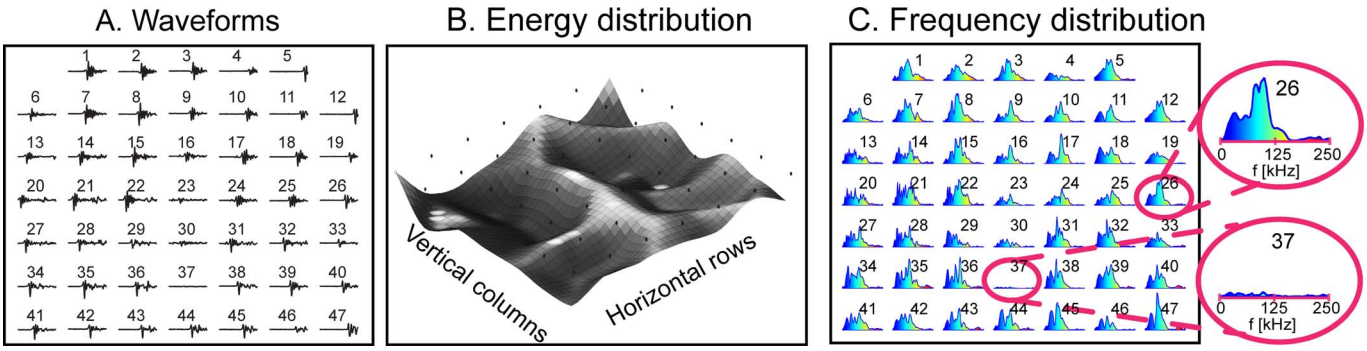


FIG. 3. (Color online) Visualization of one echolocation click acquired at the 47 hydrophone positions in the experimental setup. This click can be visualized by plotting (A) the waveforms, (B) the relative energy distribution within the beam, or (C) the color coded frequency distribution within the beam. High click energy in (B) is illustrated by letting it push the interpolated 3D surface downward, away from the echolocating dolphin.

Herzing, 1996; Herzing and dos Santos, 2004) or when multiple dolphins echolocate concurrently. The total time required to finish one burst-mode acquisition, re-trig, and start a new one is referred to here as the system's rearm time.

In order to capture the echolocation data across the entire array, the system was designed to simultaneously trigger all channels, regardless of which hydrophone was hit first by the sonar beam. Signals were acquired from 6×8 parallel and synchronized channels using six PXI-digitizer cards (NI PXI-5105, National Instruments, USA), each with eight simultaneously sampling analog-to-digital-channels and 12 bit voltage resolution. One of the digitizer channels was used as a trig-channel. The 47 remaining channels were wired to the 47 individually pre-amplified hydrophones, all amplified with either 35 or 50 dB, depending on the measurement situation. The sum of all the 47 pre-amplified hydrophone signals was wired to the pre-designated trig-channel using a separate signal summarizing hardware circuitry.

The software was optimized to ensure fast rearm time and real-time visualization and analysis of data, aspects that were typically the most time-consuming as well as important determinants of the overall system performance. This improvement was accomplished with software created in LABVIEW 8.6™ (National Instruments, USA) enabling dual-core operation of the CPU.

III. RESULTS AND DISCUSSION

The echolocation measurements of the free-swimming dolphins in the group of 19 individuals demonstrated that the system is capable of measuring the full waveforms of the spontaneous sonar activity in the group. Measurements were obtained with all 47 simultaneously sampling hydrophones at a sample rate of 1 MHz during measurement sessions of various lengths (often >15 min). Each acquisition in these tests was set to record during a time window of 150 μs around each click event with a pre-trig time period of 40 μs (see Fig. 2). The total duration of the measurement sessions was determined by the tourist activity at the facility. Sessions were never aborted due to system failure.

Figure 3(A) shows the corresponding waveforms of one single click acquired simultaneously by all 47 hydrophones. Each position of the numbered elements corresponds spatially to the hydrophones shown in Fig. 1. The level of spatial resolution and comprehensive beamwidth coverage provides new information concerning the entire cross-section of the beam during one single click. As an example, Fig. 3(B) shows the relative energy distribution of the cross-section in the beam. The energy in the click is coded as an indentation of the interpolated three dimensional (3D) surface where high energies "push" the surface downward, away from the echolocating dolphin.

A suspended scuba tank provided a way to further demonstrate the functionality of the system by shadowing the hydrophones in the center of the screen when a dolphin echolocated toward the tank from small bearing angles (i.e., close to the perpendicular to the screen). This shadowing effect is clearly seen as a ridge in the middle of the beam energy plot in Fig. 3(B).

The presented level of spatial resolution and comprehensive beamwidth coverage give an unprecedented detailed measure of the spectral content within the cross-section of the beam [Fig. 3(C)]. The minimized data flow rates make it possible to view the spectral content even in real-time. The system also allows researchers to study entire echolocation scan sequences in detail by processing all successive clicks and then re-playing them at variable frame rates. The resolution of the measurements enables detailed re-plays of the propagation of every single click across the array, further facilitating quantified detailed studies of the dynamic variations in the echolocation behavior of dolphins during prolonged periods of time.

Benchmark tests of the system performance showed that the low data flow makes possible recordings during 20 min of constantly echolocating animals with inter-click intervals of 20 ms before the data file size reaches 1 Gbyte and becomes unreasonably large for commonly used post-processing tools (such as MATLAB®, The MathWorks™ Inc., USA). This is a considerable improvement compared to previously published systems. The low data flow rate of the present system (0.83 Mbyte/s under the conditions in the benchmark tests) is even more advantageous in more realistic measurement scenarios, where free-swimming dolphins echolocate spontaneously and when minute long silent periods in the recordings are likely to occur.

In conclusion, the presented 47-element hydrophone system enables recordings with improved spatial and temporal resolutions of the cross-section of the echolocation beam

of free-swimming dolphins. Moreover, the system makes possible extended recording periods due to the minimized data flow rate. These features facilitate the reconstruction, visualization, and re-play of significant aspects of the clicks during extended echolocation sequences. The system's ability to process information from free-swimming dolphins in groups opens the door to a completely new range of studies, which will help us to better understand the functions of dolphin sonar since it eliminates the need to restrict the movement of animals in order to study the fine details in their sonar beams.

Amundin, M., Starkhammar, J., Evander, M., Almqvist, M., Lindstrom, K., and Persson, H. W. (**2008**). "An echolocation visualization and interface system for dolphin research," J. Acoust. Soc. Am. **123**, 1188–1194.

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

Ball, K. R., and Buck, J. R. (**2005**). "A beamforming video recorder for integrated observations of dolphin behavior and vocalizations," J. Acoust. Soc. Am. **117**, 1005–1008.

Buchla, D., and McLachlan, W. (**1992**). *Applied Electronic Instrumentation and Measurement* (Macmillan, New York), pp. 384–388.

Herzing, D. L. (**1996**). "Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and Bottlenose dolphins, *Tursiops truncatus*," Aquat. Mamm. **22**, 61–79.

Herzing, D. L., and dos Santos, M. E. (**2004**). "Functional aspects of echolocation in dolphins," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 386–393.

Kyhn, L. A., Tougaard, J., Jensen, J. F., Wahlberg, M., Stone, G., Yoshinaga, A., Beedholm, K., and Madsen, P. T. (**2009**). "Feeding at a high pitch: Source parameters of narrow band, high-frequency clicks from echolocating off-shore hourglass dolphins and coastal Hector's dolphins," J. Acoust. Soc. Am. **125**, 1783–1791.

Martin, S. W., Phillips, M., Bauer, E. J., Moore, P. W., and Houser, D. S. (**2005**). "Instrumenting free-swimming dolphins echolocating in open water," J. Acoust. Soc. Am. **117**, 2301–2307.

Miller, P. J., and Tyack, P. L. (**1998**). "A small towed beamforming array to identify vocalizing *resident* killer whales (Orcinus orca) concurrent with focal behavioral observations," Deep-Sea Res., Part II **45**, 1389–1405.

Moore, P. W., Dankiewicz, L. A., and Houser, D. S. (**2008**). "Beamwidth control and angular target detection in an echolocating bottlenose dolphin (Tursiops truncatus)," J. Acoust. Soc. Am. **124**, 3324–3332.

Nachtigall, P. E., and Moore, P. W. B. (**1988**). *Animal Sonar, Process and Performance* (Plenum, New York).

Sigurdson, J. E. (**1996**). "Open-water echolocation of bottom objects by dolphins (*Tursiops Truncatus*)," J. Acoust. Soc. Am. **100**, 2610.

Starkhammar, J., Amundin, M., Olsén, H., Ahlmqvist, M., Lindstrom, K., and Persson, H. W. (**2007**). "Acoustic touch screen for dolphins," in *Proceedings of Fourth International Conference of Bio-Acoustics*, Institute of Acoustics, Loughborough University, edited by S. Dimble, P. Dobbins, J. Flint, E. Harland, and P. Lepper, pp. 55–60.

Thomas, J. A., and Kastelein, R. (**1990**). *Sensory Abilities of Cetaceans* (Plenum, New York).

Villadsgaard, A., Wahlberg, M., and Tougaard, J. (**2007**). "Echolocation signals of wild harbour porpoises, *Phocoena phocoena*," J. Exp. Biol. **210**, 56–64.

Quantification of material nonlinearity in relation to microdamage density using nonlinear reverberation spectroscopy: Experimental and theoretical study

K. Van Den Abeele[a)]
Interdisciplinary Research Center, K.U. Leuven Campus Kortrijk, Etienne Sabbelaan 53, B-8500 Kortrijk, Belgium

P. Y. Le Bas
Interdisciplinary Research Center, K.U. Leuven Campus Kortrijk, Etienne Sabbelaan 53, B-8500 Kortrijk, Belgium and Los Alamos National Laboratory, EES-17, MS-D443, Los Alamos, New Mexico 87544

B. Van Damme
Interdisciplinary Research Center, K.U. Leuven Campus Kortrijk, Etienne Sabbelaan 53, B-8500 Kortrijk, Belgium

Tomasz Katkowski
Interdisciplinary Research Center, K.U. Leuven Campus Kortrijk, Etienne Sabbelaan 53, B-8500 Kortrijk, Belgium and Institute of Experimental Physics, University of Gdansk, Ulica Wita Stwosza 57, 80-952 Gdańsk, Poland

(Received 28 October 2008; revised 29 June 2009; accepted 30 June 2009)

High amplitude vibrations induce amplitude dependence of the characteristic resonance parameters (i.e., resonance frequency and damping factor) in materials with microscopic damage features as a result of the nonlinear constitutive relation at the damage location. This paper displays and quantifies results of the nonlinear resonance technique, both in time (signal reverberation) and in frequency (sweep) domains, as a function of sample crack density. The reverberation spectroscopy technique is applied to carbon fiber reinforced plastic (CFRP) composites exposed to increasing thermal loading. Considerable gain in sensitivity and consistent interpretation of the results for nonlinear signatures in comparison with the linear characteristics are obtained. The amount of induced damage is quantified by analyzing light optical microscopy images of several cross-sections of the CFRP samples using histogram equalization and grayscale thresholding. The obtained measure of crack density is compared to the global macroscopic nonlinearity of the sample and explicitly confirms that the increase in nonlinearity is linked to an increased network of cracks. A change from 1% to 3% in crack density corresponds to a tenfold increase in the signature of nonlinearity. Numerical simulations based on a uniform distribution of a hysteretic nonlinear constitutive relation within the sample support the results.
 [DOI: 10.1121/1.3184583]

PACS number(s): 43.25.Dc, 43.25.Ba, 43.25.Gf, 43.25.Zx [ROC] Pages: 963–972

I. INTRODUCTION

Safety and reliability of large and small scale engineering structures are of crucial importance. In aeronautics, for instance, the performance and behavior characteristics of airframe structures can be adversely affected by structural degradation resulting from sustained use within normal flight envelopes, as well as from exposure to severe environmental conditions or from damage due to unexpected impacts. The timely and accurate detection, characterization, and monitoring of the development of structural defects over time (e.g., cracking, corrosion, delamination, and material degradation) are a major concern in the operational environment. If the authors are to improve the accuracy of structural integrity predictions, they must minimize the uncertainty associated with critical parameters for early degradation, incipient damage, and progressive failure modes in components. Hence, huge efforts are devoted to the development of enhanced, reliable, and integrated measurement systems and protocols for identifying microcracks in structural engineering components. As part of this effort, researchers all over the world are currently developing and validating innovative microdamage inspection system based on various nondestructive testing methods within the class of nonlinear elastic wave spectroscopy (NEWS).[1,2]

NEWS techniques primarily deal with the investigation of the amplitude dependence of material parameters such as wave speed, attenuation, and spectral content. The degree to which these material properties depend on the applied dynamic amplitude can be quantified by various nonlinearity parameters. Several NEWS techniques have been developed to probe for the existence of damage (e.g., delaminations, microcracks, and weak adhesive bonds) by investigating the

[a)]Author to whom correspondence should be addressed. Electronic mail: koen.vandenabeele@kuleuven-kortrijk.be

generation of harmonics and intermodulation of frequency components,[3–11] the amplitude-dependent shift in resonance frequencies,[11–15] the nonlinear contribution to attenuation properties,[15] slow dynamic effects,[15–18] and phase modulation.[19] Laboratory tests performed on a wide variety of materials subjected to different microdamage mechanisms of mechanical, chemical, and thermal origins have shown that the sensitivity of such nonlinear methods to the detection of microscale features is far greater than that obtained with linear acoustical methods.[6,8,13,20–23]

In this paper, the authors restrict themselves to a new variation of the nonlinear resonance technique. Single mode nonlinear resonant ultrasound spectroscopy (SIMONRUS)[12,13] is a well-known frequency domain method analyzing resonance sweeps at increasing excitation amplitude; the authors here introduce its time domain variant, which analyzes the instantaneous changes in the resonance characteristics during the reverberation of an object after being excited near resonance. This technique, called nonlinear reverberation spectroscopy (NRS), is an improved version of an earlier reported time domain resonance technique (time domain SIMONRUS) that has been used to analyze mechanical fatigue in titanium and concrete.[22,24] The main advantage over SIMONRUS is that the discrete frequency sweep for several amplitudes of excitation is replaced in NRS by a simple time signal recording at a single excitation level. Hence, NRS is significantly faster and requires fewer acquisitions. After presenting the general ideas behind NRS in Sec. II, the authors apply the new version to composite laminates [carbon fiber reinforced plastics (CFRPs)] with various degrees of thermal loading, simulating the initiation of global microdamage as the result of extreme environmental conditions (Sec. III). These results will be put side by side with a measure of the crack density of the samples via image analysis in the third part. To the authors' knowledge, this is the first real experimental quantification of nonlinearity in terms of crack density. In the Conclusion (Sec. VI), the authors present a theoretical model of a resonating flexural beam including nonlinear mechanical properties to explain the observed nonlinear behavior.

II. NRS

Nonlinear resonance spectroscopy techniques investigate the resonance behavior of objects under amplitude-dependent response. Generally, a single resonance mode of the object with associated resonance frequency is selected. In SIMONRUS,[12,13] the object is subjected to a frequency sweep around this resonance frequency at constant excitation amplitude. The true resonance characteristics, frequency and damping factor (or quality factor), are then analyzed from fits of the resulting frequency response amplitude (resonance curve). This is repeated for increasing levels of excitation amplitudes. By plotting the resonance characteristics at each level versus the maximum response amplitude at the same level, one can analyze the amplitude dependence or nonlinearity of the object. Intact materials show no change in the resonance characteristics, whereas damaged materials generally show a decrease in the resonance frequency with amplitude (nonlinear softening) and an increase in the damping factor ($1/Q$, with Q as the quality factor) due to nonlinear attenuation.[15,22,24]

NRS is the time domain analogy of SIMONRUS. In NRS, a sample is excited at constant excitation amplitude and constant frequency for a certain period of time. The frequency is chosen in the neighborhood of one of the resonance frequencies of the sample. After a number of cycles, sufficient for the sample to reach its steady state response, the continuous wave excitation is stopped, say, at $t=t_0$, and the reverberation response of the sample is measured from t_0 to t_1 and stored for analysis. The reverberation signal is typically a decaying time signal, with large amplitudes near t_0 and smaller amplitudes near t_1. Appropriate synchronization allows averaging of the signal, and a feedback loop can be used to optimize the dynamic range as function of the measurement time. Several sections recorded at decreasing dynamic range are finally selected and matched to create a composed signal with adequate vertical resolution. The decay signal is then analyzed using a successive fitting of an exponentially decaying sine function,

$$A_k e^{-\alpha_k t} \sin(2\pi f_k t + \phi_k), \tag{1}$$

to small time windows (approximately 20 cycles). Here, A_k denotes the amplitude, α_k is the decay parameter, f_k is the frequency, and ϕ_k is the phase of the signal in the kth window. This allows the creation of a parametric plot of the true resonance frequency f_k and of the decay parameter α_k as function of the amplitude A_k, thereby providing information on the occurrence of nonlinearity. If the material is linear, the frequency in different windows of the reverberation signal remains constant. If the material is nonlinear, the frequency in the reverberation signal gradually increases with decreasing amplitude and thus with time, in agreement with the nonlinear softening effect on the modulus due to the presence of nonlinearity.[25,26]

In practice, depending on the sample (size and weight) and the resonance frequency of the mode, both experiments, SIMONRUS and NRS, can be performed in a fully noncontact mode by means of a loudspeaker as exciter and a laser Doppler vibrometer for the response measurement. The schematic setup and a typical NRS response and analysis of the data for one of the samples considered in Sec. III (thermal shocked CFRP) can be found in Fig. 1. The amplitude dependence of resonance frequency and damping are clear markers of the nonlinear material behavior. When the amplitudes are recalculated in terms of strain (see later), a NRS nonlinearity parameter can be deduced from the proportionality relation as the slope of the relative change.

Figure 2 shows the consistency between the two resonance methods. The analyzed resonance frequency from the frequency sweeps shows the same slope (nonlinearity) as the analyzed resonance frequency deduced from the reverberation signal. There is a small offset related to the change in experimental conditions. The advantage of the NRS method is that it requires fewer acquisitions (one time signal at a single excitation level versus a discrete frequency sweep at

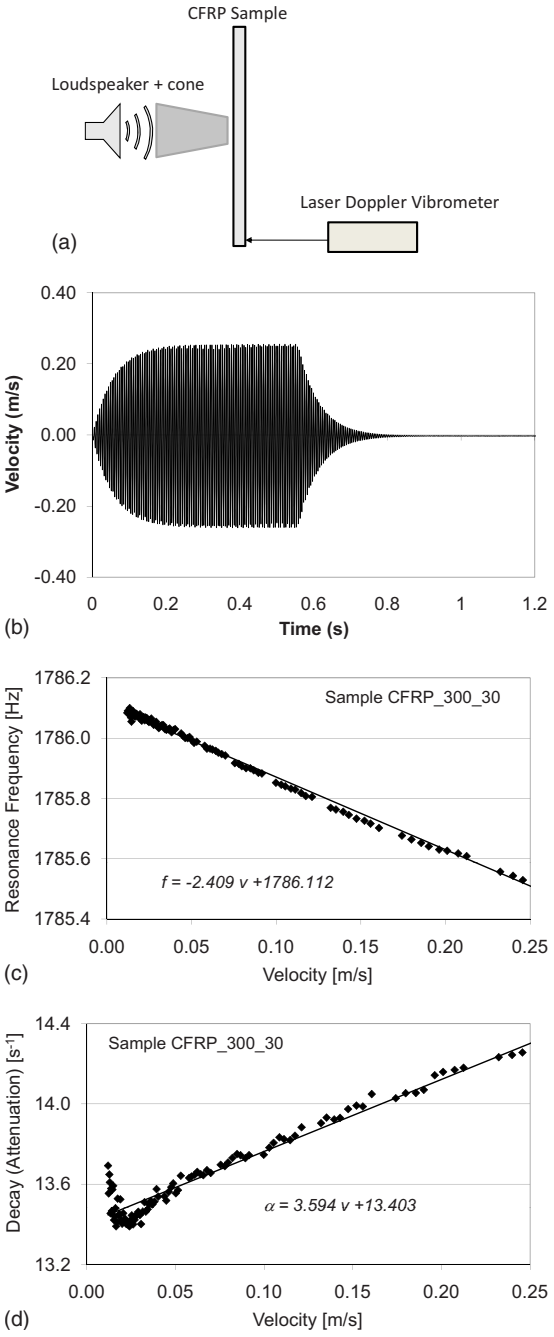


FIG. 1. (Color online) The NRS method and its typical results. (a) NRS experimental setup: The sample is excited by a loudspeaker, and the particle velocity is recorded by means of a laser Doppler vibrometer. (b) Full recorded signal. (c) Analysis of the instantaneous resonance frequency versus particle velocity amplitude for a CFRP sample shocked at 300 °C for 30 min. (d) Analysis of the instantaneous damping characteristic versus particle velocity amplitude for the same sample.

various increasing levels of excitation) and, by such, that it is faster than SIMONRUS even if more post-acquisition data analysis is needed.

The robustness of the method has been tested in several ways. Being a non-contact experiment, the only concern that could affect reproducibility is the string support of the sample. Paying particular attention to put the supporting strings near the nodes, several experiments were repeated after dismounting and remounting without major deviations in the results (errors of a few percent). In addition, even

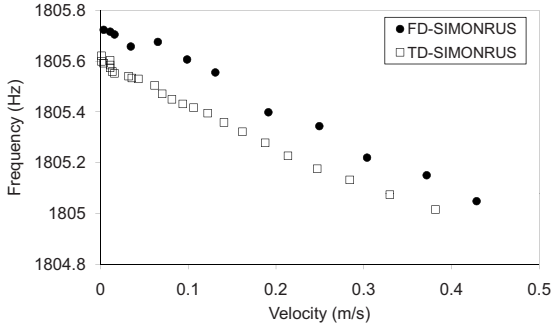


FIG. 2. Illustration of the consistency between the SIMONRUS and NRS results (sample CFRP_300_60).

though experimental conditions may significantly affect the resonance frequency, the slope of the amplitude dependence, which yields the measure of the nonlinearity parameter in the NRS experiment, is independent of the exact resonance frequency value as it merely depends on the relative changes of it with respect to amplitude. These relative changes seem to be less dependent on the experimental conditions than the values of the resonance frequency. On top of this, the authors also verified that the obtained slope in the proportionality relations is independent of the chosen initial excitation frequency and applied voltage. The results of these investigations are illustrated in Fig. 3. In Fig. 3(a), the analyzed response at three different frequencies in the neighborhood of the resonance frequency is illustrated for a fixed excitation amplitude. In Fig. 3(b), the response at a fixed excitation frequency is illustrated for three different excitation amplitudes. The conclusion is that the NRS nonlinearity parameter is independent of the initial excitation frequency (within limits in the order of the full width at half maximum of the resonance curve) and applied voltage (for regimes that do not involve slow dynamics). Together with the high sensitivity of the NRS nonlinearity parameter to damage (see Sec. III), this relative insensitivity to changes in the experimental setup and conditions, in comparison to linear resonance measurements, is of primary and practical advantage for the method.

III. APPLICATION TO THERMALLY LOADED CFRP

In this section, the authors illustrate the potential of the NRS techniques to discern heat damage in CFRPs and to validate its postulated high sensitivity to early damage and micromechanical changes in the medium.

A. CFRP and heat damage

CFRP is commonly used in the aircraft construction industry. It is expected that the next generation of airplanes will consist of more than 60% of composite structures.[27] Even though composite materials hold important advantages over aluminum, CFRP is also prone to various degradation mechanisms. The exposure to heat, for instance, induces chemical and microstructural changes affecting the mechanical behavior of the composite laminate, even at moderate temperatures.[28] Traditional nondestructive quality control techniques are often limited in their capabilities to detect and characterize subtle changes in the material properties associ-

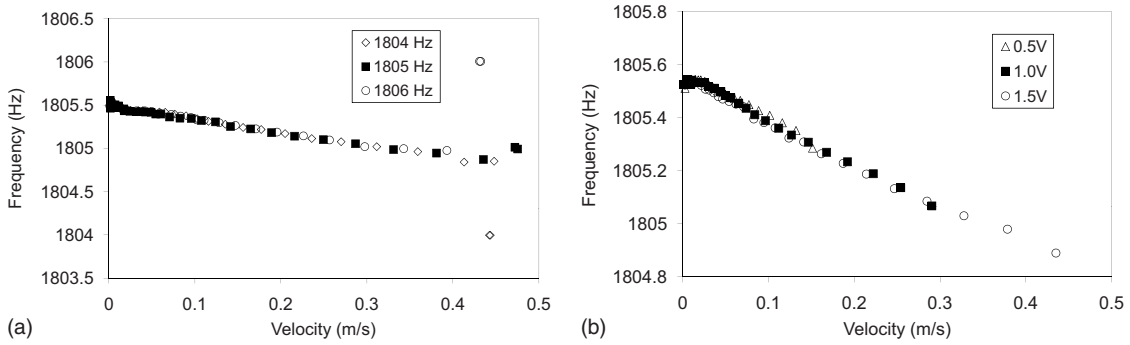


FIG. 3. Verification of the reliability of the NRS results for a single sample (CFRP_300_60) at various excitation frequencies (1804, 1805, and 1806 Hz) for fixed amplitude (1 V) (left) and at various excitation amplitudes (0.5, 1, and 1.5 V) for fixed excitation frequency (1805 Hz) (right).

ated with heat damage. A review of the mechanisms of heat damage in composites and a state-of-the-art of nondestructive evaluation (NDE) techniques currently used to evaluate heat damage is available in Refs. 28 and 29. Studies have shown that thermal degradation is typically matrix dominated since by the time fiber properties such as tensile strength and modulus are affected, all other mechanical integrity is lost. Mechanical metrics such as compressive, shear, and flexural strength and stiffness properties are believed to be the most sensitive properties for use in the early detection of thermal degradation, as opposed to nonmechanical parameters such as thermal and dielectric properties. Most of the work reported in the literature dealing with NDE for heat damage in composites is based on the following five methods: thermal (IR), ultrasonics, acoustic emission, dielectric properties, and radiography. These methods, while being readily available and generally well developed, are limited in their capabilities to detect and characterize the changes in composite material properties associated with heat damage. For instance, the detectability threshold of heat damage (1 h exposure at temperatures 200–300 °C) in unidirectional AS4-8552 CFRP laminates using conventional ultrasonics (immersed transmission C-scan imaging) was found at 290 °C.[30] Nevertheless, the measured value of the interlaminar shear strength for the same type of samples changed from 121 MPa for nonexposed samples to 114 MPa when exposed at 200 °C, to 84 MPa for 285 °C, and to 43 MPa for samples exposed at 300 °C for the duration of 1 h.

Most traditional NDE techniques are capable of detecting physical anomalies such as cracks and delaminations. However, to be effective for thermal degradation, they must be capable of detecting initial heat damage, which occurs at a microscopic scale. Review of the literature from more recent years indicates that a vast number of NDE methods are currently under development and show various degrees of promise for characterizing heat damage in composites. More extensive information on the status of development of several of these NDE methods and their capabilities for detecting heat damage in composite laminates can be found in an extended state-of-the-art review available from NTIAC.[29]

B. The NRS results

The authors examined a set of heat damaged composite laminate samples using the above described NRS technique and quantified their NRS nonlinearity parameters as function of the heating temperature and exposure time. The set of 21 CFRP (AS4/8552 quasi-isotropic lay-up) samples consisted of one reference sample, which was left unexposed, and 20 samples exposed at five different temperatures (240, 250, 260, 270, and 300 °C) for four different durations (15, 30, 45, and 60 min). The samples were cooled under ambient conditions and tested at room temperature. The nominal size of the samples was 120 mm (L) × 20 mm (W) × 4 mm (T). It is expected that thermal damage is induced in a more or less uniform manner over the sample volume.

The resonance mode under consideration in this study is the fundamental flexural mode of a beam, which has a stress concentration in the middle of the sample and displacement nodes at a distance of $0.224L$ from both edges, with L as the length of the sample (120 mm).[31,32] In the experimental setup, the sample is supported by two nylon wires at the node lines and is excited at a pure tone by a loudspeaker (diameter of 32 mm, the sound being concentrated by a converging cone of 180 mm length and 20 mm exit diameter) centered in the middle of the sample. The response is measured by a laser vibrometer (Polytec OFV303, decoder VD02) near one of the edges. All equipment is computer controlled and operated through LABVIEW and GPIB. The acquisition of the signal is realized by a 5 MHz DAQ-card.

In the NRS experiment, the authors excited the sample with a 1000 period burst excitation at a given amplitude and with a frequency close to the fundamental flexural resonance frequency. They then recorded a total of 0.6 s (120 000 points at a sampling rate of 200 kHz) of the reverberation of the sample after the excitation was stopped. Figure 1(b) shows a typical response from the start of the excitation to the steady state and the reverberation. To achieve a high accuracy in the recording of the reverberation signal, the authors implemented a variable dynamic range acquisition procedure based on an automated feedback of the instantaneous amplitude response. In this procedure, the dynamic range is decreased successively. At each range, signals are acquired and averaged ten times. The various signals recorded at decreasing dynamic range are finally matched to create a composed signal with adequate vertical resolution over the entire time axis.

The amplitude, frequency, and damping information contained in the resulting signal are then analyzed by divid-

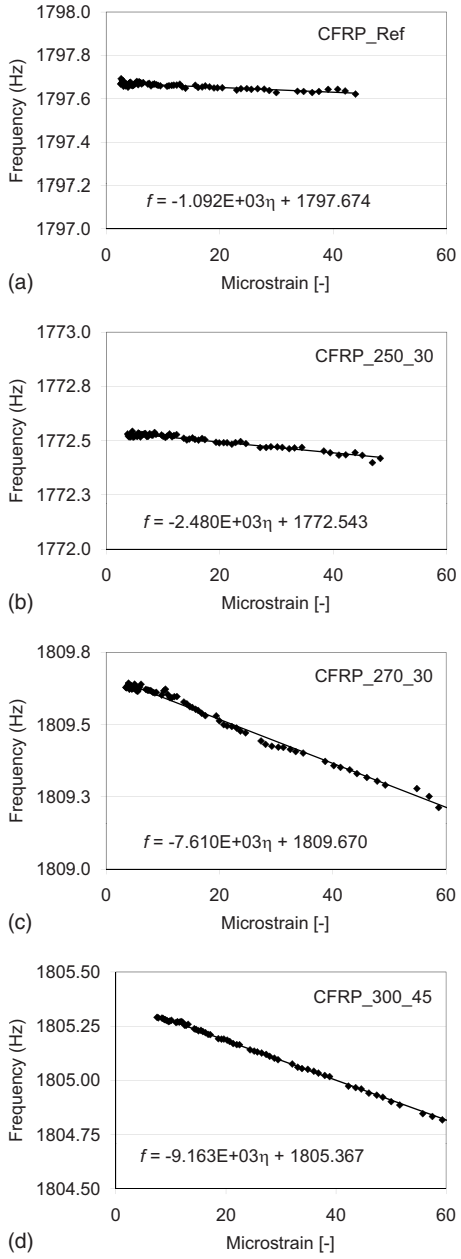


FIG. 4. NRS results showing the analyzed frequency versus microstrain amplitude for the reference sample showing almost no nonlinearity and for three samples at different heating temperatures and exposure times: 250 °C for 30 min, 270 °C for 30 min, and 300 °C for 45 min.

ing the composed signal into several windows (with fixed time duration of 10 ms, which is typically of the order of 20 periodic oscillations) and by fitting the previously described exponentially decaying sine function [Eq. (1)] to the data using a Levenberg–Marquardt algorithm to determine the parameters f_k, α_k, ϕ_k, and A_k, with k referring to the kth time window. This yields the evolution of the frequency (f_k) and damping characteristic (α_k) as function of the amplitude A_k in the decaying signal. Figure 4 shows the results for the instantaneous resonance frequency versus amplitude for the reference sample, for two samples exposed for 30 min at 250 and 270 °C, respectively, and for a sample heated at 300 °C for 45 min. The analyzed data for the reference sample nearly follow a horizontal line, meaning that there is no or minimal dependence of the frequency on the amplitude. The reference sample is thus close to being a linear material. On the other hand, the results for longer exposure and higher temperature show an increased frequency dependence on amplitude, which indicates an increase in the material nonlinearity. Changing the window size for the analysis of the reverberating signal (within limits, of course) did not influence the results.

In order to quantify the degree of nonlinearity, the authors calculated the NRS nonlinearity parameter Γ as the proportionality coefficient between the relative resonance frequency shift and the strain amplitude η,

$$\frac{\Delta f}{f_0} = \Gamma \eta, \tag{2}$$

with f_0 as the linear resonance frequency and $\Delta f = f_0 - f$. The strain amplitude values, η, were calculated from the measured particle velocity amplitude values, v, using the strain-velocity conversion expression for beams,[31,32]

$$\eta \approx 0.219 \frac{T}{f\sqrt{12}} \left(\frac{4.73}{L}\right)^2 v, \tag{3}$$

with $T=4$ mm and $L=120$ mm. It should be noted that because of the global character of the applied NEWS method, Γ only represents a global quantification of the nonlinearity, integrated over the whole sample. It contains no direct information on the localization of the defects. The values for the global NRS nonlinearity parameter Γ obtained in this study range from 0.6 to 10, and its variation as function of temperature and exposure time for all samples is summarized in Fig. 5(a). The authors observed an overall increase with increasing exposure time and heat temperature up to a factor of 10 with respect to the reference value. The obtained values are comparable to values obtained for intact samples of heterogeneous materials such as slate (30),[13] pultruded composites (5–10),[21] concrete (45),[22] and other materials (rocks and metals).[14,15]

A similar behavior can be observed when analyzing the nonlinearity in the damping characteristic,

$$\frac{\Delta \alpha}{\alpha_0} = \Upsilon \eta, \tag{4}$$

with α_0 as the linear time constant (connected to the attenuation). However, the errors in the analysis results are larger (support of the samples is very critical for attenuation), and the fits are not as clean as the ones dealing with the resonance frequency shift, which results in a less pronounced evolution [Fig. 5(b)].

C. Comparison with the linear resonance results and discussion

The NRS analysis also provides the linear resonance signatures such as linear attenuation and linear resonance frequency. Ignoring subtle geometry changes, it is possible to calculate the global stiffness (Young's modulus E_0) for the different samples from the linear resonance frequency values. However, the authors could not observe a systematic change as function of the temperature and exposure time

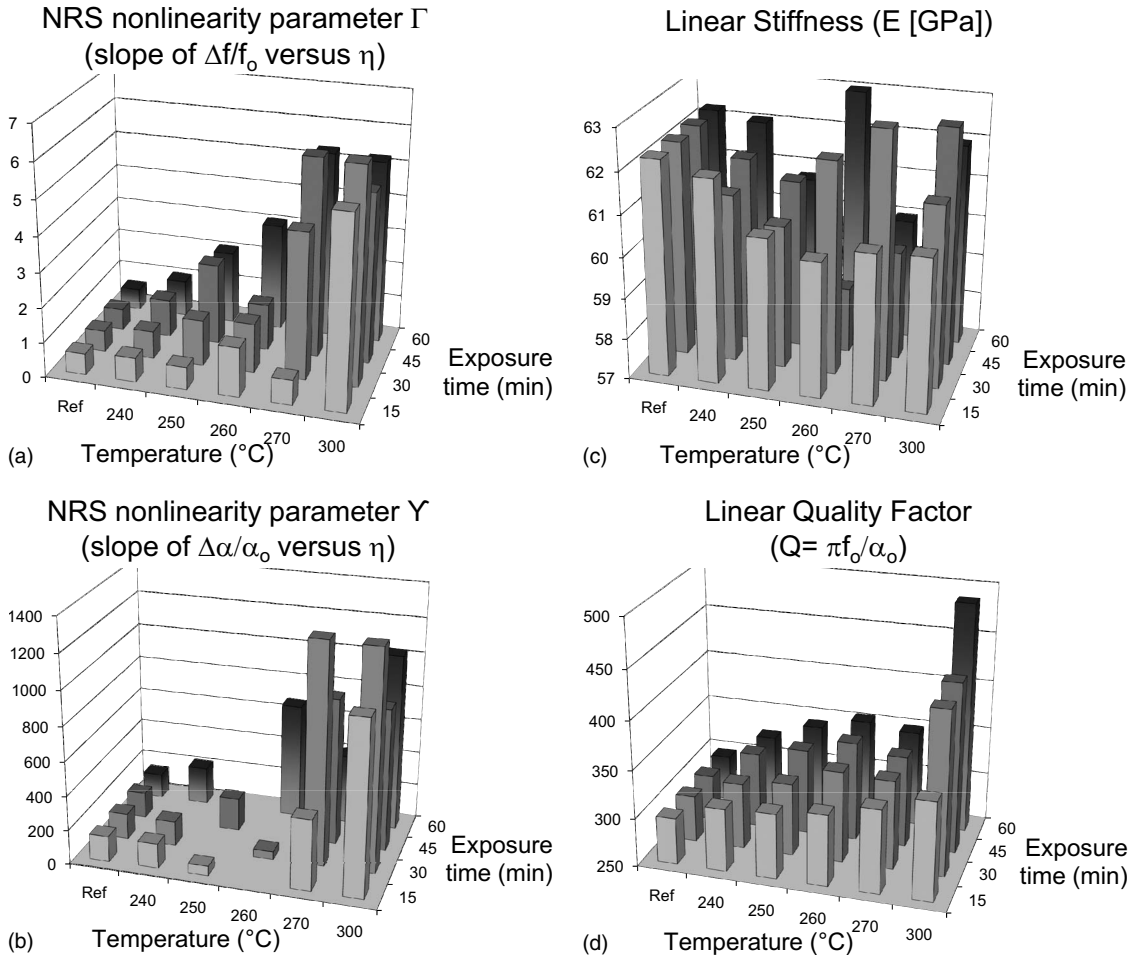


FIG. 5. Summary of the NRS results for all 21 samples as function of heating temperature and exposure time: (a) NRS nonlinearity parameter Γ deduced from the frequency response, (b) NRS nonlinearity parameter Γ deduced from the damping response (some Γ-values with low repeatability are omitted), (c) linear values of the stiffness E, and (d) linear Q-factor (inverse attenuation). The reference point was duplicated for different exposure times to help visualize the trend of the evolution of the parameters with temperature.

[Fig. 5(c)]. For the attenuation, on the other hand, they found that the linear value of the quality factor Q_0 (inverse attenuation $Q_0 = \pi f_0 / \alpha_0$) increases with temperature and exposure time [Fig. 5(d)], meaning that the attenuation (at that frequency) decreases with increasing damage. This is somewhat counterintuitive as one expects attenuation to increase with damage.

Without pretending that the authors are experts in this field, they conjecture that the decrease in the linear attenuation is associated with the chemical alteration in the matrix connected to fluids and fluid expulsion upon thermal loading. The reduction in fluids and the chemical adaptation processes generally lead to a decrease in attenuation, which in this case might dominate the expected increase in attenuation due to the formation of microcracks. In any case, it is obvious from this analysis that the nonlinear parameters derived in the NRS method show a considerable gain in sensitivity and provide a consistent interpretation of the results in contrast with the linear characteristics.

IV. QUANTIFICATION OF THE NRS NONLINEARITY PARAMETER IN RELATION TO THE MICROCRACK DENSITY

It is generally accepted that the dislocation buildup and the presence of cracks in damaged samples result in a macroscopically observed nonlinear behavior: The higher the crack density, the more pronounced the nonlinear signature of the sample will be. To check this idea and to quantify the obtained values of NRS nonlinearity parameter Γ with respect to the microcrack density, the authors sliced five of their samples in the thickness direction and extracted the crack density from each sample at the surface. The samples are first imaged using light optical microscopy (LOM) coupled to a digital camera. Images are acquired with the magnification level set to 2. For each sample, the entire surface of the transversal and longitudinal cuts is captured sequentially by imaging small size rectangles covering the entire surface. The images are then combined to generate a full picture. Two such raw images are shown in Fig. 6.

The process to quantify the crack density is performed on subsets of the entire image, the size of which is determined by the processing speed of the computer used to perform the image treatment. The "crack density" analysis for each subset consists of six standard steps:

1. selection of the zones such that only the areas with fibers oriented out of plane are analyzed,
2. edge detection using Sobel's technique[33] and conversion in grayscale,
3. equalization of the image,[34]

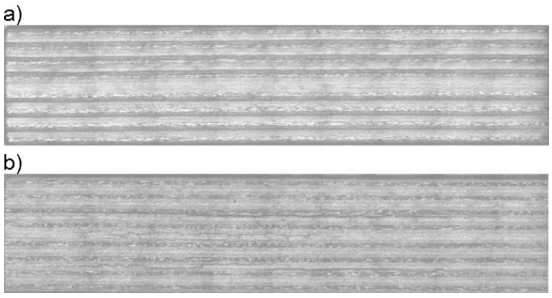


FIG. 6. Original images of CFRP layers (composed of overlapping digital shots using a LOM). (a) Transversal cut for the reference sample. (b) Transversal cut for the sample treated at 270 °C for 60 min.

4. thresholding to get binary images,
5. elimination of isolated points, and
6. determination of the crack density as the ratio of the black pixels to the total number of pixels after the final stage in the image treatment.

When performing a similar treatment for all subsets of the original image, the authors obtained Fig. 7, which compares the original raw image to the binary end result. The figure also contains a more detailed view, obtained with a scanning electron microscope, of a typical thermally induced crack in the region of the tip and in the central region.

The results of this treatment applied to all images leads to Fig. 8. Open diamonds represent the different values of the crack density obtained in several subsets of the images. The spread of the results is mainly due to the small size of the subset area, which is analyzed and illustrates the statistical variation as function of the position along the surface. As one can expect, some areas show almost no cracks, while others exhibit several. The filled circles are the average value for each sample. The authors observed a clear relationship between the NRS nonlinearity parameter Γ and the crack density.

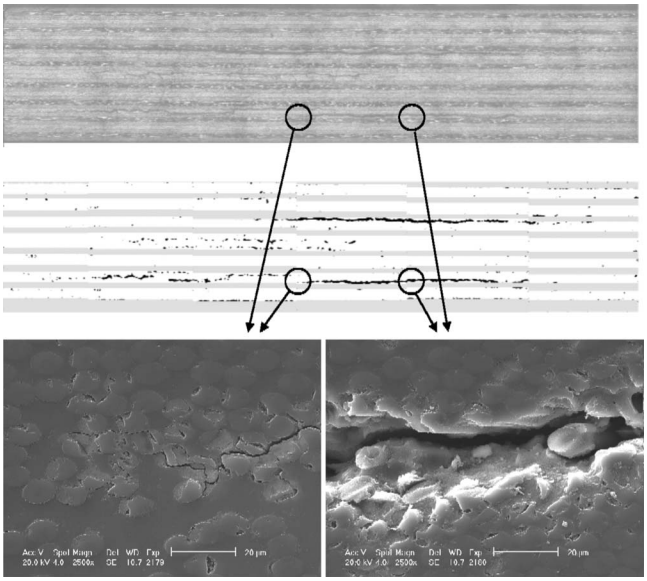


FIG. 7. Comparison between the original raw image and the post-treatment picture for crack detection. Details of the crack near its tip and its center.

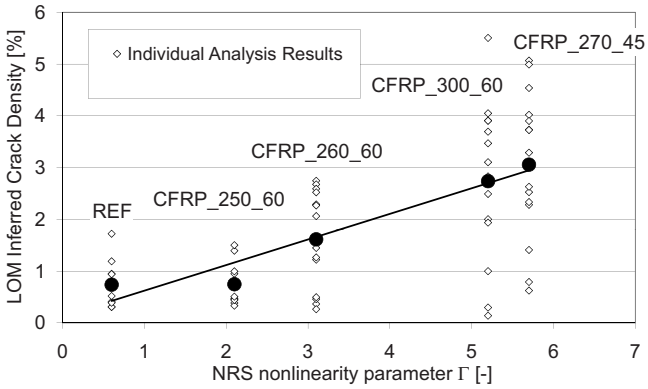


FIG. 8. Crack density versus NRS nonlinearity parameter Γ. Open diamonds are values of the crack density obtained for different parts of the images. Filled circles are the averages for the whole surface for each sample. The line is the linear trend line for the average values. The horizontal error bars for the experimentally obtained quantity of Γ are of a few percent for an individual sample.

(i) The NRS nonlinearity parameter increases with increasing crack density
(ii) The dispersion of the data increases with the NRS nonlinearity parameter. This can be explained by the non-homogeneous repartition of the cracks inside the samples.
(iii) Even though the crack density measurements for the reference sample and the sample treated at 250 °C for 60 min are not significantly different, the authors have observed a vast increase (more than a factor of 2) in the NRS nonlinearity parameter. This could imply that the authors' crack density procedure based on the image treatment is not sensitive enough to identify the very early features (e.g., increase in dislocation nuclei) that are responsible for the increased NRS nonlinearity parameter, even though they definitely exist. It again illustrates the extreme sensitivity of nonlinear techniques to early stages of damage.

V. NONLINEAR HYSTERETIC MODEL

The particular amplitude-dependent behavior of the resonance frequency after removing the external excitation can be modeled by a nonlinear extension of the Euler beam problem for flexural modes. Following the linear Euler beam theory, the true resonance frequency will be constant in the non-driven phase of reverberation. The nonlinear equation of motion in a one-dimensional flexural system, accounting for attenuation by introducing N relaxation mechanisms as was done by Blanch *et al.*,[35] reads

$$\frac{\partial v}{\partial t} = -\frac{1}{\rho}\frac{\partial \tau}{\partial x},$$

$$\frac{\partial \tau}{\partial t} = E(\tau, \dot{\tau}, \ldots)(1 + N\zeta)\left[\kappa^2 \frac{\partial^3 v}{\partial x^3} - \sum_{j=1}^{N} r_j\right] \tag{5}$$

with

$$\frac{\partial r_j}{\partial t} = -\frac{1}{\zeta_j} r_j + \frac{\zeta}{\zeta_j(1 + N\zeta)} \kappa^2 \frac{\partial^3 v}{\partial x^3},$$

$$\zeta = \frac{1}{Q_0} \frac{\int_{\omega_a}^{\omega_b} F(\omega;\zeta_1,\ldots,\zeta_N)d\omega}{\int_{\omega_a}^{\omega_b} [F(\omega;\zeta_1,\ldots,\zeta_N)]^2 d\omega},$$

and

$$F(\omega;\zeta_1,\ldots,\zeta_N) = \sum_{j=1}^{N} \frac{\omega\zeta_j}{1+\omega^2\zeta_j^2}.$$

The relaxation times ζ_j $(1 \leq j \leq N)$ and the number of mechanisms to be taken into account can be optimized to simulate a medium with constant Q_0 (quality factor) over a large frequency range $[\omega_a, \omega_b]$.[35]

In Eq. (5), v and τ are the particle velocity and the internal shear stress of the beam as functions of position x $(0 \leq x \leq L)$ and time t, E is the stress- and stress-rate-dependent Young's modulus, ρ is the mass density, and κ is the radius of gyration. In the case of the first fundamental mode, Eq. (3) provides the relation between the maximum amplitude of the internal shear strain η in the center of the beam and the maximum particle velocity v at the edges of the beam.[31,32]

In general, nonlinearity can be included by allowing the Young's modulus to depend on the shear stresses and shear stress rates [and if necessary other history-dependent variables, i.e., $E(\tau, \dot{\tau}, \ldots)$]. History and shear rate dependence of elastic moduli is typical of hysteretic media and has important consequences for the bookkeeping of the stress-strain or modulus-stress response when seeking a numerical solution since the value of the modulus needs to be updated at each time step and at each location based on (parts of) the previous shear stress history. Probably the most general method to deal with history-dependent moduli and hysteretic stress-strain relations is the Preisach–Mayergoysz (PM) approach. This technique follows the evolution of a statistical distribution of bistable elements as function of a control parameter (e.g., the shear stress) and transforms it into the evolution of the response function (e.g., the shear strain).[36–39] In the following simulations, the authors use the PM approach to take account of the nonlinearity.

Upon performing the numerical simulations, the geometrical parameters were measured for each sample, yielding input values for ρ and κ. Five relaxation mechanisms are assumed to provide a constant Q_0 value over a broad frequency range (0.1–5000 Hz). Further, the linear value of the Young's modulus E_0 and the linear quality factor Q_0 are adjusted to obtain the correct low amplitude values for each sample. They are assumed to be uniform over the beam length. The nonlinearity is introduced by specifying the statistical distribution of the bistable PM elements. The simplest way, which is most commonly used for dynamic processes, is to assume a uniform distribution of the elements. In this case, only one parameter is needed. The authors call γ the PM background density parameter [which is expressed in units of $(\text{Pa})^{-2}$][36–39] and assume that its value is uniform over the length of the sample (simulating a uniform distribution of damage). The physical meaning of the dimensionless quantity $\gamma dP_c dP_o$ is that it represents the deformation contribution of the hysteretic elements in the PM space with opening pressures between P_o and P_o+dP_o and closing pressures between P_c and P_c+dP_c upon switching from one state to the other (open to closed or closed to open). The larger γ, the larger the nonlinear strain contribution. This is the only free parameter to be used for fitting the nonlinear behavior.

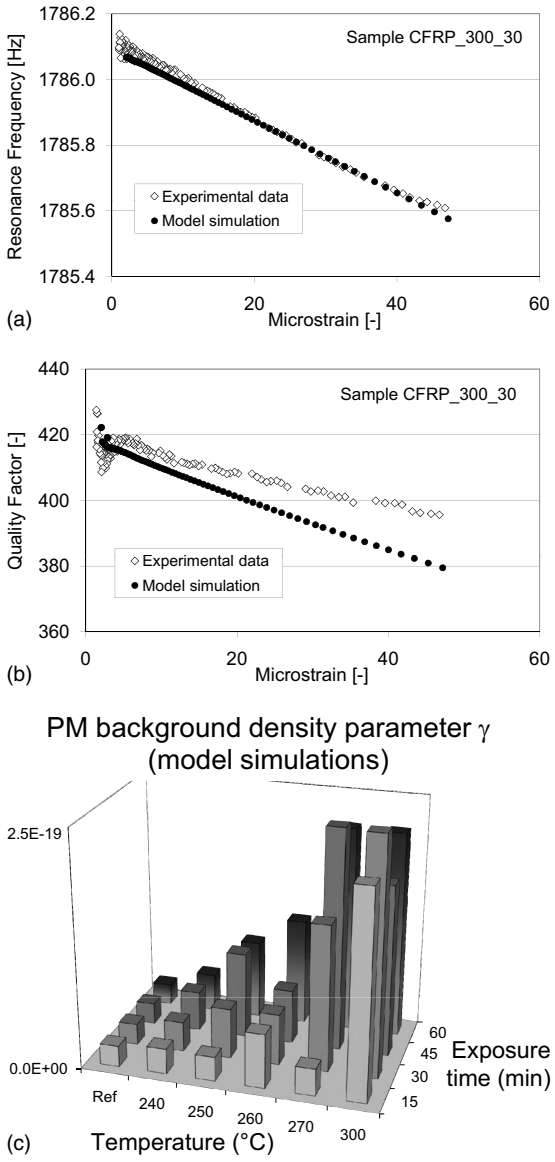


FIG. 9. Model results for CFRP_300_30 and comparison with experimental data. PM background density parameter γ used in the model simulations for all samples. The reference point was duplicated for different exposure times to help visualize the trend of the evolution of the parameters with temperature.

The numerical experiment is performed in the same way as the actual experiment. A harmonic forcing is applied for times $t < t_0$ and removed at t_0 [an external force can be added to the first equation in Eq. (5)]. The response in terms of particle velocity is calculated for $t < t_1$, with $t_0 < t_1$. Given the reverberation signal for $t_0 < t < t_1$, the authors apply the same analysis procedure, as was done for the experimental data, and fine-tune the values of E_0, Q_0, and γ so that the best agreement between model and experiment is obtained.

The comparison of the results for an exposure to 300 °C for 30 min is shown in Fig. 9. The simulations track the experimentally observed resonance frequency reduction extremely well. For the nonlinearity in the damping, the experi-

mental data are generally noisier. Nevertheless, the authors find more or less the same tendency as that predicted in the simulations (with a 5% error at 50 microstrain). The discrepancy could be due to the non-ideal experimental support of the beams by the nylon wires located at the node lines.

Three important issues about the nonlinearity parameter quantification should be noted.

(1) The use of reversible nonlinear models, such as the polynomial expansion of stress versus strain (or vice-versa),[26] would lead to a quadratic behavior of the resonance frequency shift with amplitude and does not affect the attenuation characteristic. To find the linear decrease observed in the data for the resonance frequency and the quality factor Q, it is essential to start from a hysteretic model.
(2) The PM background density parameter γ used in the numerical model is quite small. For the simulation of the nonlinear effects measured in the experiments, the authors used a value of γ between 2.2×10^{-20} Pa^{-2} (reference sample) and 2.2×10^{-19} Pa^{-2} (300 °C for 60 min). As mentioned above, $\gamma dP_c dP_o$ represents the deformation contribution of the hysteretic elements in the PM space with opening pressures between P_o and P_o+dP_o and closing pressures between P_c and P_c+dP_c upon switching from one state to the other. If the authors assumed a constant density in the statistical PM space, ranging from -5 to 5 MPa, this would amount to a total hysteretic contribution to the strain of only $\gamma\int_{-5\ \mathrm{MPa}}^{5\ \mathrm{MPa}}\int_{-5\ \mathrm{MPa}}^{P_0} 1 dP_c dP_0 = \gamma\frac{1}{2}10^{14} \approx 10^{-6}-10^{-5}$ when changing the stress from -5 to 5 MPa.
(3) Based on the PM space approach,[36–39] the relative modulus change is—at first order of approximation—proportional to the constant background density parameter γ of the PM space, the linear modulus E_0, and the stress change itself. Since stress and strain are linked by the modulus, the authors obtain $[E_0-E(\eta)]/E_0 \propto \gamma E_0^2 \eta$.

For those levels of nonlinear behavior observed in this study, giving rise to small frequency or modulus shifts, the authors indeed obtain in all cases a constant ratio between the macroscopically observed NRS nonlinearity parameter Γ and the theoretically found microscopic nonlinearity, which is expressed by the PM background density parameter γ: $(\Gamma/\gamma)(12/E_0)^2=1$.

VI. CONCLUSION

In this paper, the authors demonstrated the efficiency of the NRS technique to detect damage. NRS is the time domain analog of the SIMONRUS technique. Examples for global damage features in the case of thermal exposure on CFRP beams were given. The authors applied a crack density imaging procedure and managed to obtain the first ever quantification of the nonlinearity signature in terms of the crack density. This result confirms that the increase in nonlinearity is linked to an increased network of cracks and that the nonlinear signature is even sensitive to microscopic alterations (e.g., dislocations) in the microstructure well before the visualization threshold for microcracks by the currently used optical technique. Numerical simulations of wave resonances based on a uniform distribution of a hysteretic nonlinear constitutive relation within the sample support the results and relate the macroscopic NRS nonlinearity parameter to the microscopic PM background density parameter of hysteretic (bistable) elements.

The NRS technique has the advantage to be fast; it has low or no restrictions on the sample geometry; and in most cases it can be implemented in a fully non-contact manner (due to the low frequency nature of the method). On the other hand, it necessitates free (or at least steady, amplitude-independent) boundary condition and is only applicable for low attenuation materials.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the European FP6 Grant AERONEWS (Grant No. AST-CT-2003-502927), the Flemish Fund for Scientific Research (Grant Nos. G.0206.02, G.0554.06, and G.0443.07), the Research Council of the Katholieke Universiteit Leuven (Grant Nos. OT/07/051 and CIF1), and the institutional support of the Los Alamos National Laboratory.

[1] P. A. Johnson, "The new wave in acoustic testing," Mater. World **7**, 544–546 (1999).
[2] P. P. Delsanto, *The Universality of Nonclassical Nonlinearity with Applications to NDE and Ultrasonics* (Springer, New York, 2006).
[3] O. Buck, W. L. Morris, and J. M. Richardson, "Acoustic harmonic-generation at unbonded interfaces and fatigue cracks," Appl. Phys. Lett. **33**, 371–373 (1978).
[4] J. H. Cantrell and W. T. Yost, "Acoustic harmonic-generation from fatigue-induced dislocation dipoles," Philos. Mag. A **69**, 315–326 (1994).
[5] 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. **11B**, pp. 2025–2032.
[6] K. Van Den Abeele, A. Sutin, and P. A. Johnson, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. Part I: Nonlinear wave modulation spectroscopy (NWMS)," Res. Nondestruct. Eval. **12**, 17–30 (2000).
[7] V. V. Kazakov, A. Sutin, and P. A. Johnson, "Sensitive imaging of an elastic nonlinear wave-scattering source in a solid," Appl. Phys. Lett. **81**, 646–648 (2002).
[8] A. Zagrai, D. Donskoy, A. Chudnovsky, and E. Golovin, "Micro- and macroscale damage detection using the nonlinear acoustic vibro-modulation technique," Res. Nondestruct. Eval. **19**, 104–128 (2008).
[9] C. R. P. Courtney, B. W. Drinkwater, S. A. Neild, and P. D. Wilcox, "Factors affecting the ultrasonic intermodulation crack detection technique using bispectral analysis," NDT & E Int. **41**, 223–234 (2008).
[10] T. J. Ulrich, P. A. Johnson, and R. A. Guyer, "Interaction dynamics of elastic waves with a complex nonlinear scatterer through the use of a time reversal mirror," Phys. Rev. Lett. **98**, 104301 (2007).
[11] V. Nazarov, L. Ostrovsky, I. Soustova, and A. Sutin, "Nonlinear acoustics of micro-inhomogeneous media," Phys. Earth Planet Inter. **50**, 65–73 (1988).
[12] P. A. Johnson, B. Zinszner, and P. N. J. Rasolofosaon, "Resonance and elastic nonlinear phenomena in rock," J. Geophys. Res. **101**, 11553–11564 (1996).
[13] K. Van Den Abeele, J. Carmeliet, J. A. TenCate, and P. A. Johnson, "Nonlinear elastic wave spectroscopy (NEWS) techniques to discern material damage. Part II: Single-mode nonlinear resonance acoustic spectroscopy," Res. Nondestruct. Eval. **12**, 31–42 (2000).
[14] P. A. Johnson, B. Zinszner, P. Rasolofosaon, K. Van Den Abeele, and F. Cohen-Tenoudji, "Dynamic measurements of the nonlinear elastic parameter alpha in rock under varying conditions," J. Geophys. Res. **109**, 10129–10139 (2004).
[15] P. A. Johnson and A. Sutin, "Slow dynamics and anomalous nonlinear fast dynamics in diverse solids," J. Acoust. Soc. Am. **117**, 124–130 (2005).

[16]J. A. TenCate and T. J. Shankland, "Slow dynamics in the nonlinear elastic response of Berea sandstone," Geophys. Res. Lett. **23**, 3019–3022 (1996).
[17]J. A. TenCate, E. A. Smith, and R. A. Guyer, "Universal slow dynamics in granular solids," Phys. Rev. Lett. **85**, 1020–1023 (2000).
[18]M. Bentahar, H. El Aqra, R. El Guerjouma, M. Griffa, and M. Scalerandi, "Hysteretic elasticity in damaged concrete: Quantitative analysis of slow and fast dynamics," Phys. Rev. B **73**, 014116 (2006).
[19]M. Vila, F. Vander Meulen, S. Dos Santos, L. Haumesser, and O. Bou Matar, "Contact phase modulation method for acoustic nonlinear parameter measurement in solid," Ultrasonics **42**, 1061–1065 (2004).
[20]P. B. Nagy, "Fatigue damage assessment by nonlinear ultrasonic materials characterization," Ultrasonics **36**, 375–381 (1998).
[21]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).
[22]K. Van Den Abeele and J. De Visscher, "Damage assessment in reinforced concrete using spectral and temporal nonlinear vibration technique," Cem. Concr. Res. **30**, 1453–1464 (2000).
[23]C. Payan, V. Garnier, J. Moysan, and P. A. Johnson, "Applying nonlinear resonant ultrasound spectroscopy to improving thermal damage assessment in concrete," J. Acoust. Soc. Am. **121**, EL125–EL130 (2007).
[24]K. Van Den Abeele, C. Campos-Pozuelo, J. Gallego-Juarez, F. Windels, and B. Bollen, "Analysis of the nonlinear reverberation of titanium alloys fatigued at high amplitude ultrasonic vibration," in Proceedings Forum Acustica Sevilla 2002, (2002).
[25]R. A. Guyer and P. A. Johnson, "Nonlinear mesoscopic elasticity: Evidence for a new class of materials," Phys. Today **52**, 30–36 (1999).
[26]R. A. Guyer, K. R. McCall, and K. E. A. Van Den Abeele, "Slow elastic dynamics in a resonant bar of rock," Geophys. Res. Lett. **25**, 1585–1588 (1998).
[27]R. P. Taylor, "Fiber composite aircraft-capability and safety," Australian Transport Safety Bureau Report No. AR-2007-021, 2008, p. 6, available at http://www.atsb.gov.au/media/27758/ar2007021.pdf (Last viewed July 27, 2009); P. Trum, "Living in a composite material world," in Headway, Research Discovery and Innovation at McGill University, available at http://www.mcgill.ca/headway/winter2008-09/industrialimpact1/ (Last viewed July 27, 2009); "Composite materials and aircraft structures certificate programs overview," University of Washington, College of Engineering, available at http://www.engr.washington.edu/epp/cmc/ (Last viewed July 27, 2009).
[28]G. A. Matzkanin and G. P. Hansen, "Heat damage in graphite epoxy composites: Degradation, measurement and detection—A state-of-the-art report," Report No. NTIAC-SR-98-02, NTIAC, 1998.
[29]G. A. Matzkanin, "Nondestructive characterization of heat damage in graphite/epoxy composite: A state-of-the-art report," Texas Research Institute, Austin, TX, 1995.
[30]F. Hyllengren, Technical Report No. TEK01-0022, C. S. M. Materialteknik, Linköping, Sweden, 2001.
[31]M. Geradin and D. Rixen, *Mechanical Vibrations: Theory and Application to Structural Dynamics* (Wiley, Chichester, 1994).
[32]J. W. Strutt and B. Rayleigh, *The Theory of Sound* (Mac Millan, New York, 1894), Vol. **1**, Chap. 8.
[33]GIMP, GNU Image Manipulation Program, User Manual, Edge-Detect Filters, Sobel, The GIMP Documentation Team, 2008.
[34]T. Acharya and A. K. Ray, *Image Processing: Principles and Applications* (Wiley-Interscience, New York, 2005), pp. 111–113.
[35]J. O. Blanch, J. O. A. Robertsson, and W. W. Symes, "Modeling of a constant-Q-methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique," Geophysics **60**, 176–184 (1995).
[36]K. R. McCall and R. A. Guyer, "Equation of state and wave propagation in hysteretic nonlinear elastic materials," J. Geophys. Res. **99**, 23887–23897 (1994).
[37]R. A. Guyer, K. R. McCall, and G. N. Boittnott, "Hysteresis, discrete memory, and nonlinear-wave propagation in rock—A new paradigm," Phys. Rev. Lett. **74**, 3491–3494 (1995).
[38]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).
[39]K. Van Den Abeele and S. Vanaverbeke, "Multiscale approach and simulations of wave propagation and resonance in media with localized microdamage: 1D and 2D cases," in *The Universality of Nonclassical Nonlinearity with Applications to NDE and Ultrasonics*, edited by P. P. Delsanto (Springer, New York, 2006), Chap. 12, pp. 177–202.

Influence of the bubble-bubble interaction on destruction of encapsulated microbubbles under ultrasound

Kyuichi Yasui,[a] Judy Lee, Toru Tuziuti, Atsuya Towata, Teruyuki Kozuka, and Yasuo Iida
National Institute of Advanced Industrial Science and Technology (AIST), 2266-98 Anagahora, Shimoshidami, Moriyama-ku, Nagoya 463-8560, Japan

(Received 24 March 2009; revised 21 June 2009; accepted 23 June 2009)

Influence of the bubble-bubble interaction on the pulsation of encapsulated microbubbles has been studied by numerical simulations under the condition of the experiment reported by Chang *et al.* [IEEE Trans. Ultrason Ferroelectr. Freq. Control **48**, 161 (2001)]. It has been shown that the natural (resonance) frequency of a microbubble decreases considerably as the microbubble concentration increases to relatively high concentrations. At some concentration, the natural frequency may coincide with the driving frequency. Microbubble pulsation becomes milder as the microbubble concentration increases except at around the resonance condition due to the stronger bubble-bubble interaction. This may be one of the reasons why the threshold of acoustic pressure for destruction of an encapsulated microbubble increases as the microbubble concentration increases. A theoretical model for destruction has been proposed.
 [DOI: 10.1121/1.3179677]

PACS number(s): 43.25.Yw, 43.35.Wa [CCC] Pages: 973–982

I. INTRODUCTION

Encapsulated microbubbles have been used as contrast agents in medical applications.[1,2] There are mainly two types of the shell of an encapsulated microbubble: one is stiff (sonicated albumin or polymer) and the other is more flexible (lipid or phospholipid). The thickness of the shell ranges from 10 to 200 nm. The gas species used in microbubbles can be air or a gas with lower water solubility such as octafluoropropane.

The pulsation of a microbubble may be influenced by the surrounding microbubbles as they may modify the acoustic pressure field. It may be due to the emission of acoustic waves by their pulsation and the shielding of the incident acoustic wave.[3,4] In the present paper, the change in the microbubble pulsation due to the modification of the acoustic pressure field by the former effect is called the bubble-bubble interaction.

The bubble-bubble interaction is stronger for higher number density of microbubbles and larger size of the microbubble cloud.[5] In the previous paper by the authors,[5] it has been shown that the strength of the bubble-bubble interaction may be crudely measured by the "coupling strength" (S) defined as follows:

$$S = \sum_i \frac{1}{d_i} = \int_{l_{\min}}^{l_{\max}} \frac{4\pi r^2 n}{r} dr = 2\pi n (l_{\max}^2 - l_{\min}^2) \approx 2\pi n l_{\max}^2, \tag{1}$$

where d_i is the distance between the bubble and another bubble numbered i, the summation is for all the surrounding bubbles, $l_{\min}$ is the distance between the bubble and the nearest bubble, $l_{\max}$ is the radius of the bubble cloud, n is the number density of bubbles, r is the distance from the bubble, and it is assumed that $l_{\max} \gg l_{\min}$ in the last equation. This quantity can be experimentally estimated by measuring the number density of bubbles (n) and the radius of the bubble cloud ($l_{\max}$).

In some experiments using encapsulated microbubbles, the coupling strength (S) can be much larger than that in experiments of ultrasonic cavitation such as multibubble sonoluminescence and sonochemistry. The coupling strength can be as large as 10^9 m^{-1} for the case of encapsulated microbubbles as described later, while for the case of ultrasonic cavitation it may be 10^6 m^{-1} at most.[5] Thus, in experiments using encapsulated microbubbles, the bubble-bubble interaction can be significantly stronger than that in experiments of ultrasonic cavitation.

The bubble-bubble interaction has been studied both theoretically and experimentally since 1965.[4–32] Recently, Ilinski *et al.*[28] and Doinikov[21] reported the detailed theoretical analysis of the effect. With regard to the natural frequency of the interacting bubbles, Shima[8] first derived an analytic expression for a two bubble system. Later, Morioka[10] derived a more accurate expression and concluded that the natural frequency of a two bubble system decreases as the distance between them decreases when the two bubbles pulsate in phase. Furthermore, Morioka[10] concluded that the natural frequency of a multibubble system is smaller than the independent natural frequency of the largest bubble in the system although the analytical expression was not given. It should be noted that Weston[6] reached the same conclusion for a linear array of bubble spacing with equidistance by the theory of multiple-scattering of ultrasound. Later, Feuillade[15] also reached the same conclusion. Experimentally, Hsiao *et al.*[17] showed that the natural frequency of a two bubble system decreases as the distance between them decreases with regard to the symmetrical mode (lower frequency mode). Payne *et al.*[25] showed experimentally that the

[a]Author to whom correspondence should be addressed. Electronic mail: k.yasui@aist.go.jp

resonance frequency of a two bubble system under a glass plate decreases as the distance between the bubbles decreases. A review on the frequencies of acoustically interacting bubbles was recently written by Manasseh and Ooi.[32] In the present study, the analytic expression of the natural frequency of many interacting bubbles has been given for the first time. It decreases as the number density of bubbles increases to relatively high bubble concentration, which is qualitatively in accord with previous studies.

In the present study, numerical simulations have been performed under the condition of the experiment reported by Chang *et al.*[33] In the experiment of Chang *et al.*,[33] Albunex microbubbles, which are air bubbles encapsulated with human albumin, were used. The average radius of a microbubble was between 1.5 and 2.5 μm. An undiluted concentration was $(3-5)\times 10^8$ microbubbles/ml, which corresponds to 1000 μl/ml.[33] The Albunex was diluted in Isoton II, which is a filtered, buffered saline electrolyte diluent. The Albunex solution in a biocompatible hydrophilic polyester sample tube of 1.0 cm in diameter with 1.5 ml capacity was irradiated by 1.1 MHz ultrasound using a high intensity focused ultrasound (HIFU) transducer. Another 5 MHz transducer was used to detect the acoustic signals. In the experiment, the acoustic pressure transmitted by the 1.1 MHz HIFU transducer was increased by approximately 135 kPa every second and the threshold acoustic pressure (P_{th}) for the complete destruction of the microbubbles was measured. With intact microbubbles, the echo ultrasound from the rear wall of the tube decreased because of attenuation. At P_{th}, this echo signal returned to its initial amplitude and P_{th} was determined. They measured P_{th} as a function of the concentration of microbubbles, pulse repetition frequency (PRF), and the number of cycles per pulse. P_{th} increases as the concentration of microbubbles increases in the range of 0–100 μl/ml. P_{th} decreases as PRF increases in the range of 0–5 kHz. P_{th} decreases as the number of cycles per pulse increases in the range of 1–40. In the present paper, the effect of the microbubble concentration has been studied by numerical simulations.

Recently, many groups reported the results of experiments on destruction of encapsulated microbubbles.[33–52] Such studies are important because in the contrast-enhanced ultrasound imaging the knowledge of acoustic pressure thresholds for microbubble destruction is required as the completely destructed microbubbles do not work as contrast agents anymore. Furthermore, it is also required for therapeutic application of microbubbles because microbubbles carrying drug should be destroyed by ultrasound at the affected part.

In spite of the recent interest in the destruction mechanism of encapsulated microbubbles, there have been few theoretical or numerical studies on the subject.[53,54] Stride and Saffari[53] calculated the shell wall stresses during the microbubble pulsation under ultrasound. Postema and Schmitz[54] considered only fragmentation of a microbubble and did not study the rupture of the shell due to surface tension. In these studies,[53,54] the effect of the bubble-bubble interaction has been neglected. In the present paper, a theoretical model of destruction of a microbubble has been constructed taking into account the effect of the bubble-bubble interaction.

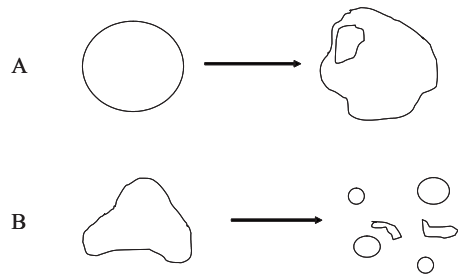


FIG. 1. Two mechanisms in destruction of a microbubble covered with a stiff shell such as an Albunex microbubble. (A) Rupture of the shell due to surface tension. (B) Fragmentation of a microbubble due to its shape instability resulting in several daughter bubbles and fragmented shell.

II. MODEL

There are mainly two mechanisms behind the destruction of a microbubble covered with a stiff shell such as an Albunex microbubble (Fig. 1).[43–45,47,48] One is the rupture of the shell due to surface tension possibly during the microbubble expansion.[43–45,47] It should be noted that diffusion of gas out of a microbubble may also result in the stronger surface tension which is proportional to the pressure difference between inside and outside a microbubble. It may result in the rupture of the shell, which is in other words the rupture during deflation of a microbubble.[34] The other is the fragmentation of a microbubble due to its shape instability, which also result in the rupture of the shell.[48]

First, the theoretical model of rupture of the shell will be discussed following the study of the membrane rupture by Evans *et al.*[55] There are three states in a membrane: defect-free state, metastable state, and ruptured state. The metastable state is that there is an annihilable defect on the membrane, which occasionally vanishes (defect-free state) or evolves to an unstable hole (ruptured state).[55] The temporal evolution of the probabilities of being in the defect-free state (S_0), metastable state (S_*), and ruptured state (S_r) is predicted by the following hierarchy of the statistical master equations:[55]

$$\frac{dS_0}{dt} = -\nu_{0\to *}S_0 + \nu_{*\to 0}S_*,$$

$$\frac{dS_*}{dt} = -[\nu_{*\to 0} + \nu_{*\to r}]S_* + \nu_{0\to *}S_0,$$

$$\frac{dS_r}{dt} = \nu_{*\to r}S_*, \tag{2}$$

where $\nu_{0\to *}$ is the rate of the defect formation, $\nu_{*\to 0}$ is the rate of the defect annihilation, and $\nu_{*\to r}$ is the rate of the unstable-hole formation from a defect. The rates are estimated by the following equations:

$$\nu_{0\to *} = \nu_0 \exp(\sigma/\sigma_\delta), \tag{3}$$

$$\nu_{*\to 0} = \nu_0 \exp(E_0/k_B T), \tag{4}$$

$$\nu_{*\to r} = \nu_\delta (\sigma/\sigma_c)^{1/2} \exp(-\sigma_c/\sigma), \tag{5}$$

where σ is the surface tension, k_B is the Boltzmann constant, T is the temperature, and there are five parameters that characterize the membrane (the shell of a microbubble); ν_0, σ_δ, E_0, ν_δ, and σ_c. According to Evans *et al.*,[55] the range of the parameters are as follows for phosphatidylcholine (phospholipid) membranes: $\nu_0 = 0.09-8.0\ \mathrm{s}^{-1}$, $\sigma_\delta = 0.003-0.0045\ \mathrm{N/m}$, $E_0 = 0-3k_BT$, $\nu_\delta = 1\times10^6-8\times10^6\ \mathrm{s}^{-1}$, and $\sigma_c = 0.03-0.22\ \mathrm{N/m}$.

According to Eq. (2), even without ultrasound, there may be a finite non-zero probability of being in metastable state as given by Eq. (6).

$$S_*(t=0) = \frac{\exp(\sigma_0/\sigma_\delta)}{\exp(\sigma_0/\sigma_\delta) + \exp(E_0/k_BT)}, \tag{6}$$

$$S_0(t=0) = 1 - S_*(t=0), \tag{7}$$

$$S_r(t=0) = 0, \tag{8}$$

where σ_0 is the surface tension of a microbubble at its ambient (initial) size. In deriving Eqs. (6)–(8), it has been assumed that the rate of rupture is negligible without ultrasound. The lifetime (τ_0) of an encapsulated microbubble due to rupture without ultrasound may be estimated by Eq. (9) when there is no gas diffusion across the shell.

$$\tau_0 = \frac{(\sigma_c/\sigma_0)^{1/2} \exp(\sigma_c/\sigma_0)(\exp(\sigma_0/\sigma_\delta) + \exp(E_0/k_BT))}{\nu_\delta \exp(\sigma_0/\sigma_\delta)}, \tag{9}$$

where ν_δ may be a function of σ_c and the membrane surface-shear viscosity (η) as Eq. (10).[55]

$$\nu_\delta \approx \sigma_c/(2\pi^{1/2}\eta). \tag{10}$$

For an Albunex microbubble, the five parameters (ν_0, σ_δ, E_0, ν_δ, and σ_c) are unknown at present. Thus, they have been determined as follows in the present paper. Without ultrasound, the probability of being in metastable state may be much smaller than that of being in defect-free state. This assumption may result in the condition $\sigma_\delta \geq \sigma_0$ according to Eq. (6). In the present numerical simulations, $\sigma_\delta = \sigma_0 = 0.04\ \mathrm{N/m}$ has been assumed.[56] In Fig. 2, the calculated lifetime (τ_0) of an encapsulated microbubble without ultrasound has been shown as a function of σ_c. It has been assumed that $E_0 = 3k_BT$ and $\nu_\delta = 5\times10^6\sigma_c\ (\mathrm{s}^{-1})$, where σ_c is in N/m. The latter equation may be derived as follows. Although the surface-shear viscosity (η) of albumin shell is unknown at present, it might be two orders of magnitude larger than that of a lipid membrane.[57,58] Furthermore, ν_δ for $E_0 = 3k_BT$ may be an order of magnitude larger than the values for $E_0 = 0$ listed by Evans *et al.*[55] Thus, ν_δ may be estimated by the above equation using the value listed by Evans *et al.*[55] and Eq. (10). From Fig. 2, it is seen that the lifetime (τ_0) by rupture without ultrasound decreases as σ_c decreases. For example, τ_0 is only 400 s for $\sigma_c = 0.7\ \mathrm{N/m}$, which is unrealistic when there is no gas diffusion across the shell.

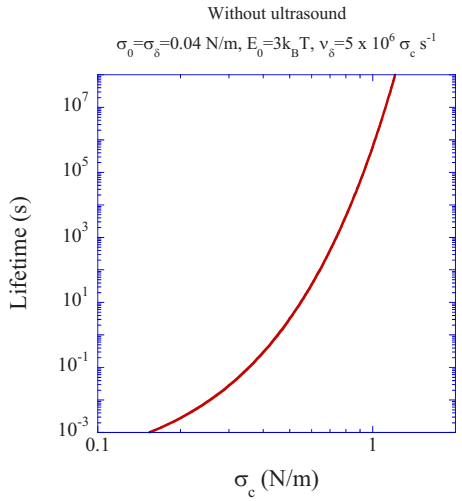


FIG. 2. (Color online) The lifetime (τ_0) of an encapsulated microbubble due to the rupture of the shell without ultrasound calculated by Eq. (9) as a function of σ_c. The parameters are as follows: $\sigma_0 = \sigma_\delta = 0.04\ \mathrm{N/m}$, $E_0 = 3k_BT$, and $\nu_\delta = 5\times10^6\sigma_c\ (\mathrm{s}^{-1})$, where σ_c is in N/m.

Thus, it may be concluded that $\sigma_c > 0.7\ \mathrm{N/m}$. The upper limit of σ_c has been determined as follows. The lifetime (τ) of an encapsulated microbubble due to the rupture of the shell under ultrasound may be estimated by Eq. (11).

$$\tau = \max\left(\frac{1}{\Delta S_r \times F_{\mathrm{PR}}}, \frac{S_0(t=0)}{|\Delta S_0| \times F_{\mathrm{PR}}}\right), \tag{11}$$

where ΔS_r and ΔS_0 are the changes in S_r and S_0, respectively, per ultrasound pulse-on and pulse-off cycles, F_{PR} is the pulse repetition frequency, and $S_0(t=0)$ is the initial value of S_0 given by Eq. (7). When the former term in the right hand side of Eq. (11) is larger, the rupture is called "cavitation-limited" where "cavitation" does not mean ultrasonic cavitation in liquid but the unstable-hole formation from a defect on the shell. When the latter is larger, it is called "defect-limited." For the comparison with the experimental data of Chang *et al.*,[33] it has been assumed that the shell is ruptured when $\tau \leq 1$ s as in the experiment the acoustic pressure was increased by approximately 135 kPa every second. According to the experimental data[33] of the destruction threshold at the microbubble concentration of 100 μl/ml, Eq. (11) requires that $\sigma_c = 1\ \mathrm{N/m}$ and $\nu_0 \geq 8\ \mathrm{s}^{-1}$ or $\sigma_c < 1\ \mathrm{N/m}$ and $\nu_0 = 8\ \mathrm{s}^{-1}$. Thus, in the present paper, $\sigma_c = 1\ \mathrm{N/m}$ and $\nu_0 = 8\ \mathrm{s}^{-1}$ have been assumed. It should be noted that qualitative conclusions in the present paper do not depend on the values of the parameters if the above conditions are satisfied. The radius-time curve of an encapsulated microbubble does not depend on the parameters at all.

Next we will discuss the model for the pulsation of an encapsulated microbubble. In the case of an isolated microbubble without any bubble-bubble interactions, the pulsation may be described by the following modified Rayleigh–Plesset equation:[59,60]

$$\left(1-\frac{\dot{R}}{c_\infty}\right)R\ddot{R}+\frac{3}{2}\dot{R}^2\left(1-\frac{\dot{R}}{3c_\infty}\right)=\frac{1}{\rho_{L,\infty}}\left(1+\frac{\dot{R}}{c_\infty}\right)\left[p_B-p_S\left(t+\frac{R}{c_\infty}\right)-p_\infty\right]+\frac{R}{c_\infty\rho_{L,\infty}}\frac{dp_B}{dt}, \tag{12}$$

where R is the instantaneous radius of a microbubble, the dot denotes the time derivative (d/dt), c_∞ is the sound velocity in the liquid, $\rho_{L,\infty}$ is the density of the liquid far from a microbubble, $p_B(t)$ is the liquid pressure on the external side of the bubble wall, $p_S(t)$ is a nonconstant ambient pressure component such as a sound field, and p_∞ is the undisturbed ambient pressure. $p_B(t)$ may be calculated by Eq. (13).[59]

$$p_B(t)=p_g(t)-\frac{2\sigma(R)}{R}-4\mu\frac{\dot{R}}{R}-4\kappa_s\frac{\dot{R}}{R^2}, \tag{13}$$

where $p_g(t)$ is the gas pressure inside a microbubble, $\sigma(R)$ is the surface tension as a function of the microbubble radius, μ is the viscosity of the liquid, and κ_s is the surface dilatational viscosity of the shell. $p_g(t)$ is calculated by van der Waals equations of state inside a microbubble. $\sigma(R)$ is assumed as Eq. (14) following the model of Marmottant *et al.*[59]

$$\text{When } R_{\text{buckling}}\leqslant R,\quad \sigma(R)=\chi\left(\frac{R^2}{R_{\text{buckling}}^2}-1\right).$$

$$\text{When } R\leqslant R_{\text{buckling}},\quad \sigma(R)=0, \tag{14}$$

where it is assumed that surface tension decreases continuously as the bubble radius decreases until surface tension vanishes. In Eq. (14), χ is the elasticity of the shell and related to the shell stiffness coefficient (S_p) defined by de Jong *et al.*[61] as follows:[59] $S_p=2\chi$. The shell friction coefficient (S_f) defined by de Jong *et al.*[61] is related to the surface dilatational viscosity (κ_s) of the shell by $S_f=12\pi\kappa_s$. For Albunex microbubbles, de Jong *et al.*[61] reported that $S_p=8$ N/m and $S_f=4\times10^{-6}$ N s/m. Thus, $\chi=4$ N/m, $\kappa_s=1.06\times10^{-7}$ N s/m have been assumed in the present numerical simulations. Assuming that the surface tension of an Albunex microbubble at ambient radius is 0.04 N/m,[56] R_{buckling} has been determined as $0.995R_0$.

When a bubble is irradiated by an acoustic wave where the wavelength is much longer than the bubble radius, the nonconstant ambient pressure in Eq. (12) is described as follows: $p_s(t)=-p_a\sin(2\pi f_a t)$, where p_a is the pressure amplitude of the acoustic wave (the acoustic amplitude) and f_a is the frequency.

In the present model, the effect of the bubble-bubble interaction is approximately taken into account by Eq. (15).[5]

$$\left(1-\frac{\dot{R}}{c_\infty}\right)R\ddot{R}+\frac{3}{2}\dot{R}^2\left(1-\frac{\dot{R}}{3c_\infty}\right)=\frac{1}{\rho_{L,\infty}}\left(1+\frac{\dot{R}}{c_\infty}\right)\left[p_B-p_S\left(t+\frac{R}{c_\infty}\right)-\rho_\infty\right]+\frac{R}{c_\infty\rho_{L,\infty}}\frac{dp_B}{dt}-S(R^2\ddot{R}+2R\dot{R}^2), \tag{15}$$

where S is the coupling strength of microbubbles and estimated by Eq. (1). The last term in the right hand side of Eq. (15) is the effect of the bubble-bubble interaction on the pulsation of a bubble. Although Eq. (15) is an approximate one with regard to the bubble-bubble interaction, it has been validated through the study of the bubble pulsation under an ultrasonic horn.[5] It should be noted that in the last term of Eq. (15) the effect of the time delays due to the finite speed propagation of acoustic waves radiated from surrounding bubbles is neglected.[23,31]

The bubble-bubble interaction may influence the microbubble pulsation through Eq. (15), which may affect the surface tension as a function of time by Eq. (14). It may influence the probability of being in ruptured state through Eqs. (3) and (5). Thus, the bubble-bubble interaction may influence the threshold acoustic pressure for rupture of the shell.

Next, fragmentation of a microbubble will be discussed. Even a microbubble covered with a stiff shell such as an Albunex microbubble may disintegrate into daughter bubbles due to its shape instability.[48] The amplitude of the shape oscillation of a microbubble is calculated as follows.[62,63] A small distortion of the spherical surface is described by $R(t)+a_n(t)Y_n$, where $R(t)$ is the instantaneous mean radius of a bubble at time t, $a_n(t)$ is the distortion amplitude, and Y_n is a spherical harmonic of degree n. The dynamics of the distortion amplitude is described by Eq. (16).

$$\ddot{a}_n+B_n(t)\dot{a}_n-A_n(t)a_n=0, \tag{16}$$

where the dot denotes the time derivative (d/dt),

$$A_n(t)=(n-1)\frac{\ddot{R}}{R}-\frac{\beta_n\sigma}{\rho_{L,\infty}R^3}-\left[(n-1)(n+2)+2n(n+2)\times(n-1)\frac{\delta}{R}\right]\frac{2\mu\dot{R}}{R^3}, \tag{17}$$

and

$$B_n(t)=\frac{3\dot{R}}{R}+\left[(n+2)(2n+1)-2n(n+2)^2\frac{\delta}{R}\right]\frac{2\mu}{R^2}. \tag{18}$$

In Eq. (17), $\beta_n=(n-1)(n+1)(n+2)$, σ is the surface tension, $\rho_{L,\infty}$ is the liquid density far from a bubble,

$$\delta=\min\left(\sqrt{\frac{\mu}{\omega}},\frac{R}{2n}\right),$$

μ is the liquid viscosity, and ω is the angular frequency of ultrasound. In the present paper, only $n=2$ and $n=3$ modes (a_2 and a_3) were numerically simulated as they are the dominant modes under most conditions. The condition for the fragmentation of a microbubble is assumed as follows:

$$a_2>R \quad \text{or} \quad a_3>R. \tag{19}$$

The other part of the model of the bubble pulsation has been described in Ref. 60 except the following modifications. The effect of thermal conduction outside a microbubble has been neglected in the present study, while the effect of thermal conduction inside a bubble has been taken into account. Furthermore, the effect of mass transfer has been neglected in the present study.

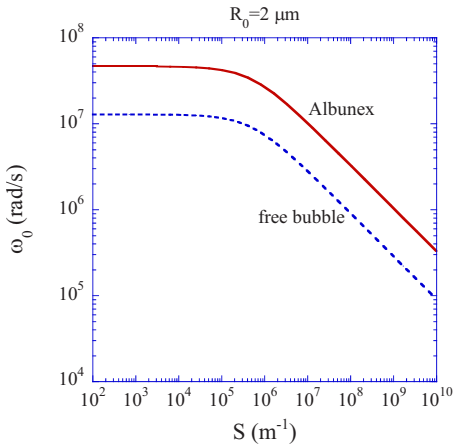


FIG. 3. (Color online) The calculated natural frequency of a free bubble [Eq. (21)] and an Albunex microbubble [Eq. (23)] as a function of the coupling strength (S) of the bubble-bubble interaction. The coupling strength is proportional to the number density of bubbles [Eq. (1)]. The ambient bubble radius is 2 μm.

III. RESULTS AND DISCUSSIONS

From Eq. (15), the natural (resonance) frequency of a microbubble may depend on the coupling strength (S) of the bubble-bubble interaction. In other words, it may depend on the microbubble number density.

First, the natural frequency of a free bubble without a shell will be discussed taking into account the bubble-bubble interaction by using Eq. (15). For this purpose, a small-amplitude linear pulsation of a bubble is considered.

$$R = R_0(1 + x(t)), \tag{20}$$

where R_0 is the ambient bubble radius and $x(t)$ is a small-amplitude sinusoidal function of time. Inserting Eq. (20) into Eq. (15) yields the natural (resonance) frequency of a microbubble taking into account the effect of the bubble-bubble interaction.

$$\omega_0 = \sqrt{\frac{3\gamma p_\infty + (3\gamma - 1)2\sigma/R_0}{\rho_{L,\infty} R_0 (R_0 + SR_0^2 + 4\mu/c_\infty \rho_{L,\infty})}}, \tag{21}$$

where ω_0 is the natural (resonance) angular frequency of a bubble, γ is the ratio of the specific heats (γ=1.4 for air), p_∞ is the undisturbed ambient pressure, σ is the surface tension, R_0 is the ambient bubble radius, $\rho_{L,\infty}$ is the liquid density far from a bubble, S is the coupling strength of the bubble-bubble interaction, μ is the liquid viscosity, and c_∞ is the sound velocity in the liquid. In the derivation of Eq. (21), it has been assumed that the bubble pulsation is adiabatic as follows:

$$p_B(t) = p_g(t) - \frac{2\sigma}{R} - \frac{4\mu\dot{R}}{R} = \left(p_\infty + \frac{2\sigma}{R_0}\right)\left(\frac{R}{R_0}\right)^{-3\gamma} - \frac{2\sigma}{R} - \frac{4\mu\dot{R}}{R}. \tag{22}$$

The natural (resonance) angular frequency of a free bubble without a shell (ω_0) has been shown as a function of the coupling strength (S) by a dashed line in Fig. 3 when R_0=2 μm, γ=1.4, σ=72.75 mN/m, μ=1.002$\times 10^{-3}$ Pa s, $\rho_{L,\infty}$=9.982$\times 10^2$ kg/m^3, and c_∞=1483 m/s, which are the values for an air bubble in water at 20 °C. It is seen that the natural frequency decreases considerably as the coupling strength increases above about 10^5 m^{-1} which corresponds to the bubble number density of 6.4$\times 10^2$ bubbles/ml when $l_{\max}$=0.5 cm. It should be noted, however, in Eq. (21) that the effect of time delays due to finite speed propagation of acoustic waves radiated from surrounding bubbles is neglected.[23,31]

With regard to an encapsulated microbubble, the natural (resonance) angular frequency (ω_0) is derived from Eq. (15) assuming the adiabatic pulsation.

$$\omega_0 = \sqrt{\frac{3\gamma p_\infty + 2\chi[(3\gamma+1)R_0/R_{\text{buckling}}^2 + (1-3\gamma)/R_0]}{\rho_{L,\infty}(R_0^2 + SR_0^3 + 4\mu R_0/c_\infty\rho_{L,\infty} + 4\kappa_s/c_\infty\rho_{L,\infty})}}, \tag{23}$$

where χ and R_{buckling} have been defined in Eq. (14), and κ_s is the surface dilatational viscosity of the shell. The natural angular frequency of an encapsulated microbubble has been shown in Fig. 3 by a solid line as a function of the coupling strength (S) when R_0=2 μm, χ=4 N/m, and κ_s=1.06 $\times 10^{-7}$ N s/m, which are the values for a typical Albunex microbubble.[61] It is seen that the natural frequency of an Albunex microbubble is higher than that of a free bubble and that it decreases considerably as the coupling strength increases above about 10^5 m^{-1} as in the case of a free bubble. In other words, the natural (resonance) frequency decreases as the concentration of microbubbles increases above about 6.4$\times 10^2$ bubbles/ml when $l_{\max}$=0.5 cm, which is the case for the experiment by Chang *et al.*[33] because the typical number density of microbubbles was larger than 10^5 bubbles/ml.

In Fig. 4, the calculated result for an isolated microbubble without any bubble-bubble interaction has been shown under a typical condition of the experiment of Chang *et al.*[33] (1.1 MHz and 1.1 MPa in ultrasonic frequency and its pressure amplitude). In Fig. 4(a), the radius of an Albunex microbubble has been shown as a function of time for 5 μs which is 5.5 acoustic cycles. The natural frequency calculated by Eq. (23) is about 7.3 MHz for an isolated Albunex microbubble, which corresponds to the cycle of about 0.14 μs. In Fig. 4(a), it is seen that the pulsation with natural frequency whose period is about 0.14 μs is superimposed on the pulsation with the driving frequency whose period is about 0.91 μs.

In Fig. 4(b), the probabilities of being in defect-free state, metastable state, and ruptured state have been shown as a function of time with the same time axis as that of Fig. 4(a). It is seen that the probability of being in defect-free state immediately falls to nearly zero by the irradiation of ultrasound and that inversely the probability of being in metastable state suddenly increases to nearly 1. The probability of being in ruptured state gradually increases especially during the bubble expansion due to stronger surface tension. Accordingly, the probability of being in metastable state gradually decreases.

In Fig. 4(c), the amplitude of the shape oscillation of n=2 mode relative to the instantaneous bubble radius has been

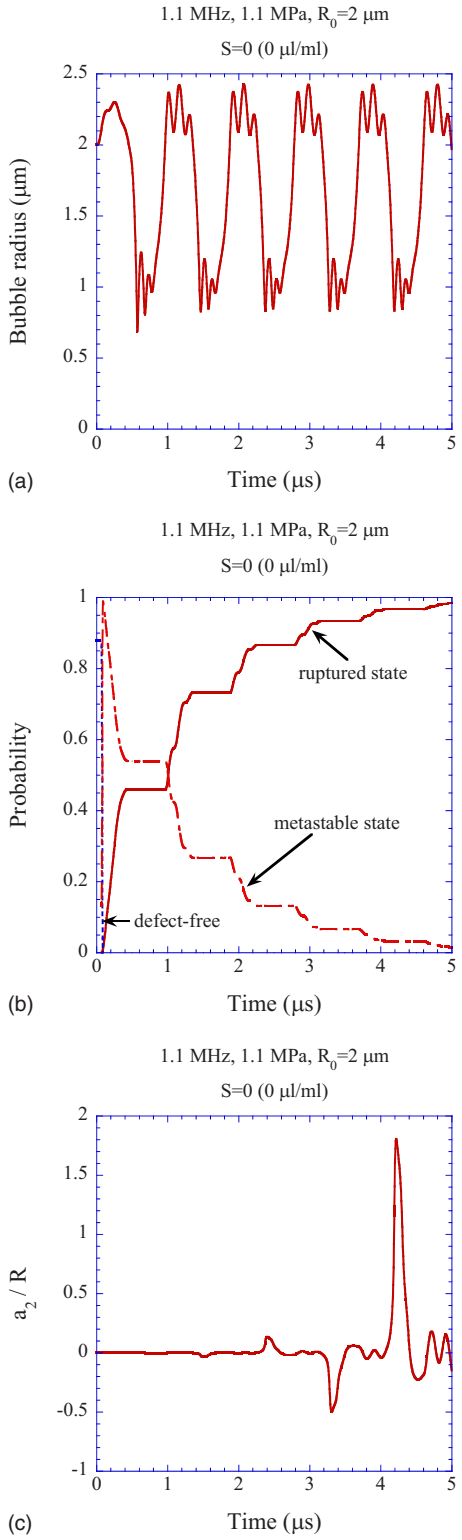


FIG. 4. (Color online) The result of the numerical simulation for an isolated encapsulated microbubble when the frequency and pressure amplitude of ultrasound are 1.1 MHz and 1.1 MPa, respectively. The ambient radius of a microbubble is 2 μm, which is the same as that in Figs. 3 and 5–8. The time axes are the same for (a)–(c) for 5 μs which corresponds to 5.5 acoustic cycles. (a) The radius of an Albunex microbubble. The pulsation with natural frequency [7.3 MHz according to Eq. (23)] is superimposed on that with driving frequency (1.1 MHz). (b) The probabilities of being in defect-free state, metastable state (with an annihilable defect), and ruptured state. (c) The distortion amplitude (n=2) of a microbubble relative to the mean bubble radius. It exceeds 1 at t=4.2 μs.

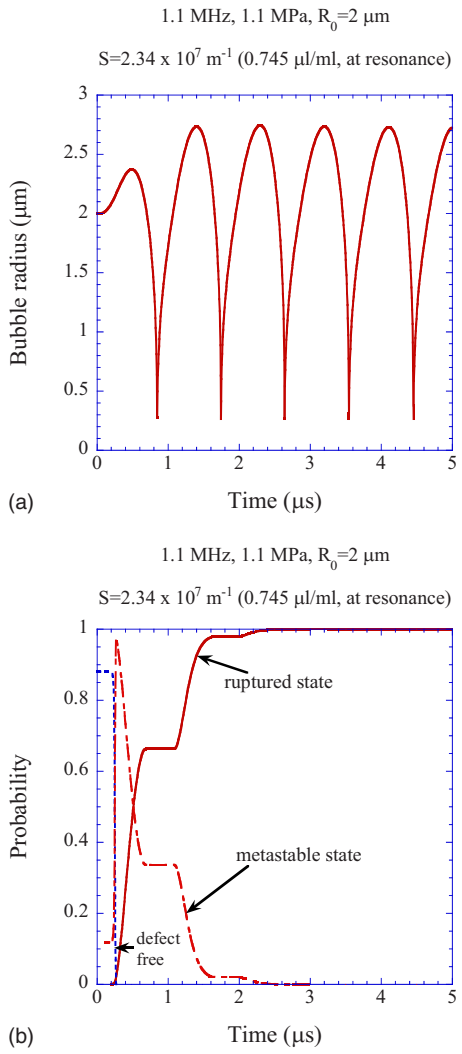


FIG. 5. (Color online) The result of the numerical simulation when the natural frequency [1.1 MHz according to Eq. (23)] coincides with the driving frequency (1.1 MHz). The coupling strength of the bubble-bubble interaction is 2.34×10^7 m^{-1}, which may correspond to the microbubble concentration of 0.745 μl/ml (Ref. 33). The pressure amplitude of ultrasound is 1.1 MPa, which is the same as that in Figs. 4 and 6. The time axes are the same for (a) and (b) for 5 μs which corresponds to 5.5 acoustic cycles. (a) The radius of an Albunex microbubble. (b) The probabilities of being in defect-free state, metastable state (with an annihilable defect), and ruptured state.

shown as a function of time with the same time axis as those of Figs. 4(a) and 4(b). It is seen that it exceeds 1 at t =4.2 μs when a microbubble may disintegrate into daughter bubbles. On the other hand, the probability of being in ruptured state at the time (t=4.2 μs) is 0.967. It means that the destruction of a microbubble in this case is mostly due to the rupture of the shell (A in Fig. 1) and that the destruction due to fragmentation has a very low probability of occurrence (0.033=1−0.967).

According to Eq. (23), the natural (resonance) frequency coincides with the driving frequency of 1.1 MHz when S =2.34×10^7 m^{-1}. In Fig. 5, the calculated result for that value of the coupling strength has been shown. In Fig. 5(a), the radius of an Albunex microbubble has been shown as a function of time for 5 μs. It is seen that the maximum radius at the expansion is much larger than that for an isolated

microbubble because it is at resonance. In contrast to the case of Fig. 4(a), the pulsation with the natural frequency and that with the driving frequency are not separated. In Fig. 5(b), the probabilities of being in defect-free state, metastable state, and ruptured state have been shown as a function of time with the same time axis as that of Fig. 5(a). Although it is not shown here, the amplitude of the shape oscillation decreases with time and a microbubble is shape stable in this case. Thus the destruction of a microbubble in this case is solely due to the rupture of the shell. The probability of being in ruptured state is 0.979 at $t=2$ μs and 0.9999 at $t=5$ μs.

According to the experiment by Klibanov *et al.*,[34] in a microbubble (Albunex) cloud, microbubbles at the border of the cloud were first destructed. At the border of the microbubble cloud, the coupling strength is roughly a half of the value at the center given by Eq. (1). Thus the value of $S=2.34\times10^{7}$ m^{-1} may correspond to the number density of microbubbles of 2.98×10^{5} bubbles/ml or 0.745 μl/ml for $l_{\max}=0.5$ cm.[33]

Typical microbubble concentration was 30 μl/ml in the experiment of Chang *et al.*,[33] which may correspond to about 1.2×10^{7} bubbles/ml. Thus, at the border of the microbubble cloud, $S=9.42\times10^{8}$ m^{-1}. In Fig. 6, the calculated result with this coupling strength has been shown when the ultrasonic frequency and the pressure amplitude are the same as those in Figs. 4 and 5. In Fig. 6(a), the radius of an Albunex microbubble has been shown as a function of time for 40 μs, which is much longer than that in Figs. 4(a) and 5(a). The natural frequency at this value of the coupling strength is about 170 kHz whose period is 5.9 μs according to Eq. (23). As the linear approximation is not strictly valid at the acoustic pressure of 1.1 MPa, the period of pulsation with natural frequency seen in Fig. 6(a) is longer than that estimated by Eq. (23). The pulsation with the driving frequency is superimposed on the pulsation with the natural frequency in contrast to the case of Fig. 4(a). In Fig. 6(b), the probabilities of being in defect-free state, metastable state, and ruptured state have been shown as a function of time with the same time axis as that of Fig. 6(a). The probability of being in ruptured state increases to 9.29×10^{-3} in 40 μs. It means that the former term in the right hand side of Eq. (11) is 1.08 s. On the other hand, the latter term is 1.04 s as the decrease in S_0 in 40 μs is 5.82×10^{-3}, that during the pulse-off time (for 9960 μs) is 2.67×10^{-3}, and the initial value of S_0 at $t=0$ is 0.8808 according to Eq. (7). On the other hand, a microbubble never disintegrates into daughter bubbles in the case of Fig. 6. It means that the destruction may be solely due to the rupture of the shell due to surface tension in this case. Thus, according to the present numerical simulation, the acoustic amplitude of 1.1 MPa in Fig. 6 may be near the threshold one for destruction because $\tau\approx1$ s.

In Fig. 7, the calculated result for an isolated Albunex microbubble has been shown when the pressure amplitude of ultrasound is 0.05 MPa. In Fig. 7(a), the radius of a microbubble has been shown as a function of time for 5 μs. It is seen that the pulsation with the natural frequency [7.3 MHz according to Eq. (23)] is superimposed on the pulsation with the driving frequency (1.1 MHz). The pulsation

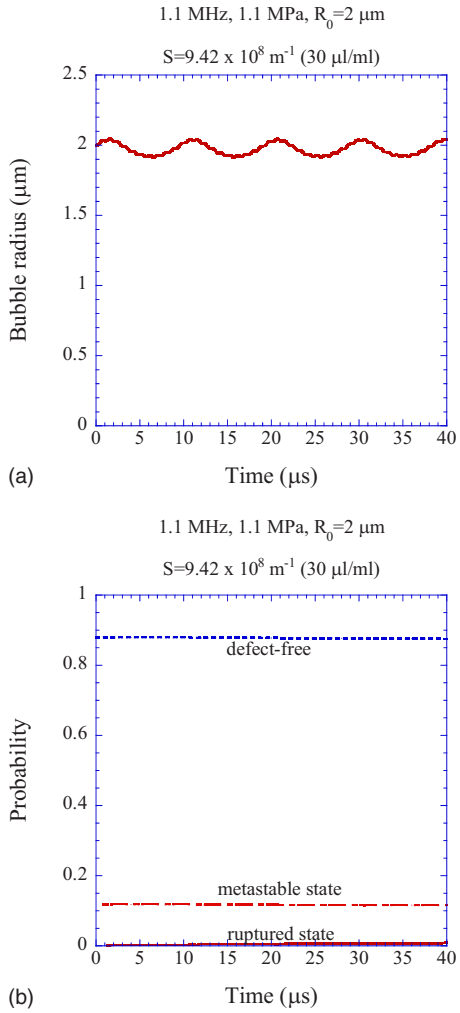


FIG. 6. (Color online) The result of the numerical simulation when the coupling strength is 9.42×10^{8} m^{-1}, which may correspond to the microbubble concentration of 30 μl/ml (Ref. 33). The frequency and the pressure amplitude of ultrasound are 1.1 MHz and 1.1 MPa, respectively, which are the same as those in Figs. 4 and 5. The time axes are the same for (a) and (b) for 40 μs which corresponds to 44 acoustic cycles. (a) The radius of an Albunex microbubble. The pulsation with the driving frequency (1.1 MHz) is superimposed on that with the natural frequency [170 kHz according to Eq. (23)]. (b) The probabilities of being in defect-free state, metastable state (with an annihilable defect), and ruptured state.

is much milder than that at 1.1 MPa in pressure amplitude shown in Fig. 4(a). In Fig. 7(b), the amplitude of the shape oscillation ($n=3$ mode) relative to the instantaneous bubble radius has been shown as a function of time for 10 μs which is longer than that in (a). It exceeds 1 at $t=9.2$ μs when the fragmentation of a microbubble may take place. At the time ($t=9.2$ μs), the probability of being in ruptured state is very low (2×10^{-4}). Thus, the destruction of a microbubble in this case may be mostly due to its fragmentation caused by its shape instability (B in Fig. 1).

Numerical simulations of destruction of microbubbles have been performed for various concentrations of microbubbles when the ultrasonic frequency is 1.1 MHz (Fig. 8). For most cases, the destruction is due to the rupture of the shell caused by surface tension mainly during the bubble expansion (A in Fig. 1) rather than the fragmentation (B in Fig. 1) according to the present numerical simulations. It is

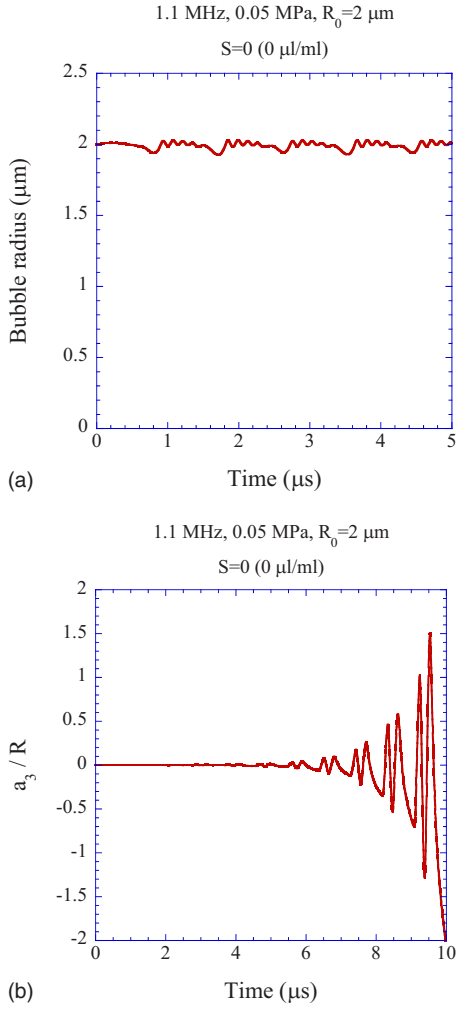


FIG. 7. (Color online) The result of the numerical simulation for an isolated Albunex microbubble when the frequency and pressure amplitude of ultrasound are 1.1 MHz and 0.05 MPa, respectively. (a) The radius of a microbubble as a function of time for 5 μs. The pulsation with natural frequency [7.3 MHz according to Eq. (23)] is superimposed on that with the driving frequency (1.1 MHz). (b) The distortion amplitude ($n=3$) of a microbubble relative to the mean bubble radius as a function of time for 10 μs. It exceeds 1 at $t=9.2$ μs. The time axis is different from that of (a).

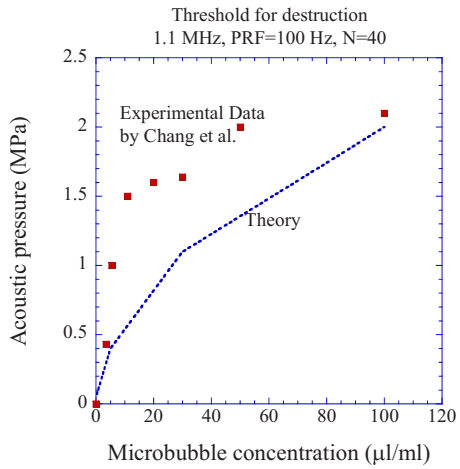


FIG. 8. (Color online) The threshold acoustic pressure for destruction of encapsulated microbubbles as a function of the microbubble concentration. The squares are the experimental data for Albunex microbubbles reported by Chang *et al.* (Ref. 33). The dotted line is the result of the numerical simulations. The ultrasonic frequency is 1.1 MHz. In the experiment, the number of acoustic cycles per pulse was 40 and the PRF was 100 Hz. The acoustic pressure was increased by approximately 135 kPa at every second to determine the threshold in the experiment.

consistent with many experimental observations of a microbubble with a stiff shell.[43–45,47] In Fig. 8, the results of the numerical simulations and the experimental data have been compared with regard to the threshold acoustic pressure for destruction as a function of the microbubble concentration. The results of the numerical simulations qualitatively agree with the experimental data. As the microbubble concentration increases, the bubble pulsation becomes milder due to the stronger bubble-bubble interaction and the threshold for destruction increases. However, there is a marked quantitative difference between the calculated results and the experimental data especially for the microbubble concentrations of 10–60 μl/ml. It may be due to the assumption in the numerical simulations that microbubbles are spatially uniformly distributed. In actual experiments, some microbubbles may aggregate each other and the local concentration may be much higher than the averaged one.[35] Thus, for more detailed discussions, the effect of inhomogeneous distribution of microbubbles should be taken into account in the numerical simulations. It means that the coupling strength (S), which is the most important parameter in predicting the strength of the bubble-bubble interaction, should be directly estimated by its definition ($S=\Sigma_i 1/d_i$, where d_i is the distance between the bubble and a bubble numbered i, and the summation is for all the surrounding bubbles). Furthermore, the effect of time delays due to the finite speed propagation of acoustic waves radiated from surrounding bubbles should be taken into account.[23,31,64] The distribution of ambient bubble radius may also be important.[65]

With regard to the attenuation of ultrasound by microbubbles, it may be negligible at the border of the microbubble cloud. However, for more detailed discussions, it should also be taken into account. Furthermore, the exact values of the five parameters in Eqs. (3)–(5), which characterize the mechanical strength of the shell, should be experimentally determined for an Albunex microbubble.[55] Nevertheless, the qualitative conclusions in the present paper are independent of the exact values of the parameters.

IV. CONCLUSION

An analytical expression has been given for the natural frequency of a microbubble with or without the shell taking into account the bubble-bubble interaction. It has been shown that the natural frequency decreases as the microbubble concentration increases to relatively high concentrations. At some microbubble concentration, the natural frequency may coincide with the driving frequency, which results in stronger pulsation. A theoretical model of destruction of encapsulated microbubbles has been constructed taking into account the effect of the bubble-bubble interaction. There are two mechanisms in the destruction of a microbubble. One is the rupture of the shell due to surface tension (A in Fig. 1). The other is the fragmentation of a microbubble due to its shape instability (B in Fig. 1). Nu-

merical simulations of destruction of microbubbles have been performed under the condition of the experiment by Chang *et al.*[33] As the microbubble concentration increases, the pulsation of a microbubble becomes milder due to the stronger bubble-bubble interaction. It may be one of the reasons why the threshold of acoustic pressure amplitude for destruction increases as the concentration increases.

[1]*Ultrasound Contrast Agents*, edited by B. B. Goldberg, J. S. Raichlen, and F. Forsberg (Martin Dunitz, London, 2001).
[2]L. Hoff, *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer Academic, Dordrecht, 2001).
[3]T. G. Leighton, *The Acoustic Bubble* (Academic, London, 1994).
[4]R. Mettin, I. Akhatov, U. Parlitz, C. D. Ohl, and W. Lauterborn, "Bjerknes forces between small cavitation bubbles in a strong acoustic field," Phys. Rev. E **56**, 2924–2931 (1997).
[5]K. Yasui, Y. Iida, T. Tuziuti, T. Kozuka, and A. Towata, "Strongly interacting bubbles under an ultrasonic horn," Phys. Rev. E **77**, 016609 (2008).
[6]D. E. Weston, "Acoustic interaction effects in arrays of small spheres," J. Acoust. Soc. Am. **39**, 316–322 (1965).
[7]O. V. Voinov and A. M. Golovin, "Lagrange equations for a system of bubbles of varying radii in a liquid of small viscosity," Fluid Dyn. **5**, 458–464 (1970).
[8]A. Shima, "The natural frequencies of two spherical bubbles oscillating in water," Trans. ASME, J. Basic Enginrg. **93**, 426–432 (1971).
[9]G. N. Kuznetsov and I. E. Shchekin, "Interaction of pulsating bubbles in a viscous liquid," Sov. Phys. Acoust. **18**, 466–469 (1973).
[10]M. Morioka, "Theory of natural frequencies of two pulsating bubbles in infinite liquid," J. Nucl. Sci. Technol. **11**, 554–560 (1974).
[11]E. A. Zabolotskaya, "Interaction of gas bubbles in a sound field," Sov. Phys. Acoust. **30**, 365–368 (1984).
[12]S. Fujikawa and H. Takahira, "A theoretical study on the interaction between two spherical bubbles and radiated pressure waves in a liquid," Acustica **61**, 188–199 (1986).
[13]A. Kubota, H. Kato, and H. Yamaguchi, "A new modeling of cavitating flows: A numerical study of unsteady cavitation on a hydrofoil section," J. Fluid Mech. **240**, 59–96 (1992).
[14]Y. A. Ilinski and E. A. Zabolotskaya, "Cooperative radiation and scattering of acoustic waves by gas bubbles in liquids," J. Acoust. Soc. Am. **92**, 2837–2841 (1992).
[15]C. Feuillade, "Scattering from collective modes of air bubbles in water and the physical mechanism of superresonances," J. Acoust. Soc. Am. **98**, 1178–1190 (1995).
[16]A. Harkin, T. J. Kaper, and A. Nadim, "Coupled pulsation and translation of two gas bubbles in a liquid," J. Fluid Mech. **445**, 377–411 (2001).
[17]P.-Y. Hsiao, M. Devaud, and J.-C. Bacri, "Acoustic coupling between two air bubbles in water," Eur. Phys. J. E **4**, 5–10 (2001).
[18]A. J. Reddy and A. J. Szeri, "Coupled dynamics of translation and collapse of acoustically driven microbubbles," J. Acoust. Soc. Am. **112**, 1346–1352 (2002).
[19]M. Ida, "Alternative interpretation of the sign reversal of secondary Bjerknes force acting between two pulsating gas bubbles," Phys. Rev. E **67**, 056617 (2003).
[20]J. S. Allen, D. E. Kruse, P. A. Dayton, and K. W. Ferrara, "Effect of coupled oscillations on microbubble behavior," J. Acoust. Soc. Am. **114**, 1678–1690 (2003).
[21]A. A. Doinikov, "Mathematical model for collective bubble dynamics in strong ultrasound fields," J. Acoust. Soc. Am. **116**, 821–827 (2004).
[22]A. A. Doinikov, "Equations of coupled radial and translational motions of a bubble in a weakly compressible liquid," Phys. Fluids **17**, 128101 (2005).
[23]A. A. Doinikov, R. Manasseh, and A. Ooi, "Time delays in coupled multibubble systems," J. Acoust. Soc. Am. **117**, 47–50 (2005).
[24]M. F. Hamilton, Y. A. Ilinski, G. D. Meegan, and E. A. Zabolotskaya, "Interaction of bubbles in a cluster near a rigid surface," ARLO **6**, 207–213 (2005).
[25]E. M. B. Payne, S. J. Illesinghe, A. Ooi, and R. Manasseh, "Symmetric mode resonance of bubbles attached to a rigid boundary," J. Acoust. Soc. Am. **118**, 2841–2849 (2005).
[26]M. Ida, "Phase properties and interaction force of acoustically interacting bubbles: A complementary study of the transition frequency," Phys. Fluids **17**, 097107 (2005).
[27]M. Ida, T. Naoe, and M. Futakawa, "Suppression of cavitation inception by gas bubble injection: A numerical study focusing on bubble-bubble interaction," Phys. Rev. E **76**, 046309 (2007).
[28]Y. A. Ilinski, M. F. Hamilton, and E. A. Zabolotskaya, "Bubble interaction dynamics in Lagrangian and Hamiltonian mechanics," J. Acoust. Soc. Am. **121**, 786–795 (2007).
[29]M. Arora, C. D. Ohl, and D. Lohse, "Effect of nuclei concentration on cavitation cluster dynamics," J. Acoust. Soc. Am. **121**, 3432–3436 (2007).
[30]V. Garbin, B. Dollet, M. L. J. Overvelde, N. de Jong, D. Lohse, M. Versluis, D. Cojoc, E. Ferrari, and E. D. Fabrizio, "Coupled dynamics of an isolated UCA microbubble pair," in 2007 IEEE Ultrasonics Symposium Proceedings (2007), Vol. **2**, pp. 757–760.
[31]A. Ooi, A. Nikolovska, and R. Manasseh, "Analysis of time delay effects on a linear bubble chain system," J. Acoust. Soc. Am. **124**, 815–826 (2008).
[32]R. Manasseh and A. Ooi, "The frequencies of acoustically interacting bubbles," Bubble Sci. Engnrng. Technol. **1**, 58–74 (2009).
[33]P. P. Chang, W. S. Chen, P. D. Mourad, S. L. Poliachik, and L. Crum, "Threshold for inertial cavitation in Albunex suspensions under pulsed ultrasound conditions," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 161–170 (2001).
[34]A. L. Klibanov, K. W. Ferrara, M. S. Hughes, J. H. Wible, J. K. Wojdyla, P. A. Dayton, K. E. Morgan, and G. H. Brandenburger, "Direct videomicroscopic observation of the dynamic effects of medical ultrasound on ultrasound contrast microspheres," Invest. Radiol. **12**, 863–870 (1998).
[35]P. A. Dayton, K. E. Morgan, A. L. Klibanov, G. H. Brandenburger, and K. W. Ferrara, "Optical and acoustical observations of the effects of ultrasound on contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 220–232 (1999).
[36]C. M. Moran, T. Anderson, S. D. Pye, V. Sboros, and W. N. McDicken, "Quantification of microbubble destruction of three fluorocarbon-filled ultrasonic contrast agents," Ultrasound Med. Biol. **26**, 629–639 (2000).
[37]W. T. Shi, F. Forsberg, A. Tornes, J. Ostensen, and B. B. Goldberg, "Destruction of contrast microbubbles and the association with inertial cavitation," Ultrasound Med. Biol. **26**, 1009–1019 (2000).
[38]J. E. Chomas, P. A. Dayton, D. May, J. Allen, A. Klibanov, and K. Ferrara, "Optical observation of contrast agent destruction," Appl. Phys. Lett. **77**, 1056–1058 (2000).
[39]J. E. Chomas, P. Dayton, J. Allen, K. Morgan, and K. W. Ferrara, "Mechanism of contrast agent destruction," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 232–248 (2001).
[40]J. E. Chomas, P. Dayton, D. May, and K. Ferrara, "Threshold of fragmentation for ultrasonic contrast agents," J. Biomed. Opt. **6**, 141–150 (2001).
[41]D. J. May, J. S. Allen, and K. W. Ferrara, "Dynamics and fragmentation of thick-shelled microbubbles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1400–1410 (2002).
[42]W. S. Chen, T. J. Matula, A. A. Brayman, and L. A. Crum, "A comparison of the fragmentation thresholds and inertial cavitation doses of different ultrasound contrast agents," J. Acoust. Soc. Am. **113**, 643–651 (2003).
[43]S. H. Bloch, M. Wan, P. A. Dayton, and K. W. Ferrara, "Optical observation of lipid-and polymer-shelled ultrasound microbubble contrast agents," Appl. Phys. Lett. **84**, 631–633 (2004).
[44]D. Koyama, W. Kiyan, and Y. Watanabe, "Optical observation of microcapsule destruction in an acoustic standing wave," Jpn. J. Appl. Phys., Part 1 **43**, 3215–3219 (2004).
[45]A. Boukaz, M. Versluis, and N. de Jong, "High-speed optical observation of contrast agent destruction," Ultrasound Med. Biol. **31**, 391–399 (2005).
[46]A. Y. Ammi, R. O. Cleveland, J. Mamou, G. I. Wang, S. L. Bridal, and W. D. O'Brien, "Ultrasonic contrast agent shell rupture detected by inertial cavitation and rebound signals," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 126–135 (2006).
[47]D. Koyama, A. Osaki, W. Kiyan, and Y. Watanabe, "Acoustic destruction of a microcapsule having a hard plastic shell," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 1314–1321 (2006).
[48]T. M. Porter, D. A. Smith, and C. K. Holland, "Acoustic techniques for assessing the Optison destruction threshold," J. Ultrasound Med. **25**, 1519–1529 (2006).
[49]C. Pecorari and D. Grishenkov, "Characterization of ultrasound-induced fracture of polymer-shelled ultrasonic contrast agents by correlation analysis," J. Acoust. Soc. Am. **122**, 2425–2430 (2007).
[50]S. Casciaro, R. P. Errico, F. Conversano, C. Demitri, and A. Distante, "Experimental investigations of nonlinearities and destruction mechanisms of an experimental phospholipids-based ultrasound contrast agent," Invest. Radiol. **42**, 95–104 (2007).

[51]D. A. B. Smith, T. M. Porter, J. Martinez, S. Huang, R. C. MacDonald, D. D. McPherson, and C. K. Holland, "Destruction thresholds of echogenic liposomes with clinical diagnostic ultrasound," Ultrasound Med. Biol. **33**, 797–809 (2007).
[52]C. K. Yeh and S. Y. Su, "Effects of acoustic insonation parameters on ultrasound contrast agent destruction," Ultrasound Med. Biol. **34**, 1281–1291 (2008).
[53]E. Stride and N. Saffari, "On the destruction of microbubble ultrasound contrast agents," Ultrasound Med. Biol. **29**, 563–573 (2003).
[54]M. Postema and G. Schmitz, "Ultrasonic bubbles in medicine: Influence of the shell," Ultrason. Sonochem. **14**, 438–444 (2007).
[55]E. Evans, V. Heinrich, F. Ludwig, and W. Rawicz, "Dynamic tension spectroscopy and strength of biomembranes," Biophys. J. **85**, 2342–2350 (2003).
[56]C. C. Church, "The effects of an elastic solid surface layer on the radial pulsations of gas bubbles," J. Acoust. Soc. Am. **97**, 1510–1521 (1995).
[57]J. Kragel, S. Siegel, R. Miller, M. Born, and K.-H. Schano, "Measurement of interfacial shear rheological properties: An automated apparatus," Colloids Surf., A **91**, 169–180 (1994).
[58]R. Dimova, B. Pouligny, and C. Dietrich, "Pretransitional effects in dimyristoylphosphatidylcholine vesicle membranes: Optical dynamometry study," Biophys. J. **79**, 340–356 (2000).
[59]P. Marmottant, S. van der Meer, M. Emmer, M. Versluis, N. de Jong, S. Hilgenfeldt, and D. Lohse, "A model for large amplitude oscillations of coated bubbles accounting for buckling and rupture," J. Acoust. Soc. Am. **118**, 3499–3505 (2005).
[60]K. Yasui, "Alternative model of single-bubble sonoluminescence," Phys. Rev. E **56**, 6750–6760 (1997).
[61]N. de Jong, R. Cornet, and C. Lancee, "Higher harmonics of vibration gas-filled microspheres. Part one: Simulations," Ultrasonics **32**, 447–453 (1994).
[62]S. Hilgenfeldt, D. Lohse, and M. P. Brenner, "Phase diagrams for sonoluminescing bubbles," Phys. Fluids **8**, 2808–2826 (1996).
[63]K. Yasui, "Influence of ultrasonic frequency on multibubble sonoluminescence," J. Acoust. Soc. Am. **112**, 1405–1413 (2002).
[64]C. Vanhille and C. Campos-Pozuelo, "Nonlinear ultrasonic propagation in bubbly liquids: A numerical model," Ultrasound Med. Biol. **34**, 792–808 (2008).
[65]K. Yasui, T. Tuziuti, J. Lee, T. Kozuka, A. Towata, and Y. Iida, "The range of ambient radius for an active bubble in sonoluminescence and sonochemical reactions," J. Chem. Phys. **128**, 184705 (2008).

A generalized statistical Burgers equation to predict the evolution of the power spectral density of high-intensity noise in atmosphere

Penelope Menounou and Aristotelis N. Athanasiadis
Department of Mechanical and Aeronautical Engineering, University of Patras, Patras 26504, Greece

(Received 6 November 2008; revised 6 June 2009; accepted 11 June 2009)

The present work is a theoretical/numerical investigation of the combined effect of nonlinearity, geometrical spreading, and atmospheric absorption on the evolution of the power spectral density of a noise field, when only the power spectral density is known at source, not the signal itself. This is often the case in aircraft noise measurements. The method presented here is based on and extends previous work [P. Menounou and D. T. Blackstock, J. Acoust. Soc. Am. **115**, 567–580 (2004)], where a recursion equation [statistical Burgers equation (SBE)] describing the evolution of the joint moments of the noise source was derived. The SBE is restricted to plane waves, thermoviscous fluids, and short propagation distances (preshock region). In the present work, the SBE is extended to include the effects of geometrical spreading and arbitrary absorption, in order to be applicable to propagation of high-intensity noise through atmosphere. A new equation is derived and termed generalized SBE, and a method for its numerical implementation is presented. Results are in good agreement with time domain calculations for propagation in atmosphere of (i) sinusoidal signals (benchmark case) and (ii) Gaussian processes with known power spectral densities at source.
 [DOI: 10.1121/1.3167393]

PACS number(s): 43.28.Bj, 43.25.Cb [LCS] Pages: 983–994

I. INTRODUCTION

The aim of the present work is to contribute to the practical problem of predicting the evolution of the power spectral density (PSD) of the noise field measured close to an aircraft at a point away from it. Aircraft, mainly through their jet engines, produce high-intensity noise, which undergoes nonlinear propagation distortion, and, accordingly, energy is transferred to the high frequency end of the spectrum.[1–6] The prediction of finite-amplitude noise propagation is in most cases based on "time domain" methods. The term time domain methods is used in a broad sense indicating that the information at source is equivalent to information provided by the pressure time waveform at source. The actual numerical prediction of noise propagation can be performed either in the time domain,[7] in the frequency domain,[8] switching between time and frequency domains,[9] or employing a computational fluid dynamics/computational aeroacoustics approach.[10–12] The references provided for each approach are only indicative. The reader is referred to Ref. 13 for a detailed presentation of published work. In many cases, however, time domain methods cannot be employed. This happens, if only the PSD is given at source, not the signal itself. Time domain methods cannot be employed, as the PSD does not contain any phase information and a unique time signal cannot be determined by the PSD.

The present work is a theoretical/numerical investigation of the combined effect of nonlinear propagation distortion, spherical spreading, and atmospheric absorption directly on the PSD of a source condition with Gaussian characteristics (that is, a Gaussian noise signal or a Gaussian stationary and ergodic stochastic process). The method presented here is based on and extends previous work of one of the authors. Details on the previous work, as well as on its categorization with respect to existing methods, are included in Ref. 14. A short outline is provided in the following as an introduction to the work presented here.

A given time waveform has a unique PSD, but a given PSD does not correspond to a unique time waveform, because the phase information is missing from the PSD. One way to tackle the problem is to create a great number of time waveforms that have the same PSD at source but different phases, predict their evolution using a time domain approach (for example, by solving the Burgers equation), and then average the resulting PSDs.[15] The second approach is to work directly with the PSD or its Fourier counterpart, the autocorrelation function. The majority of these methods are based on the so-called closure hypothesis; see, for example, Ref. 16. It is known that nonlinear equations in the time domain can be transformed into an infinite number of nested linear equations, in which the unknowns are the autocorrelation function (or PSD) and higher order joint moments (or higher order joint spectra).[17] The evolution of the autocorrelation function (a second-order joint moment) depends on the evolution of third-order joint moments, which in turn depend on fourth-order joint moments, thus, leaving the system of the nested equations open. Various closure hypotheses have been provided to close the system. In acoustical applications, the so-called quasinormal hypothesis[18,19] has been used. However, this can result in negative PSDs, which by definition are positive quantities.

Previous work[14] directly on the PSD domain resulted in a recursion equation describing the evolution of the joint moments, which can, however, be solved numerically without any closure hypothesis. Because it resulted from the sta-

tistical averaging of the Burgers equation (BE), it was termed statistical Burgers equation (SBE). Both BE and SBE are repeated here for completeness following, however, the notation that will be used in the present paper. (Symbols bearing a tilde ~ indicate dimensional quantities, and their counterparts without the tilde are the corresponding dimensionless quantities. The Latin letters s, f, t, and y denote the dimensional distance, frequency, time, and time lag, respectively, while the Greek letters σ, θ, τ, and ψ are their dimensionless counterparts.) Employing this notation, the BE can be written as

$$\text{BE:}\frac{\partial \tilde{p}}{\partial s} = \mathcal{N}\frac{\beta}{\rho_0 c_0^3}\tilde{p}\frac{\partial \tilde{p}}{\partial t} + \mathcal{A}\frac{\delta}{2c_0^3}\frac{\partial^2 \tilde{p}}{\partial t^2}, \tag{1}$$

where $\tilde{p}$ is the sound pressure, s is the range variable, t is the retarded time associated with the outgoing wave, ρ_0 is the ambient density, c_0 is the small signal sound speed, β is the coefficient of nonlinearity, and δ is the diffusivity of sound for viscosity and heat conduction. The tags $\mathcal{N}$ and $\mathcal{A}$, which can take the values of 0 or 1, mark the terms of the equation related to nonlinear effects and absorption, respectively. The SBE has the following form:

$$\text{SBE:}\frac{\partial E_{m,n}}{\partial \sigma} = \mathcal{N}\frac{\partial}{\partial \psi}\left[-\frac{m}{m+1}E_{m+1,n} + \frac{n}{n+1}E_{m,n+1}\right] + 2\mathcal{A}A_{tv}\frac{\partial^2 E_{m,n}}{\partial \psi^2}. \tag{2}$$

The normalized joint moment $E_{m,n}(\sigma,\psi)$ is defined for a stochastic process as the ensemble average and for a deterministic signal as the time average as follows:

$$E_{m,n}(\psi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p_1^m p_2^n W_{p_1p_2}(p_1,p_2;\sigma,\psi)dp_1dp_2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T p^m(\tau)p^n(\tau+\psi)d\tau, \tag{3}$$

where $p_1=p(\sigma,\tau)$ is the normalized pressure at time τ, $p_2=p(\sigma,\tau+\psi)$ is pressure p_1 delayed by time ψ, and $W_{p_1p_2}(p_1,p_2;\sigma,\psi)$ is the second-order joint probability density function at ψ. For a stationary and ergodic stochastic process, the ensemble average can be substituted by the time average. The autocorrelation function is the joint moment $E_{1,1}$. The coefficient $A_{tv}=\delta\omega_0\rho_0/2\beta p_0$ describes the relative importance between nonlinearity and thermoviscous attenuation ($1/A_{tv}$ is often called Gol'dberg number).

If all the joint moments up to order $m+n=N$ are known at source, and the computation of the joint moments starts from the higher orders to the lower, then the autocorrelation function can be computed at each propagation step $j=1,\ldots,N-1$, as shown in Fig. 1. The joint moments at source can be computed (i) if the time history of the signal is given; (ii) if the source signal is a Gaussian noise with known PSD (or equivalently $\tilde{E}_{1,1}$); or (iii) if the source condition is a Gaussian, stationary, and ergodic stochastic process with known PSD. The SBE describes propagation of plane waves in thermoviscous fluids. In previous work[14] results obtained by the SBE were shown to be in good agreement with analytical solution in the pre-shock region. Results are presented as joint moments (rather than PSDs), as the joint moments are the native quantities of the method. Results are usually presented as $\tilde{E}_{1,1}$, which is equivalent to PSD. One of the advantages of expressing results as $\tilde{E}_{1,1}$ is that the root mean-square value of the signal, and thus its intensity, is provided directly by the value of $\tilde{E}_{1,1}(0)$ [see Eq. (3)]. For details on the manifestation of nonlinear and dissipation effects on joint moments the reader is directed to Ref. 14.

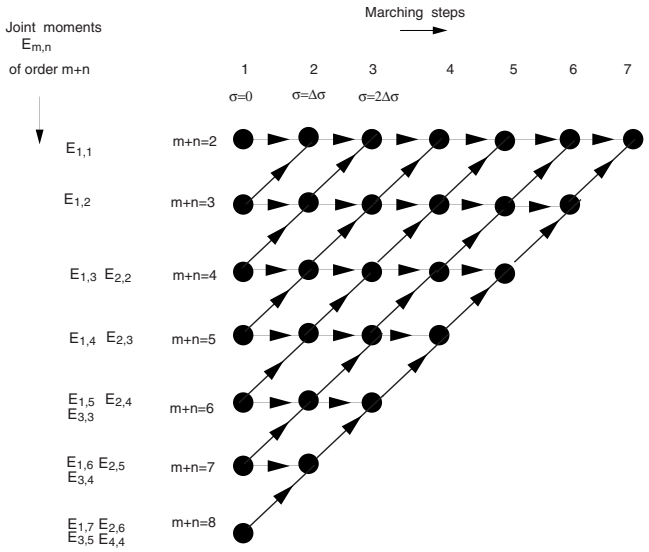


FIG. 1. Graphical illustration of the interaction between the joint moments $E_{m,n}$. At each marching step lower order moments can be computed from higher order moments known at the previous marching step. The autocorrelation function $E_{1,1}$ can be computed for $N-1$ marching steps, if joint moments up to order N are known at source.

In the present work, the SBE is extended to include the effects of geometrical spreading and arbitrary absorption, in order to be applicable to propagation of high-intensity noise through real atmosphere. The resulting equation, termed generalized statistical Burgers equation (gSBE), is presented in Sec. II, and its numerical implementation is outlined in Sec. III. Results are presented in Sec. IV. As benchmark cases, results obtained by the gSBE are compared with time domain calculations for a sinusoidal source signal. Finally, the method is applied for Gaussian processes with known PSDs at source.

II. GENERALIZED STATISTICAL BURGERS EQUATION (gSBE)

In order to model the propagation of noise through real atmosphere, the SBE [Eq. (2)] must be enhanced to include the effects of geometrical spreading and atmospheric absorption. The starting point is the following form of the generalized Burgers equation (gBE) (Refs. 13 and 20):

$$g\text{BE:}\frac{\partial \tilde{p}}{\partial s} + \frac{\mathcal{G}}{s}\tilde{p} = \mathcal{N}\frac{\beta}{\rho_0 c_0^3}\tilde{p}\frac{\partial \tilde{p}}{\partial t} + \mathcal{A}A_t(\tilde{p}), \tag{4}$$

where $t=t'-(s-s_0)/c_0$ is the retarded time (defined for diverging waves) with t' being the time and s_0 the radius of the source. The parameter $\mathcal{G}$ takes the values of 0, 1/2, and 1 for

plane, cylindrically, and spherically spreading waves, respectively. The operator $\mathcal{A}_t$ represents atmospheric absorption and dispersion that acts on pressure $\tilde{p}$. It will be shown that the gBE [Eq. (4)] can be transformed into the following gSBE:

$$\text{gSBE:}\frac{\partial E_{m,n}}{\partial \sigma} = \mathcal{N}\frac{\partial}{\partial \psi}\left[-\frac{m}{m+1}E_{m+1,n} + \frac{n}{n+1}E_{m,n+1}\right] - \mathcal{G}\frac{m+n}{\sigma}E_{m,n} + \mathcal{A}A_{\psi}(E_{m,n}), \tag{5}$$

where $E_{m,n}$ are the joint moments in their non-dimensional form and A_{ψ} is an operator accounting for atmospheric absorption that acts on $E_{m,n}$.

A. Geometrical spreading

In order to incorporate the effect of geometrical spreading, the gBE [Eq. (4)] for an ideal (non-absorbing and non-dispersive) fluid ($A_t=0$) is considered. The following dimensionless notation is introduced:

$$p = \frac{\tilde{p}}{p_0}, \quad \tau = \omega_0 t, \quad \sigma = \frac{s}{\bar{s}}, \quad \bar{s} = \frac{\rho_0 c_0^3}{\beta p_0 \omega_0},$$

where p_0 is a characteristic pressure amplitude, ω_0 is a characteristic frequency, and $\bar{s}$ is the shock formation distance for a plane wave in a non-absorbing fluid. The characteristic pressure p_0 is chosen so that the source condition has unit mean square, in other words, $p_0=\sqrt{\tilde{E}_{1,1}(0,0)}$. The characteristic frequency is either the frequency carrying the maximum energy at source or the inverse correlation time of the autocorrelation function at source $\tilde{E}_{1,1}(0,y)$. The non-dimensionless radius of the source s_0 is $\sigma_0=s_0/\bar{s}$. In terms of the dimensionless notation, Eq. (4) for ($A_t=0$) becomes

$$\frac{\partial p}{\partial \sigma} + \frac{\mathcal{G}}{\sigma}p = \mathcal{N}p\frac{\partial p}{\partial \tau}, \quad \mathcal{N}=0,1, \quad \mathcal{G}=0,1/2,1, \tag{6}$$

where the constant parameters $\mathcal{N}$ and $\mathcal{G}$ tag the terms describing the geometrical spreading and nonlinear propagation distortion, respectively.

By following the procedure outlined in Ref. 14 the statistical averaging of Eq. (6) yields the following recursive equation for the joint moment $E_{m,n}$:

$$\frac{\partial}{\partial \sigma}E_{m,n} = \mathcal{N}\frac{\partial}{\partial \psi}\left[-\frac{m}{m+1}E_{m+1,n} + \frac{n}{n+1}E_{m,n+1}\right] - \mathcal{G}\frac{m+n}{\sigma}E_{m,n}. \tag{7}$$

Similar to the SBE, the term associated with nonlinearity ($\mathcal{N}$) involves joint moments of higher order $m+n+1$. The geometrical spreading term ($\mathcal{G}$) involves joint moments of the same order as the linear propagation term on the left hand side of Eq. (7). This is expected, as geometrical spreading is a linear mechanism. Finally, it should be noted that geometrical spreading affects the higher order joint moments more than the lower ones, by reducing their amplitude faster.

B. Atmospheric absorption

A hybrid approach is chosen for the inclusion of atmospheric absorption into the gSBE, similar to the Pestorius algorithm[9] also employed by Gee *et al.*[5] in a more recent work on the gBE. In these studies, the nonlinear distortion is computed in the time domain, while the absorption in the frequency domain. Implementation of absorption in the time domain has certain restrictions regarding the types of absorption that can be handled. In the frequency domain, however, arbitrary absorption and dispersion can be introduced. Correspondingly, in the present work, nonlinear distortion and spherical spreading are implemented in the joint moments' domain ($E_{m,n}$) via Eq. (7), while absorption in the joint power spectra domain ($S_{m,n}$). It is recalled that joint power spectrum $S_{m,n}$ and joint moment $E_{m,n}$, or their dimensional counterparts $\tilde{S}_{m,n}$ and $\tilde{E}_{m,n}$, constitute Fourier transform pairs. The main advantage of the hybrid approach is that the method presented here can be applied not only for propagation in the atmosphere, but also within any propagation medium, which is weakly (or non-) dispersive and has a known absorption relation. Furthermore, absorption is applied to the dimensional joint power spectra domain $\tilde{S}_{m,n}$ rather than the dimensionless $S_{m,n}$, to further accommodate arbitrary absorption relations.

Consider an atmosphere that is homogeneous and at rest. The amplitude of a monochromatic plane wave that travels distance s in the atmosphere decreases from $\tilde{p}_1=P_1e^{j\omega t}$ to $\tilde{p}_2=P_2e^{j\omega t}$ as follows:

$$\tilde{p}_2 = \tilde{P}_2e^{j\omega t} = \tilde{P}_1e^{-\alpha s}e^{-j\gamma s}e^{j\omega t}, \tag{8}$$

where α is the absorption coefficient and is associated with amplitude attenuation, while γ is related to the phase speed $c^{\text{ph}}=\omega/\gamma$ and is associated with dispersion. Atmosphere is weakly dispersive and dispersion will be presently ignored. The absorption coefficient (in Np/m) depends on temperature, atmospheric pressure, and relative humidity and is computed using semi-empirical relations based on the work of Bass *et al.*[21–23]

In cases of spherically spreading waves, the amplitude of the pressure $\tilde{P}_1e^{-\alpha s}$ is further reduced to $\tilde{P}_1e^{-\alpha s}/s$ to account for the geometrical spreading. Geometrical spreading can therefore be dealt with together with atmospheric absorption, that is, in the joint power spectra domain. In the present method, however, the effect of geometrical spreading is implemented in the joint moments' domain (see Sec. II A) and should not be replicated in the joint power spectra domain.

The application of the above in the joint power spectra domain is outlined next. It can be shown that the joint power spectrum $\tilde{S}_{m,n}(f)$ and the spectrum $\tilde{P}(f)$ of an acoustic signal $\tilde{p}(t)$ are related as follows:

$$\tilde{S}_{m,n}(f) = \frac{\overline{\tilde{P}^m(f)}\tilde{P}^n(f)}{2}, \tag{9}$$

where the overbar denotes the complex conjugate. Consider a sinusoidal signal of frequency f_0 [$\tilde{P}(f_0)$]. The corresponding joint power spectrum $\tilde{S}_{m,n}(f)$ can be written as

$$\tilde{S}_{m,n}(f) = \frac{\overline{\tilde{P}^m(f_0)}\tilde{P}^n(f_0)}{2}, \tag{10}$$

where $\tilde{S}_{m,n}(f)$ contains the frequencies $f_0, 3f_0, 5f_0, \ldots, qf_0$ (where $q=\min\{m,n\}$), if q is odd, and the frequencies $f_0, 2f_0, 4f_0, \ldots, qf_0$, if q is even. It is noted that although only the frequency f_0 exists in the spectrum $\tilde{P}$ of the signal (hereinafter called *energy carrying* or *energy-containing frequency*), additional frequencies (hereinafter called *cross-frequencies*) appear in the joint power spectra $\tilde{S}_{m,n}(f)$. In other words, in the joint spectra $\tilde{S}_{m,n}(f)$ appear frequencies that do not exist in the signal. After propagation distance s in the atmosphere, the spectrum $\tilde{P}(f_0)$ becomes $\tilde{P}^{(\alpha)}(f_0) = \tilde{P}(f_0)e^{-\alpha(f_0)s}$, where in the following the superscript $^{(\alpha)}$ denotes quantities that have been corrected for atmospheric absorption $(\tilde{P}^{(\alpha)}, \tilde{S}^{(\alpha)}_{m,n}, \tilde{E}^{(\alpha)}_{m,n}, \ldots)$. The corrected for atmospheric absorption joint power spectrum $(\tilde{S}^{(\alpha)}_{m,n})$ of signal $\tilde{P}^{(\alpha)}$ is obtained by substituting $\tilde{P}^{(\alpha)}(f_0) = \tilde{P}(f_0)e^{-\alpha(f_0)s}$ into Eq. (10) as follows:

$$\begin{aligned}\tilde{S}^{(\alpha)}_{m,n}(f) &= \frac{\overline{[\tilde{P}^{(\alpha)}(f_0)]^m}[\tilde{P}^{(\alpha)}(f_0)]^n}{2} \\ &= \frac{\overline{\tilde{P}^m(f_0)}e^{-\alpha(f_0)sm}\tilde{P}^n(f_0)e^{-\alpha(f_0)sn}}{2} \\ &= e^{-(m+n)\alpha(f_0)s}\tilde{S}_{m,n}(f).\end{aligned} \tag{11}$$

In the equation above the same absorption coefficient, $\alpha(f_0)$, as evaluated at the energy-containing frequency f_0, is applied to *all* frequencies present in $\tilde{S}_{m,n}$. Since only frequency f_0 is present in the signal and carries energy, only the atmospheric absorption at f_0 must be considered. The cross-frequencies $(2f_0, 3f_0, \ldots)$ do not attenuate under the same physical mechanism. Therefore, applying the absorption coefficients $\alpha(2f_0), \alpha(3f_0), \ldots$ would not have a physical meaning. It should be noted, however, that the $\tilde{S}_{m,n}$ values at the cross-frequencies are reduced in amplitude, because they are generated by reduced values at f_0.

The previous result is now extended to pressure signals containing energy in multiple frequencies $f_0, f_1, f_2, \ldots, f_{K-1}$. Either one of the following approaches can be employed to predict the absorption-corrected $\tilde{S}^{(\alpha)}_{m,n}$: $\tilde{S}^{(\alpha)}_{m,n,0}(f) = e^{-(m+n)\alpha(f_0)s}\tilde{S}_{m,n}(f)$, $\tilde{S}^{(\alpha)}_{m,n,1}(f) = e^{-(m+n)\alpha(f_1)s}\tilde{S}_{m,n}(f), \ldots,$ and $\tilde{S}^{(\alpha)}_{m,n,K-1}(f) = e^{-(m+n)\alpha(f_{K-1})s}\tilde{S}_{m,n}(f)$. From the K different values of $\tilde{S}^{(\alpha)}_{m,n}$ one must compute an average. The following average is proposed:

$$\tilde{S}^{(\alpha)}_{m,n}(f) = e^{-(m+n)\bar{\alpha}s}\tilde{S}_{m,n}(f), \quad \bar{\alpha} = \frac{\sum_{k=0}^{K-1}\alpha(f_k)\tilde{S}_{1,1}(f_k)}{\sum_{k=0}^{K-1}\tilde{S}_{1,1}(f_k)}, \tag{12}$$

where $\bar{\alpha}$ is an average absorption coefficient, computed from the absorption coefficient α at the energy-containing frequencies weighted by the value of $\tilde{S}_{1,1}$ at each such frequency. Other averages can also be considered. The one proposed here is entirely based on the $\tilde{S}_{1,1}$ and is particularly advantageous in its numerical implementation as it will be outlined in Sec. III. The average absorption coefficient tends to overestimate the absorption at low frequencies and to underestimate it at high frequencies. For $K=1$, Eq. (12) collapses to Eq. (11), as expected. Equation (12) is employed in the present work to predict the absorption-corrected $\tilde{S}^{(\alpha)}_{m,n}$, instead of

$$\tilde{S}^{(\alpha)}_{m,n}(f) = e^{-(m+n)\alpha(f)s}\tilde{S}_{m,n}(f), \tag{13}$$

which is correct if only energy-containing frequencies are present in $\tilde{S}_{m,n}$. Indeed, Eq. (13) with $m=n=1$ $[\tilde{S}^{(\alpha)}_{1,1}(f) = e^{-2\alpha(f)s}\tilde{S}_{1,1}(f)]$ is employed for the prediction of the absorption correction on $\tilde{S}_{1,1}$. As only energy-containing frequencies appear in $\tilde{S}_{1,1}(f)$, each frequency is attenuated in accordance to $\alpha(f)$, and, thus, no averaging procedure is required.

In order to evaluate the proposed averaging approach, a noise signal was considered that propagates linearly as a plane wave in the atmosphere. Figure 2(a) shows the signal at source and after $s=200$ m of propagation employing Eq. (8). Figures 2(b) and 2(c) show the joint moments $E_{1,1}$ and $E_{2,2}$, respectively, after the same propagation distance as computed using Eqs. (12) and (13), and by direct computation from the pressure signal shown in Fig. 2(a). In the case of $E_{2,2}$ (an example of the general case of $E_{m\neq1,n\neq1}$, where $E_{m,n}$ contains cross-frequencies in addition to the energy carrying frequencies), Eq. (13) yields results deviating largely from the correct value of $E_{2,2}$ computed from the absorbed time waveform, while the proposed averaging procedure [Eq. (12)] provides a significant improvement. For the joint mo-

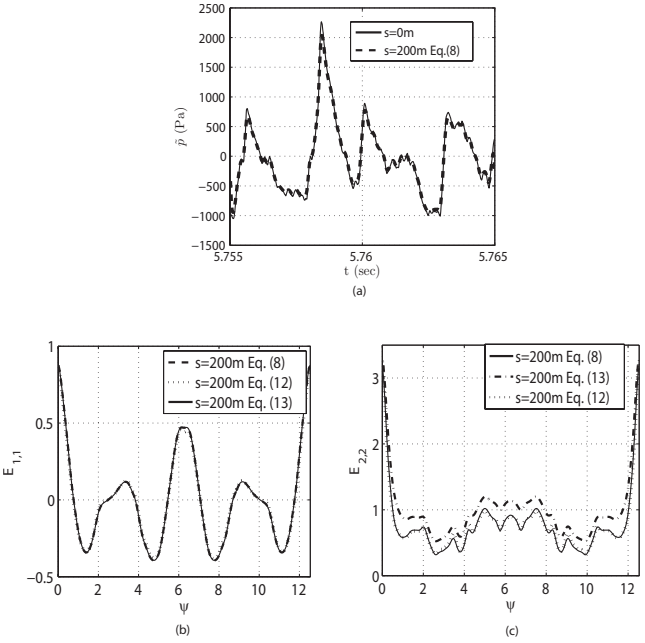


FIG. 2. Computation of the effect of atmospheric absorption on the joint moments: Time signal at source $s=0$ m and after $s=200$ m of linear propagation employing Eq. (8) (a), joint moment $E_{1,1}$ (b), and $E_{2,2}$ (c) after $s=200$ m of linear propagation employing Eqs. (12) and (13), and by direct computation of $E_{1,1}$ and $E_{2,2}$ from the pressure time signal in (a).

ment $E_{1,1}$, on the other hand, it is Eq. (13), not Eq. (12), that yields the correct result, that is, the same result as computed from the absorbed time waveform.

When incorporated into the gSBE, the absorption correction is computed at each incremental propagation step Δs as follows:

$$E_{m,n}(\psi) \to \tilde{E}_{m,n}(y) \overset{\mathcal{F}}{\to} \tilde{S}_{m,n}(f) \overset{Eq.(12)}{\longrightarrow} \tilde{S}^{(\alpha)}_{m,n}(f) \overset{\mathcal{F}^{-1}}{\to} \tilde{E}^{(\alpha)}_{m,n}(y) \to E^{(\alpha)}_{m,n}(\psi),$$

where $\mathcal{F}$ denotes Fourier transform and $\mathcal{F}^{-1}$ the inverse Fourier transform. For small propagation steps, the effect of correcting $E_{m,n}$ for absorption is numerically equivalent to incorporating absorption directly into Eq. (5). This is the same switch between time and frequency domains employed in the past.[5,9] In this case, however, the switch takes place between joint moments and joint spectra domain.

III. NUMERICAL IMPLEMENTATION

In order to develop a solution algorithm for the gSBE, a first-order Taylor expansion of $E_{m,n}(\sigma,\psi)$ is performed around $\sigma=\sigma_0+\Delta\sigma$ as follows:

$$E_{m,n}|_{\sigma_0+\Delta\sigma} = E_{m,n}|_{\sigma_0} + \Delta\sigma \left.\frac{\partial E_{m,n}}{\partial\sigma}\right|_{\sigma_0} + O(\Delta\sigma^2), \tag{14}$$

where the first derivative $\partial E_{m,n}/\partial\sigma$ is obtained from Eq. (7). Following the same procedure as in Ref. 14 one can write the discretized form of Eq. (14) as follows:

$$E^{j+1}_{m,n,i} = E^{j}_{m,n,i} + N^{j}_{m,n,i} - \mathcal{G}\left[\frac{1}{2}\frac{\Delta\sigma}{\sigma}E^{j}_{m,n,i} + \frac{1}{2}\frac{\Delta\sigma}{\sigma+\Delta\sigma}E^{j+1}_{m,n,i}\right], \tag{15}$$

$$N^{j}_{m,n,i} = \mathcal{N}\left[-\frac{m}{m+1}\frac{1}{2}\frac{\Delta\sigma}{\Delta\psi}(E^{j}_{m+1,n,i+1} - E^{j}_{m+1,n,i-1}) + \frac{n}{n+1}\frac{1}{2}\frac{\Delta\sigma}{\Delta\psi}(E^{j}_{m,n+1,i+1} - E^{j}_{m,n+1,i-1})\right], \tag{16}$$

where the index j indicates the propagation step, i is the grid point along the ψ axis ($i=1,\ldots,L$), and $N^{j}_{m,n,i}$ is the source term representing nonlinear effects ($\mathcal{N}$). The geometrical spreading term ($\mathcal{G}$) has been discretized in a manner consistent with the Crank–Nicolson method.[24,25] Finally, it should be noted that as computations progress from higher to lower order joint moments, the source term $N^{j}_{m,n,i}$, which involves joint moments of order $m+n+1$, is known at the propagation step j, when the joint moment of order $m+n$ is computed at propagation step $j+1$ (see also Fig. 1).

The computed values of $E^{j+1}_{m,n,i}$ at $\sigma_0+\Delta\sigma$ are updated to include the effect of atmospheric absorption, $E^{(\alpha)j+1}_{m,n,i}$. The dimensionless joint moment $E^{j+1}_{m,n}(\psi)$ is first put into its dimensional form $\tilde{E}^{j+1}_{m,n}(y)$. Because the absorption is independent of the amplitude of the signal, only the transformation of the time difference ($\psi=\omega_0 y$) is essential. The joint power spectrum $\tilde{S}^{j+1}_{m,n}$ (an array of $K=L$ elements) is computed next via fast Fourier transform. The absorption correction to $\tilde{S}^{j+1}_{m,n,i}$ is applied through Eq. (12), which in its digital form becomes

$$\tilde{S}^{(\alpha)j+1}_{m,n,i} = e^{-(m+n)\bar{\alpha}^{j}\Delta s}\tilde{S}^{j+1}_{m,n,i}, \quad \bar{\alpha}^{j} = \sum_{k=0}^{K-1}\alpha_k\tilde{S}^{j}_{1,1,k} \Big/ \sum_{k=0}^{K-1}\tilde{S}^{j}_{1,1,k}, \tag{17}$$

where the average absorption coefficient ($\bar{\alpha}^{j}$) has been computed in the previous known propagation step j, and K are the number of discrete frequencies that are contained in $\tilde{S}^{j}_{1,1}$. It should be noted that the dimensional propagation step Δs ($\Delta s=\bar{s}\Delta\sigma$) is used instead of $\Delta\sigma$. Equation (17) has computationally a very advantageous form, as it relates the computation of the corrected joint power spectra $\tilde{S}^{(\alpha)j+1}_{m,n}$ at propagation step $j+1$ to the PSD $\tilde{S}^{j}_{1,1}$ at the previous propagation step j. The corrected joint moment $\tilde{E}^{(\alpha)j+1}_{m,n}$ is obtained by inverse Fourier transform from $\tilde{S}^{(\alpha)j+1}_{m,n}$. Finally, $\tilde{E}^{(\alpha)j+1}_{m,n}(y)$ is put into dimensionless form to yield the final absorption-corrected $E^{(\alpha)j+1}_{m,n}(\psi)$.

The numerical solution of the gSBE, as it was the case for the SBE, has certain numerical limitations. As $m+n$ increases, the number of joint moments that must be computed increases as well, while the numerical values of the joint moments grow very large. This in turn increases the storage and memory requirements, while it also increases the truncation error in the solution of the corresponding partial differential equation. Furthermore, this truncation error is "transmitted" in the form of numerical oscillations from the higher order moments to the lower order ones, since higher order moments appear as a source term ($N_{m,n}$) in the lower order equation. Additionally, in the numerical solution of the SBE, Neumann and Dirichlet conditions had been applied for the even and odd order moments, respectively, at the boundaries of the computational domain (i.e., the end points of the ψ axis). Ghost points had been used to determine the required derivatives. The ghost point method introduced perturbations in the solution close to the boundaries of the computational domain.

Two remedies have been employed in the present work to improve the stability of the algorithm that is affected by the oscillations described above. First, a periodic boundary condition is applied for the pair of points $i=1$ and $i=L$ (the end points of the ψ axis), instead of the Neumann and Dirichlet conditions via the ghost points' method. This periodic boundary condition leads to the solution of a periodic tridiagonal system of $L-1$ unknowns, which satisfies the requirement of equal values between mesh points 1 and L,[24,25] and, thus, does not create oscillations. Second, numerical averaging[24] is applied to the source term (in order to remove numerical oscillations around the mean value) *after* the source term at propagation step j has been computed according to Eq. (16). In order to get a consistent result in the boundaries of the domain, the source term array is augmented on both sides according to the periodic condition discussed above. The above remedies have improved the stability of the algorithm and have allowed the prediction at larger propagation distances.

IV. RESULTS

Results from the method described here are compared with time domain calculations for the following two source conditions: (i) a sinusoidal signal, and (ii) a Gaussian noise signal (or equivalently, a Gaussian, stationary, and ergodic process) with known $S_{1,1}$ (or equivalently $E_{1,1}$) at source. In both cases, all joint moments $E_{m,n}$ at source can be computed from $E_{1,1}$ at source via the following recursive relations:[14,26]

$$\frac{m+n}{2}E_{m,n} = nE_{1,1}E_{m-1,n-1} + (m-1)E_{m-2,n},$$

$$E_{m,n} = nE_{1,1}E_{m-1,n-1} + (m-1)E_{m-2,n}, \tag{18}$$

for sinusoidal signals and Gaussian signals/Gaussian stochastic processes, respectively, both having zero mean and unit mean-square value. The comparisons with the sinusoidal source condition are benchmark comparisons to test the accuracy of the presented method for each one of the propagation effects added. Subsequently, the method is employed to predict the nonlinear evolution of a Gaussian stationary and ergodic stochastic process with known $S_{1,1}$ at source. PSDs ($S_{1,1}$) derived from measurements taken 18 m from an F/A-18E/F military aircraft engine[27] are employed to describe the considered processes.

Results from the present method are compared with time domain calculations obtained by the numerical solution of the following form of the generalized BE:[13,20]

$$\frac{\partial \tilde{p}}{\partial s} = -\frac{\mathcal{G}}{s}\tilde{p} + \mathcal{N}\frac{\beta}{\rho_0 c_0^3}\tilde{p}\frac{\partial \tilde{p}}{\partial t} + \mathcal{A}\left[\frac{\delta}{2c_0^3}\frac{\partial^2 \tilde{p}}{\partial t^2} + \sum_{\nu}\frac{c_\nu'}{c_\nu^2}\int_{-\infty}^{\tau}\frac{\partial^2 \tilde{p}'}{\partial l^2}e^{-(t-l)/t_\nu}dl\right]. \tag{19}$$

Equation (19) is a variation of Eq. (4) specific to atmospheric absorption. As before, the tags $\mathcal{G}$, $\mathcal{N}$, and $\mathcal{A}$ mark the terms related to geometrical spreading, nonlinear effects, and atmospheric absorption, respectively. The atmospheric absorption term describes the combined effect of thermoviscous attenuation (δ) and attenuation due to relaxation of oxygen and nitrogen, with $\nu=1,2$ being the index of the two relaxation processes, each characterized by a relaxation time t_ν and the corresponding net increase in phase speed (c_ν'), as frequency varies from zero to infinity. Equation (19) is solved numerically in the time domain employing the "Texas algorithm."[7,28]

Finally, in the present section, dimensionless nonlinear coefficients are presented that quantify the relative importance of nonlinear and dissipation effects for the two source conditions considered.

A. Sinusoidal source signal

The propagation of a sinusoidal source signal in real atmosphere is examined to evaluate the method presented. Consider the source signal $\tilde{p}=p_{\max}\sin(2\pi f_0 t)$, with $p_{\max}=10\sqrt{2}$ Pa and $f_0=1000$ Hz. Its evolution is predicted by Eq. (19) for (i) $\mathcal{G}=1$, $\mathcal{N}=1$, and $\mathcal{A}=0$; (ii) $\mathcal{G}=0$, $\mathcal{N}=1$, and $\mathcal{A}=1$; and (iii) $\mathcal{G}=1$, $\mathcal{N}=1$, and $\mathcal{A}=1$. The yielded time waveforms are then used to compute the joint moments via Eq. (3). The joint moments computed from the time domain calculations are subsequently compared with the joint moments as predicted directly in the joint moment domain by the present method. The characteristic pressure for this signal is $p_0=10$ Pa, the characteristic frequency $f_0=1000$ Hz, and the atmospheric absorption corresponds to temperature $T=20$ °C and relative humidity $h=70\%$.

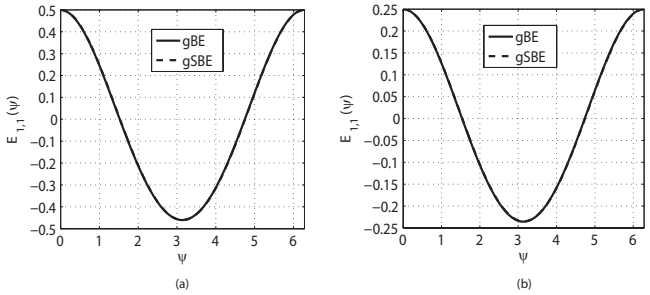


FIG. 3. Comparison between numerical results obtained by the gSBE and the gBE under the combined effect of nonlinear propagation distortion and geometrical spreading $\mathcal{G}=0.5/1$, $\mathcal{N}=1$, and $\mathcal{A}=0$ [cylindrical spreading, $\mathcal{G}=0.5$, shown on the left, and spherical, $\mathcal{G}=1$, on the right] for a sinusoidal source signal; $E_{1,1}$ labeled as gSBE is computed via the present method, and $E_{1,1}$ labeled as gBE by Eq. (19) and subsequent averaging of the time waveform; source radius $\sigma_0=0.5$; and propagation distance $\sigma=0.5$.

1. Comparison with time domain results

The case of a finite-amplitude cylindrically or spherically spreading wave in a nondissipative fluid is considered first ($\mathcal{G}=0.5/1$, $\mathcal{N}=1$, and $\mathcal{A}=0$). The evolution of $E_{1,1}$ is computed via Eq. (5) and is compared with $E_{1,1}$ obtained from the time domain calculations. Figure 3 illustrates the very good agreement.

The combined effect of nonlinear distortion and atmospheric absorption is examined next ($\mathcal{G}=0$, $\mathcal{N}=1$, and $\mathcal{A}=1$). Figures 4(a) and 4(b) show time domain results for distances of up to $\sigma=1.5$. It can be observed that the absorption mechanism reduces the wave amplitude, while it does not allow the formation of shocks. In Fig. 4(c) the corresponding solution in the joint moment domain with the proposed method is shown. Very good agreement with the time waveform solutions can be observed for distances of up to approximately $\sigma=1.2$ [see Fig. 4(a)]. Thereafter, the agreement deteriorates slightly around $\psi=0$ and $\psi=2\pi$. The maximum discrepancy does not exceed $(E_{1,1}^{gBE}|_{\psi=0} - E_{1,1}^{gSBE}|_{\psi=0})/E_{1,1}^{gBE}|_{\psi=0}=5.5\%$ and is due to numerical instabilities attributed to the nonlinear term. The compensating smoothing that must be applied reduces the accuracy [Fig. 4(e)].

Finally, the combined effect of nonlinear distortion, spherical spreading, and atmospheric absorption is tested ($\mathcal{G}=1$, $\mathcal{N}=1$, and $\mathcal{A}=1$). In Fig. 5(a) the pressure waveforms are illustrated for similar propagation distances σ as for the planar case. Compared to the plane wave case, spherical spreading causes a larger reduction in the wave amplitude, which in turn causes less distortion in the waveform. The corresponding solution in the joint moment domain is shown in Fig. 5(c), where very good agreement with the time domain simulations is observed. In Figs. 5(d) and 5(e) compari-

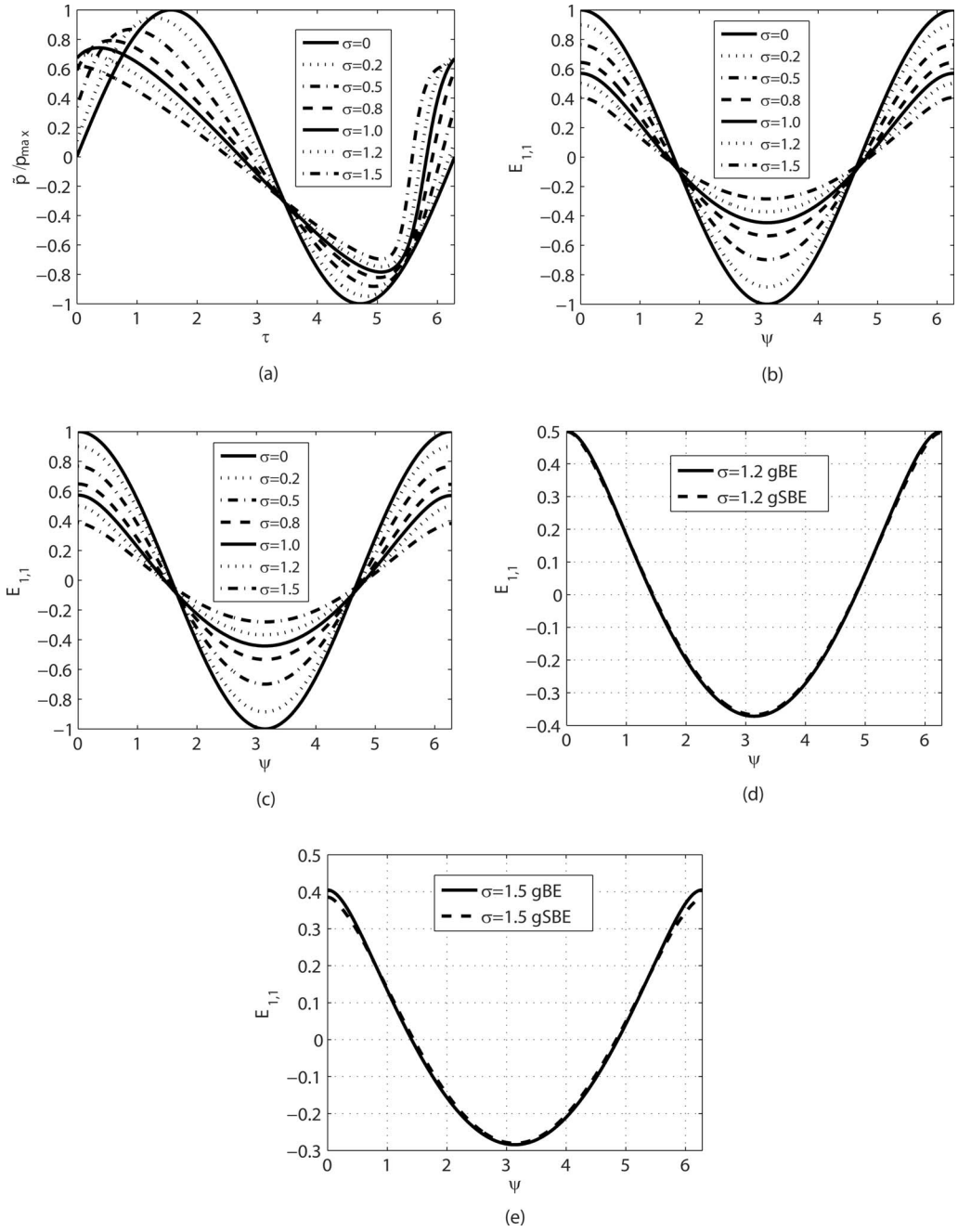


FIG. 4. Comparison between numerical results obtained by gSBE [Eq. (5)] and gBE [Eq. (19)] under the combined effect of nonlinear distortion and atmospheric absorption ($\mathcal{G}=0$, $\mathcal{N}=1$, and $\mathcal{A}=1$) for a sinusoidal source signal; (a) evolution of the pressure time waveform obtained by gBE, (b) $E_{1,1}(\psi)$ from averaging of time signals in (a), (c) from gSBE, and [(d) and (e)] comparison of $E_{1,1}(\psi)$ obtained by gBE (b) and gSBE (c) at selected propagation distances σ; atmospheric conditions $T=20\,^{\circ}\mathrm{C}$ and $h=70\%$.

sons are shown between the time waveform simulation and the joint moments' simulation for propagation distances of up to $\sigma=4$.

2. Relative importance of nonlinear effects—Nonlinear coefficient Π

Nonlinearity and dissipation are competing effects for the eventual formation of shocks in the pressure time waveform. The coefficient A_{tv} often used in the dimensionless form of the BE (it also appears in the SBE) $A_{tv}=\delta\omega_0\rho_0/2\beta p_0$ can be thought of as an inverse Reynolds number that describes the relative importance between nonlinear and dissipation effects. However, A_{tv} describes only thermoviscous dissipation effects and is not appropriate for propagation in atmosphere. The following dimensionless nonlinear coefficient Π is proposed instead:

$$\Pi=\frac{1}{\bar{s}a(f_0)}=\frac{\beta p_0 2\pi f_0}{\rho_0 c_0^3 a(f_0)}, \tag{20}$$

where $a(f_0)$ is the absorption coefficient for the single frequency f_0 of the sinusoidal source signal,[21–23] and $\bar{s}$ is the shock formation distance of a sinusoidal plane wave. In the following it is examined whether this coefficient can be used in the present method.

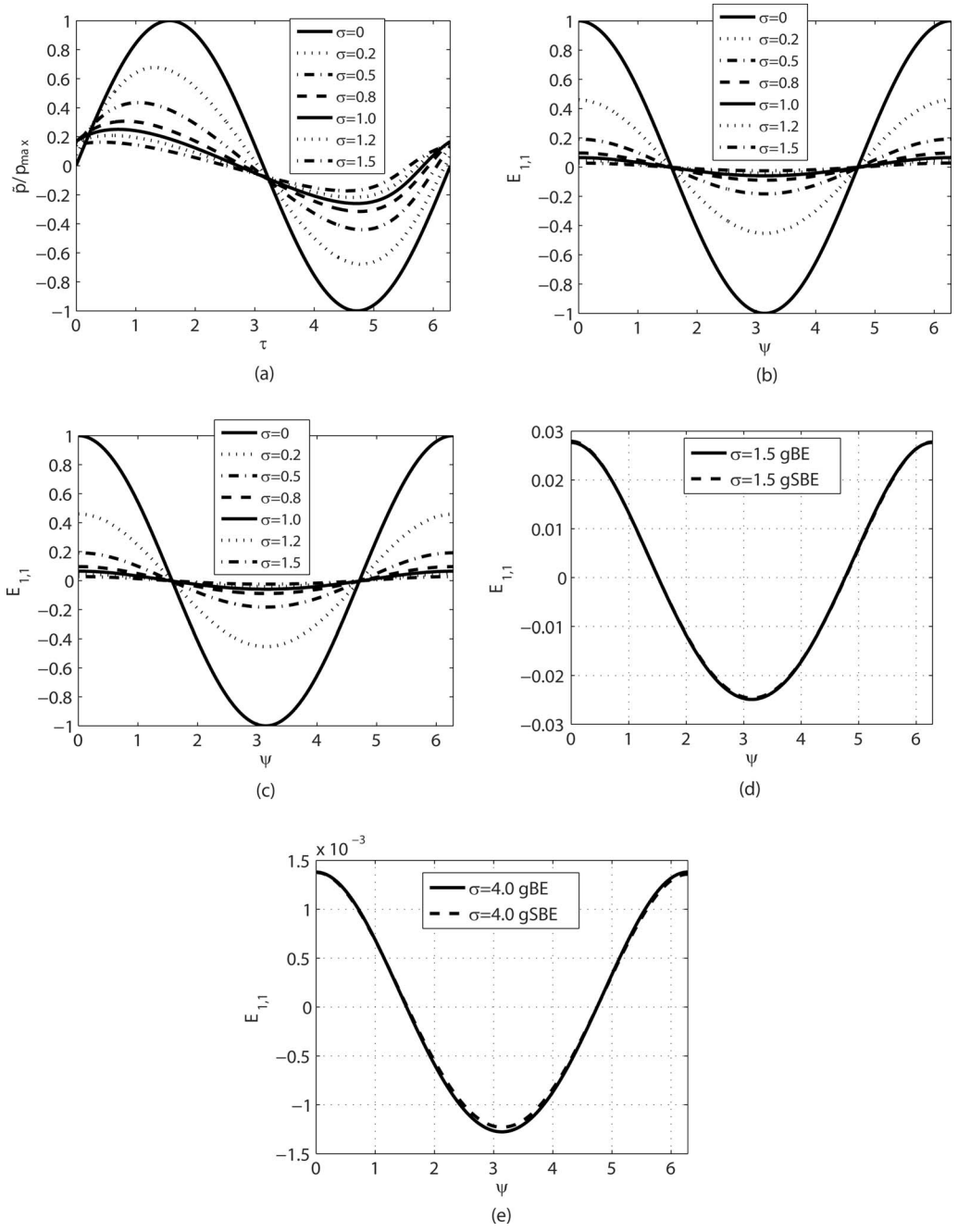


FIG. 5. Comparison between numerical results obtained by gSBE [Eq. (5)] and gBE [Eq. (19)] under the combined effect of nonlinear distortion, spherical spreading, and atmospheric absorption ($\mathcal{G}=1$, $\mathcal{N}=1$, and $\mathcal{A}=1$) for a sinusoidal source signal; (a) evolution of the pressure waveform obtained by gBE, (b) $E_{1,1}(\psi)$ from averaging of time signals in (a), (c) $E_{1,1}(\psi)$ from gSBE, and [(d) and (e)] comparison of $E_{1,1}(\psi)$ obtained by gBE (b) and gSBE (c) at selected propagation distances σ; source radius at $\sigma_0=0.5$; atmospheric conditions $T=20\,^{\circ}\mathrm{C}$ and $h=70\%$.

The nonlinear coefficient Π increases monotonically with increased p_0 indicating amplification of nonlinear effects. The absorption effects, on the other hand, depend on the frequency f_0 in a non-monotonic way. Figure 6 shows Π/p_0 as a function of frequency for atmospheric conditions $T=20\,^{\circ}\mathrm{C}$ and $h=30\%$. The curve has two minima at 250 Hz and 20 kHz, the resonance frequencies of nitrogen and oxygen, respectively, for the specific atmospheric conditions considered.

Figure 7 illustrates results obtained by the present method for a planar wave at propagation distance $\sigma=0.5$ for three sinusoidal source conditions with different p_0 and f_0,

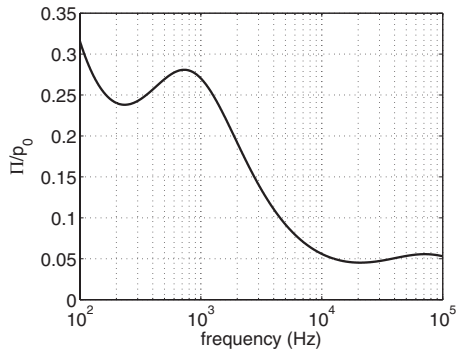


FIG. 6. Variation of Π/p_0 for a sinusoidal source signal as a function of frequency; atmospheric conditions $T=20\,^{\circ}\mathrm{C}$ and $h=30\%$.

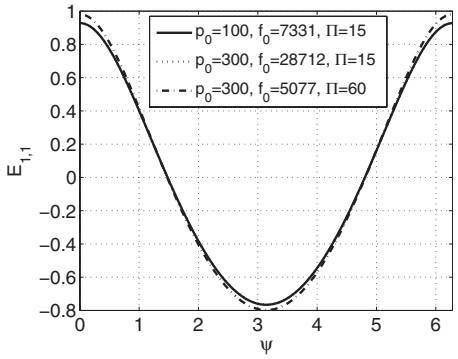


FIG. 7. Autocorrelation function $E_{1,1}(\psi)$ predicted by the gSBE at propagation distance $\sigma=0.5$ for three different planar sinusoidal waves; source conditions with larger values of Π exhibit stronger nonlinear distortion; source conditions with the same Π yield identical results; atmospheric conditions $T=20$ °C and $h=70\%$.

with two having the same Π value, namely, (i) $p_0=100$ Pa, $f_0=7331$ Hz, and $\Pi=15$; (ii) $p_0=300$ Pa, $f_0=28\,712$ Hz, and $\Pi=15$; and (iii) $p_0=300$ Pa, $f_0=5077$ Hz, and $\Pi=60$. It can be observed that the source condition with the larger Π exhibits stronger nonlinear distortion, indicating that Π can be used as a measure of the relative importance of nonlinear effects versus absorption. For details on the manifestation of nonlinear effects on $E_{1,1}(\psi)$ the reader is directed to Ref. 14. It should further be noted that source conditions with the same value of Π yield identical results, which renders Π a quantitative measure of nonlinearity for sinusoidal source conditions in the solution of the gSBE.

The nonlinear coefficient Π can be further modified to include the effect of source radius by substituting the shock formation distance for a sinusoidal plane wave, $\bar{s}$, with the corresponding shock formation distance of the spherical wave

$$\bar{s}_{\text{sph}} = s_0 e^{\rho_0 c_0^3/\beta p_0 \omega_0 s_0}. \tag{21}$$

Finally, it is noted that Π is a quantitative measure specific to the presented method, not for the general case of nonlinear propagation of a sinusoidal signal in real atmosphere.

B. Gaussian process

For the case of Gaussian noise signals three PSDs were considered. A reference PSD was obtained [see PSD_{high} in Fig. 8(a)] from the segment of the pressure time signal measured at 18 m from an F/A-18E/F aircraft engine with afterburners engaged[27] [see Fig. 8(b)]. Because this PSD corresponds to the noisiest condition considered, it is called PSD_{high}. Two derivative PSDs were created by subtracting 6 and 12 dB, respectively, from PSD_{high}. The resulting PSDs (called hereafter $\text{PSD}_{\text{medium}}$ and PSD_{low}) have the same "shape" (i.e., frequency distribution). Accordingly, in the time domain, they correspond to pressure time waveforms with the same shape but different amplitudes (half and one quarter of the original amplitude, respectively).

It should be noted that the segment of the pressure time signal [see Fig. 8(b)] employed for the derivation of the

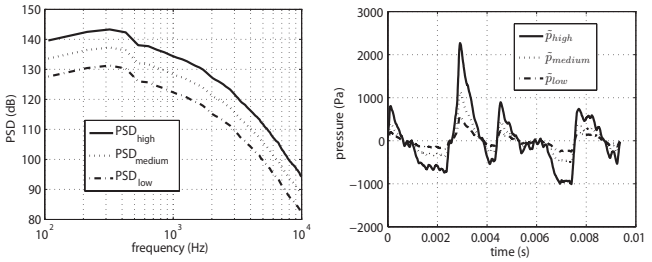


FIG. 8. The PSDs considered as source conditions (left); PSDs obtained from the time pressure waveforms shown on the right; $\tilde{p}_{\text{high}}$ from Fig. 1 of Ref. 27.

PSDs is short and the derived PSDs [see Fig. 8(a)] cannot be considered as representative of real aircraft noise. The derived PSDs are nevertheless employed, as the purpose of the present section is to demonstrate the accuracy of the method, not to extract results regarding aircraft noise propagation. It also noted that although the PDSs are not characteristic of aircraft noise, the pressure time signal $\tilde{p}_{\text{high}}$, shown in Fig. 8(b), is a segment of real aircraft noise time signal, and will be used to show the effect of employing non-Gaussian (instead of Gaussian) source conditions.

1. Relative importance of nonlinear effects—Nonlinear coefficient Π

The present method is first checked qualitatively. The three PSDs described above are considered. Nonlinear effects are expected to be more pronounced for PSD_{high} and less pronounced for $\text{PSD}_{\text{medium}}$ and PSD_{low}. Indeed, results obtained by the present method follow this trend.

Figure 9 shows results obtained by the present method after 7 m of propagation ($r_0=18$ m and $r=25$ m) expressed as the difference $\Delta\tilde{S}_{1,1}=\tilde{S}_{1,1}(f)|_{r_0}-\tilde{S}_{1,1}(f)|_r$. It can be observed that at the low frequency end of the spectrum $\Delta\tilde{S}_{1,1}$ is the same for all three cases and is predominately affected by spherical spreading. At higher frequencies $|\Delta\tilde{S}_{1,1}|$ is larger for the higher amplitude case. It should be further noted that $\Delta\tilde{S}_{1,1}$ becomes negative at high frequencies for the high and moderate amplitude cases, indicating a transfer of energy to high frequencies, as expected.

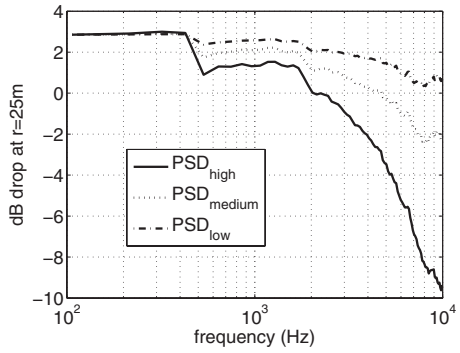


FIG. 9. $\Delta\tilde{S}_{1,1}$ [$\Delta\tilde{S}_{1,1}=\tilde{S}_{1,1}(f)|_{r_0}-\tilde{S}_{1,1}(f)|_r$; $r_0=18$ m, $r=25$ m] for the three PSDs shown in Fig. 8; spherically spreading propagation in the atmosphere ($T=20$ °C and $h=70\%$).

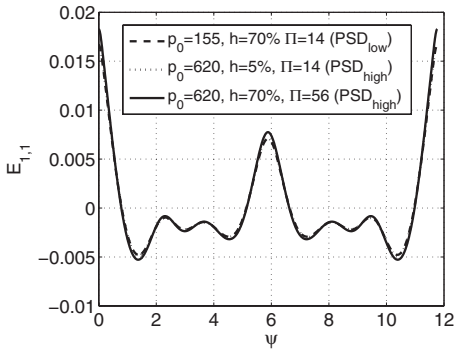


FIG. 10. Autocorrelation function $E_{1,1}(\psi)$ predicted by the gSBE for three different Gaussian source conditions based on $\mathrm{PSD}_{\mathrm{high}}$ and $\mathrm{PSD}_{\mathrm{low}}$; source radius $\sigma_0=0.1$ and propagation distance $\sigma=0.6$; source conditions with larger values of the nonlinear coefficient Π exhibit stronger nonlinear distortion; and source conditions with the same Π yield identical results.

For a Gaussian noise signal or a Gaussian stochastic process, the nonlinear coefficient Π can be defined as

$$\Pi = \frac{1}{\bar{s}\bar{a}} = \frac{\beta 2\pi p_0 f_0}{\rho_0 c_0^3 \bar{a}}, \tag{22}$$

where p_0 is the characteristic pressure that renders the mean-square value of the process unit, f_0 is a characteristic frequency, either the frequency carrying the most acoustic energy, or the inverse correlation rime, and $\bar{a}$ is the average absorption coefficient as defined in Eq. (12). All three required quantities for the evaluation of Π (p_0, f_0, and $\bar{a}$) can be obtained directly from the PSD at source. Further, it should be noted that Eq. (22) collapses to Eq. (20), when the PSD at source contains a single frequency.

In terms of the coefficient of nonlinearity and for atmospheric conditions of $T=20$ °C and $h=70\%$, the three PSDs yield $\Pi_{\mathrm{high}} \simeq 56$, $\Pi_{\mathrm{medium}} \simeq 28$, and $\Pi_{\mathrm{low}} \simeq 14$, respectively. It is, therefore, observed that Π can be used to qualitatively predict the relative importance of nonlinear to absorption effects for Gaussian noise signals or stochastic processes. The nonlinear coefficient Π can be further employed as a quantitative measure in the solution of the gSBE. Source conditions with the same value of Π yield identical results in the solution of the gSBE. Figure 10 shows the evolution of $\mathrm{PSD}_{\mathrm{high}}$ and $\mathrm{PSD}_{\mathrm{low}}$ under the same atmospheric conditions (their Π values being different) compared with the evolution of $\mathrm{PSD}_{\mathrm{high}}$ but under different atmospheric conditions ($T=20$ °C and $h=5\%$) that render the Π value for $\mathrm{PSD}_{\mathrm{high}}$ the same as for $\mathrm{PSD}_{\mathrm{low}}$. Indeed it is observed that source conditions with the same Π value yield identical results. This observation also indicates that in dry environments a high-intensity noise source such as $\mathrm{PSD}_{\mathrm{high}}$ exhibits the same nonlinear behavior as a lower-intensity noise source such as $\mathrm{PSD}_{\mathrm{low}}$.

As in the case of the sinusoidal source condition, the nonlinear coefficient Π can be further modified to include the effect of the source radius by substituting into Eq. (22) the shock formation distance for a sinusoidal plane wave, $\bar{s}$, with the corresponding shock formation distance of the spherical wave $\bar{s}_{\mathrm{sph}}$ [Eq. (21)].

2. Comparison with time domain results

For cases of noise signals or stochastic processes the comparison with time domain calculations is more complex than for the case of a sinusoidal signal. Based on the PSD at source a great number of pressure time waveforms must be reconstructed, all having the same PSD but each one with a different random phase distribution. If this random phase distribution is uniform between $[-\pi, \pi]$, the reconstructed time waveforms constitute a Gaussian process. Each time waveform is numerically propagated by employing the gBE [Eq. (19)] and the resulting PSDs are averaged. The averaged PSD is compared with the evolution of the PSD of the stochastic process predicted by the present method directly in the joint moments' domain. The comparison procedure comprises the essence of the new method, which is to replace the reconstruction of many time waveforms, the prediction of their propagation, and the subsequent averaging of the results after propagation.

Before proceeding to the comparisons, a quick note is merited on the number of time waveforms that must be reconstructed so that the average PSD after propagation converges. The number of time waveforms needed for convergence increases with propagation distance and with increased nonlinearity (larger values of Π). In other words the further away from the source the solution is sought, and the more intense the noise source is, the more time waveforms are needed for convergence. It should also be recalled that a Gaussian stochastic process that undergoes nonlinear propagation ceases to be Gaussian as it propagates. It nevertheless remains stationary and ergodic.

For the comparisons presented here the average of 12 time signals was considered. Stability considerations, as in the case of the sinusoidal source condition, limit the total propagation distance that can be achieved. The simulation was performed using joint moments with maximum order $m+n=50$ and $L=4096$ sampling points to maintain accuracy for the high frequencies. Furthermore, the propagation step was $d\sigma \simeq 2.7\times 10^{-3}$ and the total simulation time 360 s on a PC. The short execution time should be emphasized as a major advantage of the method. In Fig. 11(a) the joint moment $E_{1,1}$ obtained by the new method is compared to the average $E_{1,1}$ from the 12 time domain simulations. Excellent agreement is observed between the new method involving only one simulation in the joint moment domain and the average of a (potentially large) number of time domain simulations. Similar agreement can also be observed in the corresponding PSDs [shown in Fig. 11(b)], where the energy shift to higher frequencies due to the nonlinear effects can be identified.

Similar comparisons have been performed for $\mathrm{PSD}_{\mathrm{medium}}$ and $\mathrm{PSD}_{\mathrm{low}}$, and are shown in Figs. 11(c) and 11(d) and Figs. 11(e) and 11(f), respectively. Again, the good agreement can be observed. It should be added that even $\mathrm{PSD}_{\mathrm{low}}$, which is the source condition with the lowest intensity tested, corresponds to a quite strong noise condition.

The presented results demonstrate that the gSBE can be employed to predict the evolution of the PSD based solely on the PSD at source (that is, without knowledge of the phase information), provided that the noise source is Gaussian.

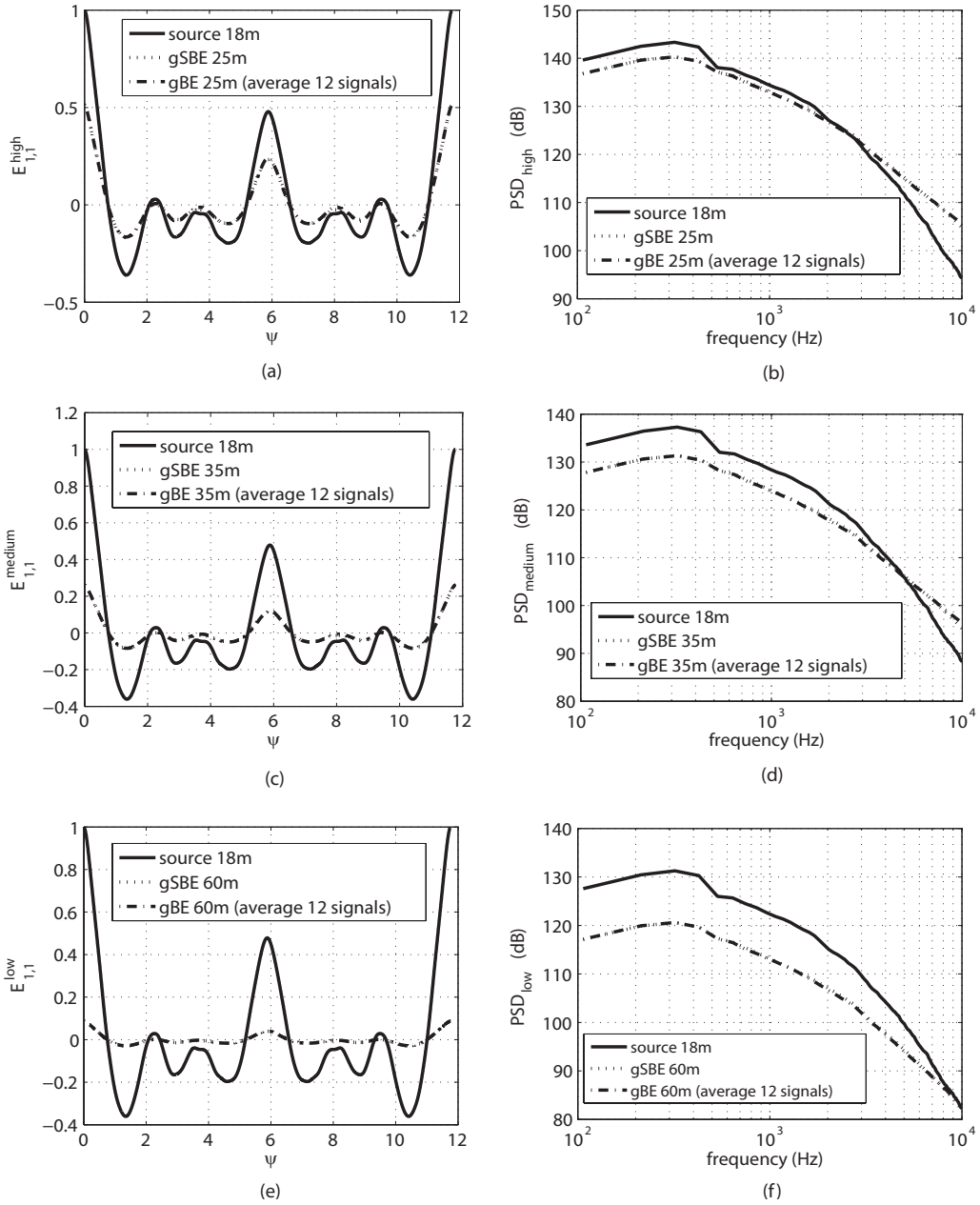


FIG. 11. Comparison between results obtained by the gSBE (dotted lines) and by averaging of time domain calculations employing the gBE (dashed lines); results shown for $E_{1,1}$ (left) and PSD (right); three source conditions (solid lines) are considered [PSD_{high} (top row), PSD_{medium}(middle), and PSD_{low}(bottom row)].

Non-Gaussian noise source conditions are now briefly discussed. Figure 12(a) shows the segment of real aircraft noise time signal [same shown as $\tilde{p}_{high}$ in Fig. 8(b)], which is non-Gaussian, as well as one of the infinite in number Gaussian noise signals that has the same PSD. Figure 12(b) shows the PSD at source and its evolution obtained: (i) by the gSBE assuming a Gaussian noise source [the results are the same as shown in Fig. 11(b)], and (ii) by the gBE using as source signal the non-Gaussian signal in Fig. 12(a). Deviation between the results obtained for Gaussian and non-Gaussian noise source conditions can be observed. This is expected and it is recalled that the method is applicable only to Gaussian noise source conditions.

C. Summary of limitations

The present work is a contribution toward a particularly challenging problem: to predict the evolution of the PSD without knowledge of the phase information at source. Many modifications have been implemented to improve its applicability compared to the original development presented in Ref. 14. The method in its current stage has certain limitations. First, the method makes use of an average absorption coefficient. This introduces an approximation in the computation of higher order moments [recall Fig. 2(c)]. Furthermore, dispersion is not taken into account. This restricts the method to non-dispersive or weakly dispersive fluids. Second, the method is restricted to Gaussian noise source conditions and cannot be applied to cases of non-Gaussian noise

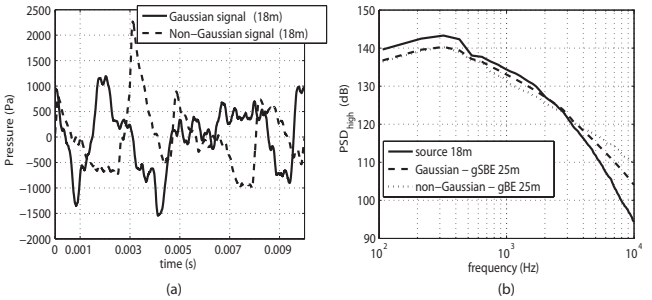


FIG. 12. Comparison between Gaussian and non-Gaussian source conditions both having the same PSD at source (PSD_{high}); (a) Gaussian and non-Gaussian noise source time signals having the same PSD at source, (b) PSDs obtained by the gSBE (dashed lines) assuming Gaussian noise source and time domain calculations obtained by the gBE (dotted lines) and the non-Gaussian noise source signal shown in (a).

sources [recall Fig. 12(b)]. Finally, numerical instabilities prevent the prediction of the PSD at long propagation distances [recall Figs. 4(e) and 5(e)].

V. CONCLUSIONS

A new equation was derived and termed gSBE that describes the combined effect of nonlinear propagation distortion, geometrical spreading, and arbitrary absorption in a weakly (or non-) dispersive medium on the PSD of the noise source condition. The numerical solution of the gSBE is based on a method previously developed for the solution of the SBE. The gSBE extends the SBE, which is restricted to plane waves, thermoviscous fluids, and short propagation distances. The geometrical spreading is accounted for in the joint moments' domain, while the absorption is accounted for in the joint power spectra domain via Fourier transforms between joint moments and joint power spectra. A dimensionless coefficient Π has also been proposed, which is applicable to the gSBE and can be used to provide the relative importance between nonlinear and absorption effects. Results from the present method are in good agreement with time domain calculations for sinusoidal source signals and Gaussian processes with known PSDs at source.

Future work will be undertaken in both the theoretical and numerical aspects of the method. The aim of the numerical work is to improve the stability of the numerical algorithm, which will allow the prediction of the noise PSD at distances much larger than the ones considered in the present study. The aim of the analytical work is to include additional propagation effects, such as dispersion, refraction, or diffraction. Furthermore, future work will include modifications to the method to make it applicable to cases, where the noise source has characteristics that deviate from Gaussian. The final scope is the application of the method to real, practical problems for predicting the evolution of high-intensity noise spectra into real fluids.

[1]C. L. Morfey and G. P. Howell, "Nonlinear propagation of aircraft noise in the atmosphere," AIAA J. **19**, 986–992 (1981).
[2]D. G. Crighton and S. Bashforth, "Nonlinear propagation of broadband jet noise," in *The Sixth Aeroacoustic Conference* (Institute of Aeronautics and Astronautics, Hartford, Conn., 1980).
[3]S. McInerny, K. L. Gee, M. Dowing, and M. James, "Acoustical nonlinearities in aircraft flyover data," AIAA Paper No. 2007-3654 (2007).
[4]H. H. Brouwer, "Numerical simulation of nonlinear jet noise propagation," AIAA Paper No. 2005-3088 (2005).
[5]K. L. Gee, V. W. Sparrow, M. M. James, J. M. Dowing, C. M. Hobbs, T. B. Gabrielson, and A. A. Atchley, "The role of nonlinear effects in the propagation of noise from high-power jet aircraft," J. Acoust. Soc. Am. **123**, 4082–4093 (2008).
[6]K. L. Gee, T. B. Gabrielson, A. A. Atchely, and V. W. Sparrow, "Preliminary analysis of nonlinearity in military jet aircraft noise propagation," AIAA J. **43**, 1398–1401 (2005).
[7]Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–916 (1995).
[8]F. H. Fenlon, "A recursive procedure for computing the nonlinear spectral interactions of progressive finite-amplitude waves in nondispersive fluids," J. Acoust. Soc. Am. **50**, 1299–1312 (1971).
[9]F. M. Pestorius and D. T. Blackstock, "Propagation of finite-amplitude noise," in *Finite-Amplitude Wave Effects in Fluids*, edited by L. Bjorno (IPC Science and Technology, England, 1974).
[10]V. W. Sparrow and R. Raspet, "A numerical method for general finite amplitude wave propagation in two dimensions and its application to spark pulses," J. Acoust. Soc. Am. **90**, 2683–2691 (1991).
[11]Y. Kallinderis, M. Manolesos, and P. Menounou, "A flow/acoustics interaction method for the prediction of sound propagation," Int. J. Aeroacoust. **6**, 171–197 (2007).
[12]M. S. Wochner, A. A. Atchley, and V. W. Sparrow, "Numerical simulation of finite amplitude wave propagation in air using a realistic atmospheric absorption model," J. Acoust. Soc. Am. **118**, 2891–2898 (2005).
[13]M. F. Hamilton and D. T. Blacktock, *Nonlinear Acoustics* (Academic, San Diego, CA, 1998).
[14]P. Menounou and D. T. Blackstock, "A new method to predict the evolution of the power spectral density for a finite-amplitude sound wave," J. Acoust. Soc. Am. **115**, 567–580 (2004).
[15]F. M. Pestorius, S. W. Williams, and D. T. Blackstock, "Effect of nonlinearity on noise propagation," in *The Second Interagency Symposium on University Research in Transportation Noise* (North Carolina State University, Raleigh, NC, 1974).
[16]R. H. Kraichnan, "The structure of isotropic turbulence at very high Reynolds numbers," J. Fluid Mech. **5**, 497–543 (1959).
[17]S. Gurbatov, A. Malakhov, and A. Saichev, *Nonlinear Random Waves and Turbulence in Nondispersive Media: Waves, Rays and Particles* (Manchester University Press, Manchester, NY, 1991).
[18]C. L. Morfey, *Nonlinear Propagation of Jet Noise in the Atmosphere* (Royal Aircraft Establishment, Technical Report 80004, 1980).
[19]G. P. Howell and C. L. Morfey, "Non-linear propagation of broadband noise signals," J. Sound Vib. **114**, 189–201 (1987).
[20]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1994).
[21]H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmopheric absorption of sound: Further developments," J. Acoust. Soc. Am. **97**, 680–683 (1995).
[22]H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Erratum: Atmospheric absorption of sound: Further developments," J. Acoust. Soc. Am. **99**, 1259–1259 (1996).
[23]ANSI, "Method for calculation of the absorption of sound by the atmosphere," S1.26-1996 (1996).
[24]D. A. Anderson, J. C. Tannehill, and R. H. Pletcher, *Computational Fluid Mechanics and Heat Transfer* (McGraw-Hill, New York, 1984).
[25]Ch. Hirsch, *Numerical Computation of Internal and External Flows: Fundamentals of Computational Fluid Dynamics*, 2nd ed. (Butterworth-Heinemann, Oxford, 2007).
[26]Y. K. Lin, *Probabilistic Theory of Structural Dynamics* (McGraw-Hill, New York, 1967).
[27]K. L. Gee, V. W. Sparrow, T. B. Gabrielson, and A. A. Atchley, "Nonlinear modeling of F/A-18E noise propagation," in The 26th AIAA Aeroacoustics Conference, Monterey, CA (2005).
[28]R. O. Cleveland, M. F. Hamilton, and D. T. Blackstock, "Time-domain modeling of finite-amplitude sound in relaxing fluids," J. Acoust. Soc. Am. **99**, 3312–3318 (1996).

Flow effects on the acoustic end correction of a sudden in-duct area expansion

Susann Boij[a)]
The Marcus Wallenberg Laboratory for Sound and Vibration Research and the Linné Flow Centre, KTH, Teknikringen 8, SE-100 44 Stockholm, Sweden

(Received 7 October 2008; revised 15 June 2009; accepted 16 June 2009)

For scattering of plane waves at a sudden area expansion in a duct, the presence of flow may significantly alter the reactive properties. This paper studies the influence of a mean flow field and unstable separated flow on the reactive properties of the expansion, formulated as an end correction. Theoretical and experimental results show that the expansion end correction is significantly affected by the flow and hydrodynamic waves excited at the edge of the expansion. The effects are different in three regions where the Strouhal number is small, of order 1, and large. The influence is most significant at Strouhal numbers of the order 1, with specific limiting values for large and small Strouhal numbers, respectively. In the analytic model, an important feature is the shear layer at the edge modeled as a vortex sheet with the unsteady Kutta condition applied at the edge. The influence of Mach number, Helmholtz number, and area expansion ratio is studied, and a quasistationary, small Strouhal number, approximation yields an expression for the end correction. Further, the influence of edge condition is explored, emphasizing the importance of interaction between sound and unsteady vorticity shedding at the edge of the area expansion.
 [DOI: 10.1121/1.3177263]

PACS number(s): 43.28.Py, 43.20.Mv, 43.20.Fn [AH] Pages: 995–1004

I. INTRODUCTION

Exhaust and ventilation systems are examples of devices that transmit noise and consequently are subject to acoustic treatment. The primary purpose of these systems is gas transport, most often at relatively low Mach numbers. The objective of this paper is to further investigate some effects of the flow field on the acoustic properties of a duct system consisting of two semi-infinite ducts connected by a sudden area expansion with a sharp edge where flow separation occurs. Here, the interest is in the reactive parts of the acoustic field generated by scattering of incident plane waves at this edge. To study the phenomena, the argument of the reflection coefficient normalized by the Helmholtz number is selected, to arrive at an end correction formulation. The presence of the flow separation is shown to have a significant effect on the end correction, and the influence of different parameters, such as frequency, Mach number, and area expansion ratio, is explored. The results are general to other concepts describing the reactive part of the scattered field, such as added impedance and added mass. Also, the phenomena observed are expected to occur for the other concepts describing the acoustic propagation properties of the duct area change. The work builds on the studies of the reflection and transmission properties of the flow duct area expansion published by Boij and Nilsson[1,2] and Boij.[3]

Existing low frequency models give good predictions of the acoustic behavior of area expansions in the absence of flow, and when the only effect of the flow is convection. However, the flow may significantly affect the acoustic properties of a duct area expansion through influence of flow acoustic coupling for separated flows. A deeper knowledge of the phenomena due to flow acoustic interaction is a requirement to construct simple models of the end correction for low frequencies including these effects. Among the first to study the influence of mean flow and flow separation on the acoustic plane wave propagation in more detail was Bechert,[4,5] who from experimental data revealed for the first time that part of the acoustic energy was dissipated while incident on the open end of a duct with superimposed mean flow forming a jet downstream of the pipe termination. A theoretical explanation was presented by Howe,[6] showing that part of the acoustic energy was converted into hydrodynamic energy when vortex shedding is induced by incident acoustic waves. Several models that include the interaction of acoustic waves with a vortex sheet—a jump in mean flow velocity—have then been published,[7–9] stressing the importance of unstable vortex waves—corresponding to a Helmholtz instability—and the application of the unsteady Kutta condition at the edge where flow separation occurs. For the unflanged flow-pipe termination, scattering properties with flow are treated theoretically by Munt.[10] Several authors[11,12] treated the case of an aperture in a diaphragm, i.e., an orifice, in a flow duct. Effects of grazing flow over perforated plates or the neck of Helmholtz resonators and side branches have been studied, by, e.g., Ronneberger,[13] Cummings,[14] and Ajello.[12]

A scattered sound field generated from plane waves incident on an aperture of some kind, consists of a near field—a reactive field—and a propagating component. Determination of the reactive properties is a classical acoustic problem, with applications ranging from musical instruments to design of automobile mufflers. This phenomenon was described already by Rayleigh[15] who recognized that the acoustic length

a)Electronic mail: sboij@kth.se

of, e.g., organ pipes and flutes, differed from the physical length, the difference denoted the end correction corresponding to the extension of the pipe where the boundary condition of zero acoustic pressure is achieved. The end correction is an important property, as it relates to resonance phenomena. This is of great importance for all types of devices that are designed for a specific behavior around a given frequency, as the acoustic length determines the resonance properties. Such properties are of interest both for pipe terminations and for in-duct elements applying both to area changes along the duct, such as area expansions and orifices, and openings in the duct walls like those of perforated plates or side branches.

For the in-duct case, the end correction corresponds to excitation of evanescent higher order modes. The first results that were published on the topic treat the no flow case, originating both from studies of acoustic and electromagnetic wave propagation. An analytical solution for the sound reflected at an open pipe termination without mean flow was first presented by Levine and Schwinger.[16] For the case of area discontinuities in ducts the works by Miles[17] and Karal[18] are classical.

The effect of flow separation on the reactive part of the acoustic properties is a topic of more recent interest. A model including the reactive part of the scattering properties for the pipe termination was presented by Munt,[10] and an overview and experimental results are found in Ref. 19. They concluded that with the mean flow present in the duct, the end correction is best described as a function of the Strouhal number (*St*), rather than as a function of, for example, the Helmholtz number. Theoretical predictions for small[8] and large *St* (Ref. 20) are verified experimentally. In the intermediate region where the Strouhal number is of order 1, the end correction increases quite rapidly from the small *St* limit to the large *St* limit. Recent experimental results and simulations of the exact Munt's model verify the results.[21] Similar results for the area expansion are found in this paper.

For a sudden area expansion in a duct, a number of simplified models have been put forward. The influence of the convective effects have been investigated, and shown negligible.[22] The same conclusion is drawn by Davies,[23] where it is suggested that the end correction for the case of no mean flow can be a substitute. However, the coupling between the acoustic field and the flow field at the edge is overlooked in these works and the behavior for small Strouhal numbers, i.e., the hydrodynamic region, has not been explored. A model for cylindrical ducts with an area expansion was presented by Nilsson and Brander.[24] Recently, data for the phase of the reflection coefficient were published by Boij and Nilsson[1] and by Kooijman *et al.*[25] Based on experimental[26] verification of the model used by Boij and Nilsson,[1] this paper aims at investigating the parameter dependence and influence on higher order mode properties on the reactive part of the acoustic properties of the sudden area expansion, concentrating on the end correction.

While purely convective effects may be negligible for the acoustic end correction of a duct area expansion, the results presented in this paper show that the presence of flow may have a non-negligible influence on the scattering properties both for intermediate and small Strouhal numbers. Flow-acoustic interaction at Strouhal numbers around 1 can substantial change the end correction and, at lower Strouhal numbers, the wave propagation properties deviates from that of the no flow case. A similar behavior is also found for the open pipe termination.[8,19,21] In a certain Strouhal number regime, the experimentally deduced end correction is only a third of the end correction for the no flow case, a non-negligible difference if the end correction is to be considered at all.

II. THE END CORRECTION

The reactive part of the acoustic properties of a sudden area expansion in a flow duct is influenced by mean flow, flow separation, and excitation of hydrodynamic waves. To study these effects, we chose the concept of end correction, as defined below. Concepts both of impedance and scattering matrices may be disputed in the presence of mean flow. However, as a parallel to the studies of end correction and radiation impedance for open pipe terminations mentioned above, the concept of end correction is of interest. The formulation reveals the significant influence of the Strouhal number, *St*, and further it explores the non-uniform limit for small Helmholtz number *ka*, i.e., the product of the wave number k and a typical cross-dimension of the duct a, and Mach number, M, already acknowledged for the open pipe.[8,9] As the argument of the reflection coefficient is proportional to the Helmholtz number, a normalization by this number uncovers the dependence on M and *St* already seen[1,25] for the case of constant Helmholtz number and varying M.

When sound waves encounter a sudden increase in cross-sectional duct area, evanescent higher order modes are excited and constitute an important part of the sound field close to the area expansion. In the plane wave regime, i.e., when the frequency of the propagating wave is so low that only the fundamental mode is propagating in the duct, this can be observed in the reflection coefficient for the plane wave as an additional phase shift. The additional phase shift can be reformulated as an added length, the end correction. Assuming a time dependence $\exp(-i\omega t)$ and with $k_{\pm}=\omega/c_{\pm}$, where $c_{\pm}$ is the effective speed of sound for the incident and reflected plane wave, respectively, the end correction, Δl, used in this paper is defined as[11]

$$R = |R|e^{i\phi} = -|R|e^{i(k_+ + k_-)\Delta l}. \tag{1}$$

An alternative way of defining an end correction, or added length, is from the impedance jump at the expansion

$$Z_{\text{area\_change}} = Z_2 - i\omega m, \quad m = k\Delta L/\rho/S_1, \tag{2}$$

where m is denoted the added mass, and the subscript 1 indicates the upstream and 2 indicates the downstream duct parts, respectively. In the low frequency limit, the relation between the added length, ΔL, from Eq. (2), and the end correction, Δl, from Eq. (1), is then, using the standard relation between reflection coefficient and impedance, given by

$$k\Delta L = k\Delta l(1-\eta^2), \quad \eta = S_1/S_2,$$

where S is the cross section area and η denotes the area expansion ratio. For the case of no mean flow, with the speed of sound, c_0, and $k=\omega/c_0$, the end correction Δl is defined by

$$R = -|R|e^{i2k\Delta l}, \tag{3}$$

yielding a relation between the phase shift and the end correction as $\phi = 2k\Delta l \pm \pi$, with the sign of the term π chosen so that ϕ is positive and less than π. With a mean flow Mach number defined as $M=U_0/c_0$ where U_0 is the mean flow speed, the respective wave numbers for the plane waves become $k_+=k/(1+M)$ and $k_-=k/(1-M)$, and the definition of the end correction is

$$R = -|R|e^{i(k_+ + k_-)\Delta l} = -|R|e^{i2k(1-M^2)^{-1}\Delta l}. \tag{4}$$

Hence, the relation between the phase shift and the end correction is given by

$$\phi = 2k(1-M^2)^{-1}\Delta l \pm \pi. \tag{5}$$

In the following the dimensionless end correction, δ, is defined as $\delta=\Delta l/a$ where a is a typical cross-dimension of the duct.

Expressions for the low frequency limit for the end correction of an area expansion without flow are found, e.g., in Ref. 27, for a cylindrical duct as a function of area expansion ratio η, as

$$\delta_c = \left(\frac{\Delta l}{a}\right)_c = \frac{1}{1-\eta^2}\sqrt{\frac{(1-\sqrt{\eta})^2(1-\eta)(15-2\sqrt{\eta}-\eta)}{24}}, \tag{6}$$

where a is the radius of the smaller duct. For a rectangular duct, Miles[17] derived a plane wave reflection coefficient, from which the following end correction can be obtained:

$$\delta_r = \left(\frac{\Delta l}{a}\right)_r = \frac{1}{\pi}\left\{-\frac{2\ln(4\eta)}{(1-\eta)(1+\eta)} + \frac{(1+\eta)\ln(1+\eta)}{\eta(1-\eta)} - \frac{(1-\eta)\ln(1-\eta)}{\eta(1+\eta)}\right\}, \tag{7}$$

where a is the width of the smaller duct. As the difference in area increases toward infinity, i.e., $\eta\to 0$, the end correction for a cylindrical duct expansion tends to a finite value of 0.82,[28] the value for a baffled duct termination. The end correction of the rectangular geometry, on the other hand, tends logarithmically to infinity. This difference is related to the fundamental difference between the fields generated by two- and three-dimensional acoustic sources.[11]

As shown later in this paper, the concept of low frequency values in the presence of flow is more complex. For large Strouhal numbers, the Helmholtz number is the governing parameter and the expressions in Eqs. (6) and (7) are relevant. However, the end correction values for small Strouhal numbers will depend on the Mach number, whereas for Strouhal number of order unity the end correction is dependent on the Strouhal number.

III. FLOW INTERACTION EFFECTS

The classical concepts describing the acoustics of duct elements, such as scattering coefficients and impedance, are formulated for the case of zero mean flow. In the case of non-zero mean flow the physical interpretation may thus not be as straightforward. However, the concept of end correction has been used to explore the influence of mean flow on the properties of open pipe terminations. Therefore we apply the same concept to the case of a sudden area expansion in a flow duct. Experimental as well as analytical results are presented to show the influence of the mean flow and the unstable shear layer, revealing the flow-acoustic coupling effects on the end correction. The acoustic field studied covers low frequencies, below the cut-on frequency for excitation of higher order modes. The flow speed is up to Mach number 0.5, so the Strouhal number, as defined below, ranges from small to large. The experimental results are for a cylindrical duct, whereas the theory is derived for a two-dimensional, rectangular geometry. A comparison is performed assuming[11] that the effect of area expansion ratio dominates for small Helmholtz numbers. In the analysis, the dimensionless *Strouhal number* is an important parameter, and is defined as

$$St = \frac{ka}{M}, \tag{8}$$

where a is a typical cross-dimension of the duct. For a cylindrical geometry the Strouhal number is defined as

$$St_c = \frac{(ka)_c}{M}, \tag{9}$$

where a is the radius of the small duct, and for a two-dimensional, rectangular geometry,

$$St_r = \frac{(ka)_r}{M}, \tag{10}$$

where a is the width of the small duct. The low frequency limits for the no flow case presented above corresponds to the limit when $M\to 0$ for a fixed (small) value of ka; hence it is the value as the Strouhal number tends to infinity for finite ka.

A. Experimental results

The end correction is calculated from values of the complex reflection coefficient, determined experimentally. These experimental values have not previously been used to determine the end correction, but have been presented as the phase of the reflection coefficient in the work by Ronneberger,[26] presenting detailed measurements of the complex scattering coefficients. The high quality of these phase measurements allows for studies of the flow effects on the end correction. The duct geometry in the experiment is cylindrical, with a radius of 25 mm for the smaller duct and 42.5 mm for the larger duct, and the experiment is performed at frequencies 0.5, 1.0, 2.0, 2.5, and 3.0 kHz and Mach numbers between 0 and 0.45. Note that the experimental setup is designed so that the first higher order non-radial mode is not excited.

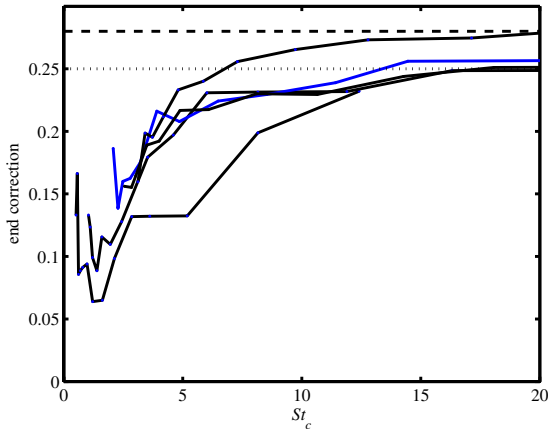


FIG. 1. (Color online) The end correction, $\Delta l/a_c$, as a function of the Strouhal number St_c [see Eq. (9)], calculated from experimental data for the phase of the reflection coefficient (Ref. 26). The theoretical value for the no flow low frequency limit: dashed line, Eq. (6), and dotted line (Ref. 27). The dots on the solid lines indicates data for, from below, $f=0.5,1,2,2.5$, 3 kHz, $a=25$ mm, and M is between 0.01 and 0.5, for an area expansion ratio $\eta=S_1/S_2=0.35$.

The end correction based on the experimental data is calculated as

$$\delta = \frac{\phi + \pi}{2(ka)_c}(1 - M^2), \tag{11}$$

from the experimental results for the phase ϕ of the plane wave reflection coefficient. Curves for the end correction for a set of different Helmholtz numbers are presented as a function of the Strouhal number in Fig. 1.

For large Strouhal numbers (decreasing Mach number for fixed Helmholtz number), the experimental data agree with the low frequency limit for the no flow case, Eq. (6), indicated by the dotted line. For Strouhal numbers below St_c around 5, however, the end correction deviates significantly from the no flow values. Furthermore, the end correction decreases with decreasing Strouhal number until it reaches a minimum and increases again for small enough Strouhal numbers. It is observed that the end correction can be as small as 30% of the value for large Strouhal numbers, i.e., a reduction in the end correction of 70%. The end correction has a clear minimum when the Strouhal number is between 1 and 2. The qualitative shape of the end correction curves is a function of the Strouhal number, a quantity related to flow-acoustic interaction. The end correction is influenced both by the Mach number, i.e., the flow conditions, and the Helmholtz number, i.e., the acoustic properties of the incident wave in relation to the duct geometry. Thus, the variation is dependent on the relationship between the three parameters frequency, duct dimensions, and flow speed, rather than on each parameter separately. A further observation is that for large Strouhal numbers the no flow approximation that is often used seems appropriate, whereas for Strouhal numbers of order 1 and for small Strouhal numbers the assumption that the influence of the flow field is negligible is no longer valid. The results show that commonly accepted methods to compute the end correction in flow ducts,[23] i.e., to ignore the influence of flow, can give significant errors for certain parameter combinations.

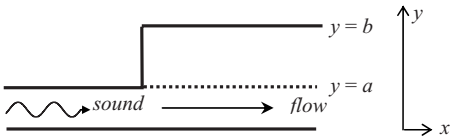


FIG. 2. A sound wave incident on an area expansion. A mean flow is present in the lower part of the duct, the dotted line indicating an infinitely thin shear layer.

B. Theoretical results

Theoretical results are computed from simulation by the model for scattering of sound at area discontinuities in flow ducts, presented by Boij and Nilsson.[1] The geometry of the area expansion is shown in Fig. 2. The model is a full mode analytical model, and the scattering properties are determined using Wiener–Hopf technique. It is a linear model, where viscous as well as thermal dissipation is neglected. The flow is assumed subsonic. Downstream of the area expansion the flow field is modeled by a vortex sheet emanating from the edge of the duct expansions, separating the moving fluid in the lower part from the quiescent fluid in the upper part. The solution of the wave equation in this part of the duct includes two hydrodynamic modes, propagating downstream with a speed proportional to the mean flow speed. One of these corresponds to the Kelvin–Helmholtz instability of the shear layer, and is exponentially growing, while the other is a damped mode. Modeling the shear layer as a vortex sheet is an approximation valid when the thickness of the shear layer forming at the edge is much smaller than the wavelength of the incident sound, i.e., for small shear layer Strouhal numbers. As the flow is confined to the lower part of the duct along the entire duct, the effects of the jet expansion, expected to influence the acoustic wave propagation at high frequencies, is neglected in the model. In order to determine a unique solution for the scattering, the unsteady Kutta condition is applied at the edge.[29] This edge condition allows for coupling between the acoustic field and the instability waves induced by the vortex sheet downstream of the sharp edge. Earlier studies[1,2] show that this model indeed incorporates the main effects of the flow-acoustic interaction in plane wave scattering at a sudden area expansion. These results indicate that, in the plane wave region, the effect of the jet expansion on the back scattering is negligible for both reflection and transmission of sound incident on an area expansion, and that the flow acoustic interaction is governed by the conditions at the edge. The model is thus suited to model the end correction since it is a low frequency concept, applicable in the plane wave range.

The end correction is computed from the phase, ϕ, of the plane wave reflection coefficient, according to

$$\delta = \frac{\phi + \pi}{2(ka)_r}(1 - M^2). \tag{12}$$

For the theoretical results, the geometry is rectangular and asymmetric as depicted in Fig. 2. The number of interacting modes included in the computation of the scattering coefficients is 10, including the plane waves and the hydrodynamic modes. Theoretical results for the end correction for the same parameter values as for the experimental data in

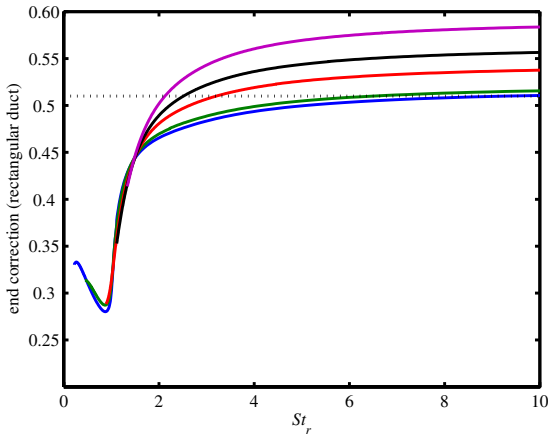


FIG. 3. (Color online) The end correction, $\Delta l/a_r$ as a function of the Strouhal number St_r [see Eq. (10)], computed with the vortex sheet model (Ref. 1). The theoretical value for the no flow, low frequency limit (7) is indicated by the dotted line. The curves correspond to (from below) ka_r =0.11,0.22,0.45,0.56,0.67. The curves are generated by varying M between 0.01 and 0.5 for fixed ka, for an area expansion ratio $\eta=S_1/S_2$ =0.35.

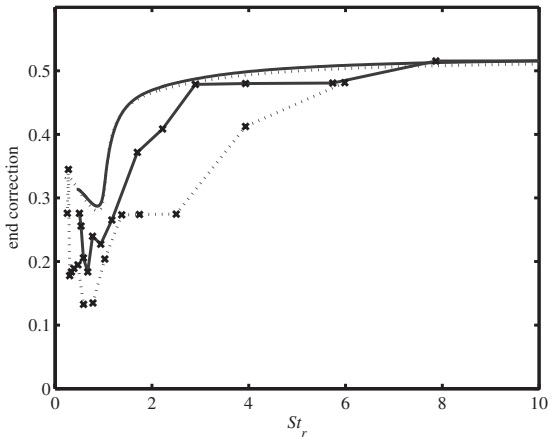


FIG. 4. The end correction, $\Delta l/a_r$, as a function of the Strouhal number St_r [see Eq. (10)]. For the experimental values, the Helmholtz number is scaled according to Eq. (13). The Mach number M is between 0.01 and 0.5, and the area expansion ratio $\eta=S_1/S_2=0.35$. Experimental data: x, vortex sheet model: continuous line. The curves correspond to $(ka)_r=0.11$ (dotted); 0.22 (solid)

Sec. III A are displayed in Fig. 3. Note the difference in geometry with the experimental, cylindrical duct. This difference is expected to have a negligible influence on the acoustic properties in the low frequency regime, as discussed above.

In Fig. 3, the predictions of the end correction for the area expansion ratio 0.35 are presented as functions of the Strouhal number for certain values of ka. Note that the curves are generated by varying the Mach number between 0.01 and 0.5, while the Helmholtz number is kept constant; thus the curves start at different Strouhal numbers. The graph shows that for large St (decreasing Mach number for constant ka), there is good agreement between the end correction calculated for small values of ka and the low frequency no flow limiting value, indicated by the dashed line. In this case, the low frequency limit of the end correction[17] without flow is 0.51, Eq. (7). For Strouhal numbers of the order of 1, the end correction decreases with decreasing Strouhal number. A distinct minimum for the end correction occurs slightly below $St_r=1$, and the Strouhal number of the minimum does not vary significantly with frequency. When the Strouhal number is smaller, the end correction increases again, but to a much lower value than for the large Strouhal numbers.

C. Comparison between experimental and theoretical results

Now, a closer comparison between experimental and predicted results is presented. An equivalent Helmholtz number is used for the frequency scaling[1,30] and for the calculation of the end correction. An equivalent Helmholtz number $(ka)_r$ corresponding to a two-dimensional rectangular geometry is related to the Helmholtz number of the cylindrical duct, $(ka)_c$, by

$$(ka)_r = \frac{\pi}{\kappa_0}\sqrt{\eta}(ka)_c. \tag{13}$$

This corresponds to a normalization where, for a given area expansion ratio, the Helmholtz number kb of the larger duct is normalized by the cut-on frequency of that duct, in the cylindrical case $k_0 b_c=\kappa_0\approx 3.832$ and in the two-dimensional rectangular case $k_0 b_r=\pi$, where b is the radius and width, respectively, of the large duct. The underlying assumption is that the cut-on frequency of the first higher order mode in the larger duct section characterizes the plane wave frequency dependence. Note that the factor $\sqrt{\eta}$ originates from the fact that the cylindrical case is three-dimensional, whereas the rectangular case is two-dimensional. This scaling has proven to give good agreement for the reflection coefficient[1] and for the transmission coefficients as well as for predictions of acoustic dissipation due to vortex shedding[2] at a trailing edge, and have been used by several other authors.[25,31]

The end correction is calculated using Eq. (12) for both calculated and measured values of the phase of the reflection coefficient, where the equivalent Helmholtz number $(ka)_r$ is used for the experimental calculations. The values of the end correction in regimes where the flow affects the results are not expected to scale qualitatively; however, the interest is to verify the trends rather than absolute numbers.

Experimental and theoretical results are shown together in Fig. 4, where the end correction is presented as a function of the Strouhal number, as defined by Eq. (10). The fundamental properties of the end correction observed in the experimental results are captured by the theoretical model, and the minimum end correction occurs at approximately the same St for experiments and predicted values. For large Strouhal numbers, the curves approach the no flow results. The overall behavior is distinct in the experimental as well as the theoretical results, verifying the variations. The good correspondence of theory with experiments suggests that the applied model captures the important interaction effects between the mean flow and vortex shedding, and the acoustic field. This allows us to use the model for parameter studies, and some results are presented in later sections of this paper.

There are certain deviations between the experimental and the theoretical results. An important difference is that in the experimental data, the *St* corresponding to a minimum in the end correction shows a slight dependence on the Mach number for the experimental curves, whereas the theory misses this effect and predicts a minimum at a fixed *St*. Here, the simplifications in the model concerning the mean flow field (neglect of the shear layer expansion) may be one main reason as to why the shift in Strouhal number is not predicted by the model. Further studies are required to determine how the Strouhal number corresponding to the minimum varies with M and ka. Concerning the minimum values of the end correction, the shapes of the curves are similar, but the magnitude at the minimum relative to the no flow value differs for the cylindrical/experimental and the rectangular/predicted case. This difference may well be due to the differences in geometry between the cylindrical and the rectangular (two-dimensional) duct. In this regime, the flow-acoustic coupling is dominating and the details of this may be different for the cylindrical and a rectangular geometry. Another aspect is that the scaling used only accounts for acoustic properties and do not reflect the effects of flow-acoustic interaction. Also, the difficulties to accurately measure the phase in the presence of flow may induce measurement errors. Note that the end correction is calculated from the phase of the reflection coefficient divided by the Helmholtz number. Thus, if the measurement error in the phase is independent of frequency, the relative error of the end correction will increase with decreasing frequency.

In summary, the end correction of an area expansion can be significantly influenced by the presence of mean flow and vortex shedding, and the difference in end correction between the largest and smallest values can be of the order of the end correction itself. Thus, neglecting the variations in the small Strouhal number region gives an error that is of the same order as when the end correction is completely neglected. This effect is predicted by the vortex sheet model, results that are verified by the comparison above.

IV. STUDY OF PARAMETER DEPENDENCE

In the presence of mean flow, parameters that influence the end correction of a sudden area expansion are the Helmholtz number, the Mach number, and area expansion ratio. The problem formulation sets some limits to the parameter range of interest. First the end correction, as stated earlier, is a concept that is applicable in the plane wave range, i.e., for frequencies where only the fundamental duct mode is propagating. This corresponds to a maximum Helmholtz number for the larger duct, kb, of π for a two-dimensional rectangular duct and ≈ 3.832 for an axi-symmetric cylindrical geometry. Also, subsonic flow is considered with Mach numbers usually well below one. Thus, the parameter range is limited both in ka, related to kb and η, and M. In the presence of mean flow the acoustic scattering properties in general and the end correction in particular is not only dependent on ka and M, but also on the Strouhal number, i.e., the quotient between the Helmholtz number and the Mach number, as concluded in Sec. III. Note that with the mentioned restrictions on the values of ka and M, the Strouhal number can still take any value from zero to infinity. The value of the area expansion ratio, which influence the end correction also without flow, is important for the degree of flow-acoustic coupling as well.

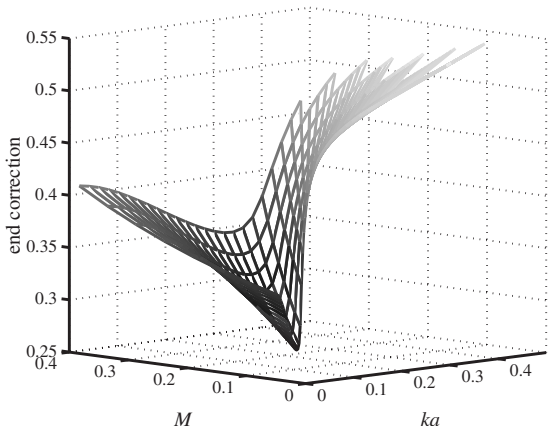


FIG. 5. The end correction as a function of Mach number and Helmholtz number, for an area expansion ratio of $\eta = S_1/S_2 = 0.35$.

A. Influence of Strouhal number

The rate of influence of each of the three parameters, ka, M, and *St*, is varying depending on the parameter combination. In particular, different behaviors are identified for large, small, and intermediate Strouhal numbers, respectively, both in experimental and theoretical results. The different behaviors depending on Strouhal number imply that different mechanisms are involved in the three regimes. This trend is already observed for the reflection and transmission properties of the area expansion,[1,2] as well as for an open pipe termination.[8,19,21] However, the effect of interaction for Strouhal number of order 1 appears to be particularly strong for the end correction. To illustrate the dependence on the Mach number and the Helmholtz number separately, the end correction for an interval of ka and M is computed and shown as a surface plot in Fig. 5 and in a contour plot in Fig. 6. The area ratio is 0.35. A straight line from the origin represents a constant Strouhal number. From the graphs the end correction for very small Helmholtz numbers shows to

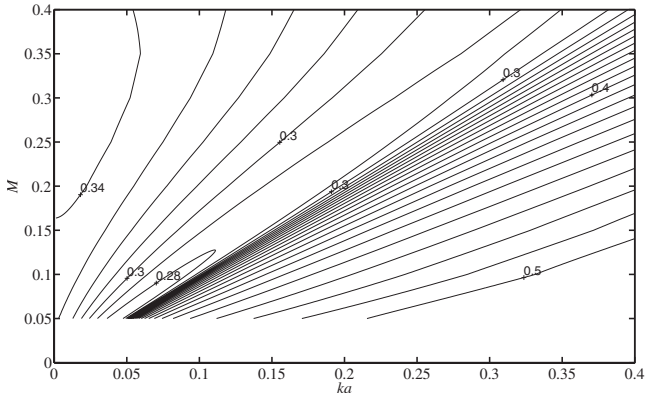


FIG. 6. A contour plot of the value of the end correction as a function of Mach number and Helmholtz number, area expansion ratio $\eta = S_1/S_2 = 0.35$. The difference between adjacent lines is 0.01.

be rather constant, and in the interval 0.33–0.35, a result that is in accordance with the quasi stationary expression given at the end of this section. The minimum value of the end correction in the plot varies between 0.27 and 0.29 and occurs at an almost constant Strouhal number of about 0.85. From the contour plot, Fig. 6, a rapid variation is shown close to the Strouhal number of order 1 where the end correction is strongly dependent on both M and ka. On the whole, these results indicate that the end correction, for small enough M and ka, is to a great extent determined by the value of the Strouhal number. Specifically, it is ambiguous to talk about a value for small M and ka, as this criterion does not correspond to a specific value or a trend for the end correction.

In the context of sound waves incident on a point of flow separation, the Strouhal number represents the relationship between the typical frequencies, i.e., the inverse of the time scales, of the acoustic field and the flow field. For large Strouhal numbers the acoustic time scale is much shorter than that of the flow field, and convection is the main effect of the flow on the acoustic properties. This limit, when St tends to infinity, is in the following called the *acoustic limit*. When the Strouhal number tends to zero, the acoustic field is almost stationary in comparison with the unsteady flow field, and this limit is here denoted the *hydrodynamic limit*. In the region where the Strouhal number is around 1, the time scales of the two fields, the acoustic and the hydrodynamic, are of the same order. The distinction between the two fields becomes less pronounced, and this indicates that energy can flow from one field to another, i.e., there is interaction between the acoustic field and the flow field. It is interesting to note that in this region it is the Strouhal number rather than the Helmholtz number and the Mach number that determines the value of the end correction, indicating that the interaction effects are indeed dominating.

B. Influence of the area expansion ratio

An initial study of the effects of variation in area expansion ratio was presented by Boij,[32] where area ratios 0.35 and 0.5 were compared, showing a negligible influence of flow in the 0.5 case. An extended study is presented here. The end correction for a fixed Mach number of 0.1 is calculated for a number of area expansion ratios between 0.1 and 0.9. The result is presented as function of St_b based on the width of the downstream duct, Fig. 7, as this Strouhal number seem to be the most suitable to predict the minimum. It is observed that the behavior is quite different for the parameter regimes $\eta<0.5$ and $\eta>0.5$. At $\eta=0.5$, the value in the hydrodynamic and the acoustic regimes is almost constant. For area expansions $\eta<0.5$, the extent of the regime for the hydrodynamic limit seem to be determined by St_b, whereas the regime where the acoustic low frequency limit is valid seems to be determined by the Strouhal number St_a.

For $0.001<\eta<0.5$ and $M=0.1$, the minimum end correction is achieved for $2.1<St_b<3.0$, and the minimum is $0.26<\Delta l/a<0.33$, providing an estimate of the value in this region. Note that the results in Fig. 6 indicate a weak variation in minimum end correction with increasing Mach number. For area ratios $\eta>0.5$ the Strouhal number of the minimum seems to have a different behavior. The value in the intermediate regime is here a maximum, at around the same value as the minimum for smaller area ratios.

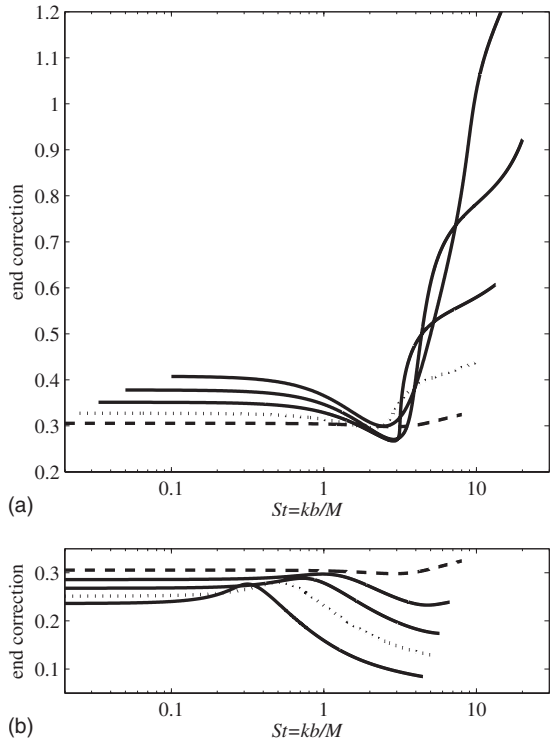


FIG. 7. The end correction as a function of the Strouhal number based on the downstream duct. Varying area ratio $\eta=S_1/S_2$ and $M=0.1$. Upper graph (from above) $\eta=S_1/S_2=0.1,0.2,0.3,0.4$ (dotted line), 0.5 (dashed line) Lower graph $\eta=S_1/S_2=0.5$ (dashed line), 0.6, 0.7, 0.8 (dotted line), 0.9. Cut-on for higher order modes occur at $St=\pi/M$.

A low frequency, small Strouhal number version of the theory described in Ref. 1 yields an expression for the end correction, as defined in Eq. (12), as

$$\frac{2\ln 2}{(1+\eta)\pi}+\frac{4\eta(1-\eta)\ln 2}{(1+\eta)^2\pi}M + \frac{(1-2\eta+23\eta^2-10\eta^3-4\eta^4)\ln 2}{(1+\eta)^3\pi}M^2+O(M^3). \quad (14)$$

In the hydrodynamic limit and for η tending to zero, i.e., a flanged duct termination, the end correction as defined here tends to 0.44 with a correction of order M^2. Note that in this limit, the behavior is different from that of the "acoustic" low frequency limit, i.e., large Strouhal numbers, and the comparison with a cylindrical geometry is not straightforward. However, the expression is useful for verification of simulations for two-dimensional geometries.

C. Effects of wave number and modal properties

To get further understanding of the variations in end correction, the vortex sheet model is studied. In the model, the pressure field in the duct is described as a modal sum. This formulation allows for a physical interpretation of the modes and corresponding wave numbers, such that the modes in the solution can be divided into three different wave type categories, acoustic plane waves, hydrodynamic

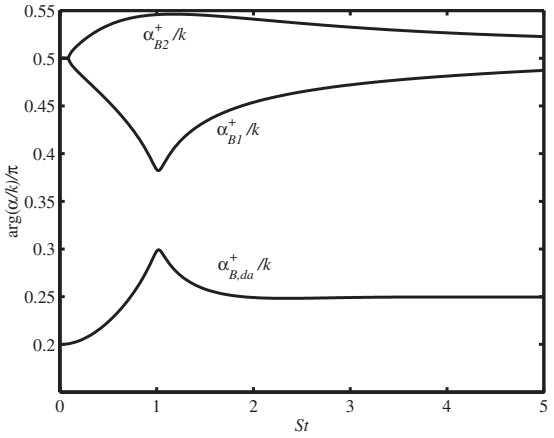


FIG. 8. The phase of the normalized wave numbers for $\eta=S_1/S_2=0.35$ and $M=0.1$. The corresponding modes are the first higher order mode $(B1)$, the second higher order mode $(B2)$, and the damped hydrodynamic mode (B,da).

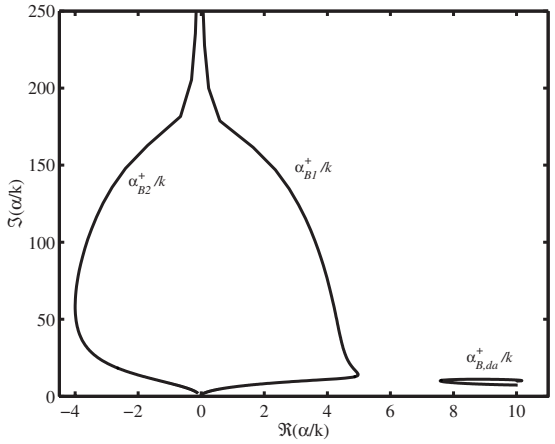


FIG. 9. Normalized wave numbers α/k for $\eta=S_1/S_2=0.35$ and $M=0.1$, with increasing ka. The location for $St_a=1$ is indicted by ●. The corresponding modes are the first higher order mode $(B1)$, the second higher order mode $(B2)$, and the damped hydrodynamic mode (B,da). Notation as in Fig. 8.

waves (one growing and one damped), and an infinite number of higher order modes that are evanescent for low frequencies.[1] Characteristic for the hydrodynamic waves is that they propagate with a speed proportional to the mean flow speed, corresponding to a wave number proportional to $1/M$. The interpretation of the modes, valid for large enough Strouhal numbers, also shows that in a certain Strouhal number region the modal characteristics diverge from this categorization. Connections between the variations in end correction and in wave number characteristics are therefore studied.

The theoretical expression for the end correction explicitly contains the wave numbers of the higher order modes in the larger duct, such that the influence of the modal wave numbers on the value of the end correction may be investigated. From formulas presented in earlier papers,[1,2] it can be shown that the argument of the axial wave numbers of the higher order non-propagating modes in the larger duct are the wave numbers that influence the end correction. The variation in end correction with Strouhal number is mainly associated with the variation in these wave numbers, or rather the phase of the wave numbers of the acoustic higher order modes of the larger duct. The hydrodynamic wave numbers, however, do not have an explicit influence on the value of the end correction, but it will show that the presence of the vortex sheet has indeed an impact on the results for the end correction. The variation with St_a, ka, and M related to the variations in the end correction to the flow effects on the wave numbers are now investigated.

Studies of the wave numbers show that for all area expansion ratios, the higher order modes are affected by the presence of the vortex sheet, in particular, around a Strouhal number of order 1. For an area expansion ratio of $\eta=0.35$, the two first higher order modes are the ones significantly affected by the flow velocity profile with the vortex sheet, and the phase of these is displayed in Fig. 8. The phase of all three wave numbers varies significantly in the same region as the end correction varies, and the phase of α^+_{B1} and $\alpha^+_{B,da}$ are closest for $St_a=1$. The magnitude of the wave numbers of the first higher order mode and the damped hydrodynamic mode are in fact equal when $St_a=1$. The solution of the wave numbers clearly changes character at this Strouhal number.

To further illustrate the wave number behavior, the two first higher order wave numbers, normalized by k, are shown in Fig. 9 and in more detail in Fig. 10, together with the wave number of the damped hydrodynamic mode. Large deviations from the limiting values are observed for this area expansion ratio. It is observed that for these parameter values and St_a around 1, the location of the wave number of the first higher order mode, α^+_{B1}, is very close to the location of the hydrodynamic wave number $\alpha^+_{B,da}$. In a similar fashion, the location of α^+_{B2} deviates from the limiting values, but rather in the opposite direction. Thus, the wave number of the first higher order mode has properties that resembles those that characterize a hydrodynamic wave. It is worth noting that the hydrodynamic wave numbers turn from the small Strouhal number behavior to a large Strouhal number behavior around $St_a=1$.

The hydrodynamic wave numbers do not affect the end correction explicitly as mentioned earlier. The presence of

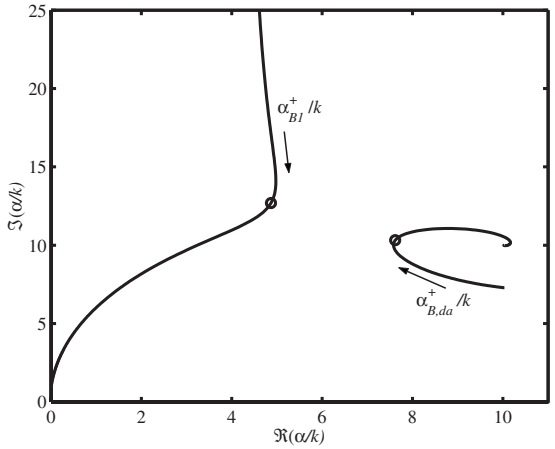


FIG. 10. Normalized wave numbers α/k for $\eta=S_1/S_2=0.35$ and $M=0.1$, with ka increasing in the direction of the arrow. The ○ indicates the wave number location at $St_a=1$. The corresponding modes are the first higher order mode $(B1)$ and the damped hydrodynamic mode (B,da).

the vortex sheet, however, affects the higher order modes, in this case α_{B1}^{+}, for St_a close to 1. In this regime, this higher order mode changes its properties from that typical of acoustic behavior to become more of a "hydrodynamic" type of mode. The effect on the end correction is a decrease, i.e., that the acoustic inertia of the mass at the area change decreases.

There are some additional interesting remarks concerning the nearness of the first higher order mode and the damped hydrodynamic mode for St_a close to 1. The inverse of the distance between the wave numbers is a measure of the strength of the coupling, the smaller the distance the stronger the coupling. If the two wave numbers are equal the dispersion relation has a double root; however, this only occurs for complex k. Such merge of the wave number of the growing hydrodynamic mode and a higher order mode related to upstream propagation would correspond to an absolute instability. The conditions for this are further explored by Nilsson and Brander.[24]

The influence of the mode shapes and modal properties are left for a coming publication, but a discussion on the topic is found in Refs. 3 and 33. It is noted that depending on the area expansion ratio, different higher order modes may be affected by the presence of the vortex sheet. From asymptotic expressions for the wave numbers of the higher order modes,[1,34] it can be shown that only the two first higher order wave numbers may be close to the hydrodynamic wave number for a value of St_a of the order of one when the area expansion ratio is not too small nor too large. Also, for the end correction it is the deviation in phase of the wave numbers that is important, so the larger the damping—imaginary part—a modal wave number has, the larger variation in the real part is required to give a significant effect on the phase.

D. Effects of the edge condition

Another interesting parameter in the modeling of the sharp edge is the so called edge condition, i.e., the criterion set to the edge of the expansion when solving the scattering problem. The results presented in this paper are all calculated applying the unsteady Kutta condition. This edge condition states that the velocity of the acoustic field at the edge is parallel to the splitter plate surface. The unsteady Kutta condition corresponds to a case when the free shear layer downstream of the edge is unstable, which in practice is true for low values of the shear layer Strouhal number. However, for higher Strouhal numbers the shear layer will become stable. In the vortex sheet model, the instability is present through the growing hydrodynamic mode in the solution. One way of modeling the effect of a stable shear layer is to, through a change in the edge condition, suppress the excitation of this growing hydrodynamic mode. The resulting edge condition is termed the relaxed Kutta condition. This method is applied for calculations of the reflection and transmission coefficients in an earlier paper by Boij and Nilsson.[2]

The end correction with the two different edge conditions is calculated, and the result is shown in Fig. 11. From Fig. 11, it is observed that in the region where the relaxed Kutta condition should be valid, above, say, $St=5$, the two models are rather similar. For small and vanishing Strouhal

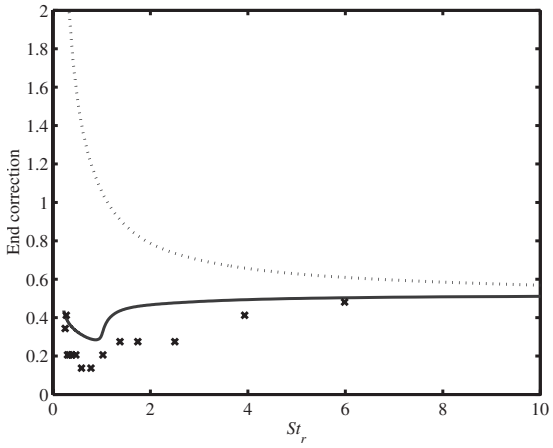


FIG. 11. The end correction as a function of the Strouhal number with fixed $ka=0.11$ and area expansion ratio $\eta=S_1/S_2=0.35$. — unsteady Kutta condition, $(\cdots)$ relaxed Kutta condition, and $(\times)$ experimental results.

number, the result with the Kutta condition is far superior to the version with the relaxed Kutta condition, a result in accordance with the physics of the shear layer instability. This result implies that the excitation of the growing hydrodynamic wave is an important part of the near field, and that the instability effects are well reproduced through the vortex sheet model. In this region, the choice of edge condition is important. It is also concluded that the thickness of the shear layer at the edge and the resulting dynamic properties may have a great influence on the end correction.

V. SUMMARY AND CONCLUSIONS

A variation with the Strouhal number of the reactive part of the impedance or reflection properties of a sudden area expansion in a flow duct is investigated. The results previously presented in Refs. 1 and 25 for the end correction of a two-dimensional, rectangular duct, were extended to a wider Strouhal number, St, regime and the influence of the area expansion ratio was studied. The end correction show clear difference in behavior in three different regions: the hydrodynamic regime for small St, the acoustic regime for large St, and the intermediate regime where St $O(1)$. The hydrodynamic and the acoustic limit correspond to results presented for open pipe terminations. However, the behavior in the intermediate regime is distinctly different and not a mere transition region. For area ratios where the smaller duct is 50% or less of the larger duct, a dip is observed in the end correction for intermediate Strouhal numbers, the variation here is sometimes of the same order of magnitude as the end correction in the acoustic limit. The results indicate that the minimum end correction is achieved for a Strouhal number, based on the wider duct, of kb/M around 2–3 with a minimum value varying around 0.3. The value seems marginally influenced by the Mach number. Also, analysis with the vortex sheet model shows that in the intermediate Strouhal number regime, the hydrodynamic mode and a higher order nonpropagating mode are very similar for large area expansions. For an area expansion ratio of 0.5, no difference is observed across the three different regions. However, for area expansion ratios where the smaller duct is more than 50% of the

larger a different behavior is observed. In the hydrodynamic limit, i.e., for small Strouhal numbers, a low frequency, quasistationary approximation is presented, that predicts a finite value for the end correction that is proportional to the Mach number and tends to 0.44 when the larger duct dimension tend to infinity with a correction factor or order Mach number squared.

Finally, the importance of the instability of the shear layer downstream of the expansion is investigated through the use of different edge conditions. The conclusion is that excitation of the unstable shear layer has an important impact on the acoustic properties of the sudden area expansion.

To conclude, the end correction may be strongly influenced by duct mean flow and the presence of the unstable shear layer downstream of an area expansion. The influence, however, is dependent on the conditions such as frequency, flow speed, area ratio, and shear layer thickness, making the flow acoustic interaction effects on the end correction, and other reactive parts of the scattering coefficients, nonintuitive phenomena that is difficult to predict a-prior. An interesting future work is to experimentally investigate the connection between the area expansion ratio and the degree of influence from the flow, and to study the low frequency behavior in the limits of large and small Strouhal numbers.

ACKNOWLEDGMENT

This work was funded by the Swedish Research Council Grant No. 621-2003-3741.

[1]S. Boij and B. Nilsson, "Reflection of sound at area expansions in a flow duct," J. Sound Vib. **260**, 477–498 (2003).
[2]S. Boij and B. Nilsson, "Scattering and absorption of sound at flow duct expansions," J. Sound Vib. **289**, 577–594 (2006).
[3]S. Boij, "Acoustic scattering in ducts and influence of flow," Ph.D. thesis, KTH, Stockholm (2003).
[4]D. W. Bechert, U. Michel, and E. Pfizenmaier, "Experiments on the transmission of sound through jets," AIAA Paper No. 77–1278.
[5]D. W. Bechert, "Sound absorption caused by vortex shedding, demonstrated with a jet flow," J. Sound Vib. **70**, 389–405 (1980).
[6]M. S. Howe, "The dissipation of sound at an edge," J. Sound Vib. **70**, 407–411 (1980).
[7]R. M. Munt, "The interaction of sound with a subsonic jet issuing from at semi-infinite cylindrical pipe," J. Fluid Mech. **83**, 609–640 (1977).
[8]S. W. Rienstra, "A small Strouhal number analysis for acoustic wave-jet flow-pipe interaction," J. Sound Vib. **86**, 539–556 (1983).
[9]A. M. Cargill, "Low frequency acoustic radiation from a jet pipe—a second order theory," J. Sound Vib. **83**, 339–354 (1982).
[10]R. M. Munt, "Acoustic transmission properties of a jet pipe with subsonic jet flow: I. The cold jet reflection coefficient," J. Sound Vib. **142**, 413–436 (1990).
[11]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[12]G. Ajello, "Mesures acoustiques dans les guides d'ondes en présence d'écoulement: mise au point d'un banc demesure, application à des discontinuités (Acoustic measurements in waveguides with the prescence of flow: Design and adaption of an experimental setup, for discontinuities)," These de doctorat du l'Université du Maine (1997).
[13]D. Ronneberger, "The acoustical impedance of holes in the wall of flow ducts," J. Sound Vib. **24**, 133–150 (1972).
[14]A. Cummings, "The effects of grazing turbulent pipe-flow on the impedance of an orifice," Acustica **61**, 233–242 (1986).
[15]J. W. S. Baron Rayleigh, *The Theory of Sound* (Macmillan, London, 1877).
[16]H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," Phys. Rev. **73**, 383–406 (1948).
[17]J. W. Miles, "The analysis of plane discontinuities in cylindrical tubes. Part II," J. Acoust. Soc. Am. **17**, 272–284 (1946).
[18]F. Karal, "The analogous acoustical impedance for discontinuities and constrictions of circular cross section," J. Acoust. Soc. Am. **25**, 327–334 (1953).
[19]M. C. A. M. Peters, A. Hirschberg, A. J. Reijnen, and A. P. J. Wijnands, "Damping and reflection coefficient measurements for an open pipe at low Mach number and low Helmholtz number," J. Fluid Mech. **256**, 499–534 (1993).
[20]M. S. Howe, "Attenuation of sound in a low Mach number nozzle flow," J. Fluid Mech. **91**, 209–229 (1979).
[21]S. Allam and M. Åbom, "Full plane wave decomposition using microphone arrays in ducts," J. Sound Vib. **292**, 519–534 (2006).
[22]K. S. Peat, "The acoustical impedance at discontinuities of ducts in the presence of a mean flow," J. Sound Vib. **127**, 123–132 (1988).
[23]P. O. A. L. Davies, "Practical flow duct acoustics," J. Sound Vib. **124**, 91–115 (1988).
[24]B. Nilsson and O. Brander, "The propagation of sound in cylindrical ducts with mean flow and bulk-reacting lining—I. Modes in an infinite duct," J. Inst. Math. Appl. **26**, 269–298 (1980).
[25]G. Kooijman, P. Testud, Y. Auregan, and A. Hirschberg, "Multimodal method for scattering of sound at a sudden area expansion in a duct with subsonic flow," J. Sound Vib. **310**, 902–922 (2002).
[26]D. Ronneberger, "Theoretische und experimentelle Untersuchung der Schallausbreitung durch Querschnittssprünge und Lochplatten in Strömungskanälen (Theoretical and experimental investigation of sound propagation at sudden area changes and diaphragms in flow ducts)," DFG-Abschlussbericht, Drittes Physikalisches Institut der Universität Göttingen (1987).
[27]Y. Auregan, A. Debray, and R. Starobinski, "Low frequency sound propagation in a coaxial cylindrical duct: Application to sudden area expansions and to dissipative silencers," J. Sound Vib. **243**, 461–473 (2001).
[28]A. N. Norris and I. C. Sheng, "Acoustic radiation from a circular pipe with an infinite flange," J. Sound Vib. **135**, 85–93 (1989).
[29]D. G. Crighton, "The Kutta condition in unsteady flow," Annu. Rev. Fluid Mech. **17**, 411–445 (1985).
[30]S. Boij, "Mean flow effects on the acoustics of silencers," Licentiate thesis, KTH, Stockholm (1999).
[31]S. Dequand, S. J. Hulshoff, Y. Aurégan, J. Huijnen, R. ter Riet, L. J. van Lier, and A. Hirschberg, "Acoustics of 90 degree sharp bends. Part I: Low frequency acoustical response," Acta. Acust. Acust. **89**, 1025–1037 (2003).
[32]S. Boij, "Parameter dependence of flow acoustic interaction," Proceedings of the Second Conference Math Modelling of Wave Phenomena [AIP Conf. Proc. **834**, 100–108 (2006)].
[33]S. Boij, "An analysis of the acoustic energy in a flow duct with a vortex sheet," Proceedings of the Third Conference Math Modelling of Wave Phenomena [AIP Conf. Proc. **1106**, 130–139 (2009)].
[34]B. Nilsson, "Instability waves and causality," Mathematical modelling in physics, engineering and cognitive sciences, **5**, 431–449 [Proceedings of the conference "Foundations of probability and physics - 2," June 2002 (Växjö University Press, Sweden, 2003)].

Two-dimensional model of low Mach number vortex sound generation in a lined duct

S. K. Tang[a] and C. K. Lau[b]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, China

(Received 22 April 2009; revised 17 June 2009; accepted 5 July 2009)

The sound generated by a vortex moving across a duct section lined with porous materials and the corresponding vortex dynamics are studied numerically in the present investigation. The combined effects of the effective fluid density, the flow resistance, the length, and the thickness of the porous linings on the vortex dynamics and sound generation are examined in detail. Results show that stronger sound radiation will take place when the length and the thickness of the porous linings are increased or when the effective fluid density is reduced. The flow resistance can only result in stronger sound radiation within a range whose width depends on the abovementioned system parameters. Such sound amplification cannot be achieved when the initial vortex height gets closer and closer to the duct centerline. The present results also indicate the strong correlation between vortex acceleration and the sound radiation under the actions of the porous linings.
 [DOI: 10.1121/1.3192332]

PACS number(s): 43.28.Ra, 43.50.Nm, 43.50.Gf [JWP] Pages: 1005–1014

I. INTRODUCTION

Commercial buildings in a sub-topical city nowadays are very heavily serviced. A significant portion of the noise inside these buildings comes from the air conditioning and ventilation systems. The low Mach number turbulent flows inside the ductwork are also sound producing, especially when they interact with obstacles.[1,2] These aerodynamic noises are of low frequency and they propagate into the occupancy zones together with the system noises through the ductwork.

In order to achieve a reasonable level of indoor noise, dissipative silencers and acoustical linings,[3] which consist of porous materials like fiberglass, are commonly used to attenuate the system noise before it reaches the occupancy zones. There are studies on modeling the sound attenuation by dissipative silencers in the presence of a mean flow (for instance, Cummings and Chang[4] and Peat and Rathi[5]) and on the design of high attenuation silencers (for instance, Selamat *et al.*[6]). However, there are evidences showing that the low Mach number turbulent flow can interact with an absorbent liner to produce sound.[7,8] The low frequency characteristic of this noise makes it very difficult to attenuate by conventional methods. It is therefore important to study how this aerodynamic noise will reduce the performance of dissipative silencers and the wall linings.

Flow turbulence is very difficult to model analytically. However, the low Mach number flow condition inside the building air ductwork makes it possible to model the turbulent eddies as discrete vortices moving in an incompressible flow, which can then be handled analytically by the potential theory.[9] Once the motions of the vortices are obtained, the vortex sound theory[10] can basically be used to estimate the vortex sound so generated. Though the vortex is a drastic simplification of the flow turbulence, this semi-analytical approach has attracted the attention of many researchers as it is expected that this vortex analogy can provide insights into the aeroacoustics of more complicated low Mach number turbulent flows. Typical examples include the works of Cannell and Ffowcs Williams,[11] Crighton,[12] and Obermeier.[13] More examples can be found in Ref. 14.

Many of the works in the existing literature deal with vortex sound in the presence of a rigid boundary. Recently, Tang[15] worked out the vortex dynamics in the presence of porous surfaces. The authors have also extended the vortex sound study to include the porous wedge and cylinder.[16,17] Lau and Tang[17] showed that the dipole sound generation under the effect of the porous cylinder can be stronger than that generated in the presence of a rigid cylinder. The more recent work of the authors examines the vortex sound radiation under the influence of a porous lining in an opened space.[18]

In the present investigation, the vortex sound generation inside an infinitely long two-dimensional rigid wall duct with porous linings of finite length on both sides of the duct is studied. This configuration is analogous to the situation of a lined duct in the air conditioning and ventilation ductwork. It is hoped that the present study could enhance the general understanding on vortex sound generation and be able to provide insights for further development of aeroacoustic modeling of wall linings and dissipative silencers.

II. THEORETICAL DEVELOPMENT

Figure 1 illustrates the schematics and the essential nomenclature adopted in the present study. An inviscid vortex initially at a distance very far away from the porous materials propagates toward the latter under the effect of the rigid duct walls in an incompressible fluid medium. Crighton[12] showed explicitly that the near field incompressible solution can be

[a]Author to whom correspondence should be addressed. Electronic mail: besktang@polyu.edu.hk
[b]Present address: Ove Arup & Partners, Level 5, Festival Walk, Kowloon, Hong Kong, People's Republic of China.

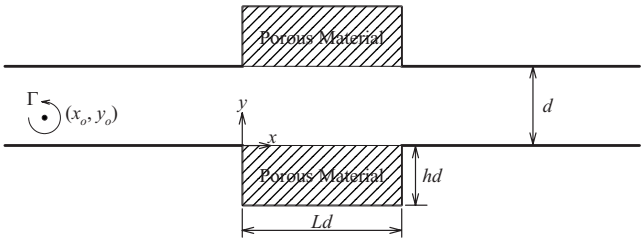


FIG. 1. Schematic of the vortex-lined duct system.

used to estimate the far field acoustic radiation through a matching technique in the low Mach number condition. This is also the approach adopted in the present study.

All length scales in the present study are normalized using the width of the duct d. The time and velocities are normalized by d^2/Γ and Γ/d, respectively. As in the previous studies of the authors,[15–18] the air density inside the duct is denoted by ρ while the properties of the porous materials are characterized using the effective density ρ_e and the flow resistance R_f inside its lattice.[19] The flow resistance is normalized by $\rho\Gamma/d^2$. Under the low Mach number condition, the flow inside the porous material, which is expected to be very weak, is incompressible. To simplify the analysis, the porous material is assumed to be a continuum and thus the potential theory applies.

Denoting the streamfunctions within the duct $(0 \le y \le 1)$, in the upper porous layer $(1 \le y \le 1+h, 0 \le x \le L)$, and in the lower porous layer $(-h \le y \le 0, 0 \le x \le L)$ as ψ, ψ_{pu}, and ψ_{pl}, respectively, one finds that

$$\nabla^2 \psi = -\delta(x - x_o)\,\delta(y - y_o), \tag{1}$$

and

$$\nabla^2 \psi_{pu} = \nabla^2 \psi_{pl} = 0, \tag{2}$$

where ∇^2 and δ are the Laplacian operator and delta function, respectively. Since the normal fluid velocities at the interfaces between the porous material and the rigid walls vanish, one observes that

$$\left.\frac{\partial \phi_{pl}}{\partial x}\right|_{x=0} = \left.\frac{\partial \psi_{pl}}{\partial y}\right|_{x=0} = \left.\frac{\partial \phi_{pl}}{\partial x}\right|_{x=L} = \left.\frac{\partial \psi_{pl}}{\partial y}\right|_{x=L} = -\left.\frac{\partial \psi_{pl}}{\partial x}\right|_{y=-h} = 0, \tag{3}$$

and

$$\left.\frac{\partial \phi_{pu}}{\partial x}\right|_{x=0} = \left.\frac{\partial \psi_{u}}{\partial y}\right|_{x=0} = \left.\frac{\partial \phi_{pu}}{\partial x}\right|_{x=L} = \left.\frac{\partial \psi_{pu}}{\partial y}\right|_{x=L} = -\left.\frac{\partial \psi_{pu}}{\partial x}\right|_{y=1+h} = 0, \tag{4}$$

where ϕ_{pl} and ϕ_{pu} denote the flow potentials within the lower and upper porous layers, respectively. It is straightforward to show from Eqs. (2)–(4) that

$$\psi_{pl} = \sum_{n=1}^{\infty} A_n e^{\alpha_n h} \sin(\alpha_n x)\sinh[\alpha_n(h + y)], \tag{5}$$

and

$$\psi_{pu} = \sum_{n=1}^{\infty} B_n e^{\alpha_n (1+h)} \sin(\alpha_n x)\sinh[\alpha_n(1 + h - y)], \tag{6}$$

where n is a non-zero integer, $\alpha_n = n\pi/L$, and A_n and B_n the mode magnitudes. The Fourier transform with respect to x of Eq. (1) gives

$$\Psi = \int_{-\infty}^{\infty} \psi e^{ikx} dk = \begin{cases} G_1 e^{-|k|y} + G_2 e^{|k|y}, & 0 \le y \le y_o \\ H_1 e^{-|k|y} + H_2 e^{|k|y}, & y_o \le y \le 1, \end{cases} \tag{7}$$

where G_1, G_2, H_1, and H_2 are functions of k. The continuity of Ψ and the vorticity jump $\partial\Psi/\partial y$ at $y = y_o$ lead to

$$H_1 - G_1 = \frac{1}{2|k|} e^{ikx_o + |k|y_o} \text{ and } G_2 - H_2 = \frac{1}{2|k|} e^{ikx_o - |k|y_o}. \tag{8}$$

The continuity of normal fluid velocity at the porous layer surface at $y=0$, $0 \le x \le L$ gives

$$-\left.\frac{\partial \psi}{\partial x}\right|_{y=0} = -\left.\frac{\partial \psi_{pl}}{\partial x}\right|_{y=0} \Rightarrow \Psi|_{y=0} = \Psi_{pl}|_{y=0}. \tag{9}$$

The same condition at the upper porous layer surface at $y = 1$, $0 \le x \le L$ gives

$$\Psi|_{y=1} = \Psi_{pu}|_{y=1}. \tag{10}$$

It follows from Eqs. (9) and (10) that

$$G_1 = \sum_{n=1}^{\infty} \alpha_n A_n e^{\alpha_n h} \sinh(\alpha_n h) \frac{(-1)^n e^{ikL} - 1}{k^2 - \alpha_n^2} - G_2, \tag{11}$$

and

$$H_1 = \sum_{n=1}^{\infty} \alpha_n B_n e^{\alpha_n (1+h)} \sinh(\alpha_n h) \frac{(-1)^n e^{ikL} - 1}{k^2 - \alpha_n^2} e^{|k|} - H_2 e^{2|k|}. \tag{12}$$

G_1, G_2, H_1, and H_2 can thus be obtained in terms of the mode magnitudes α_n, h, L, and k by solving Eqs. (8), (11), and (12) together and are shown in the Appendix. It can then be shown with the use of Eq. (7) that

$$\Psi = \frac{e^{ikx_o}\sinh(|k|y)\sinh(|k|(1-y_o))}{|k|\sinh(|k|)} + \frac{1}{\sinh(|k|)}\sum_{n=1}^{\infty}\left\{\alpha_n e^{\alpha_n h}\sinh(\alpha_n h)\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}[A_n\sinh(|k|(1-y)) + e^{\alpha_n}B_n\sinh(|k|y)]\right\}. \quad (13)$$

Following the previous study of the authors,[15,18] the vortex velocities are

$$\dot{x}_o = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\partial}{\partial y}\left(\Psi - \frac{1}{2|k|}\exp(ikx_o + |k|y - |k|y_o)\right)e^{-ikx}dk\Big|_{x=x_o,y=y_o}$$
$$= \frac{1}{4}\cot(\pi y_o) - \sum_{n=1}^{\infty}\alpha_n e^{\alpha_n h}\sinh(\alpha_n h) \times \int_0^{\infty}\frac{(-1)^n\cos[k(L-x_o)] - \cos(kx_o)}{\pi(k^2-\alpha_n^2)\sinh(k)/k} \times (A_n\cosh[k(1-y_o)] - B_n e^{\alpha_n}\cosh(ky_o))dk \quad (14)$$

and

$$\dot{y}_o = -\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\partial}{\partial x}\left(\Psi - \frac{1}{2|k|}\exp(ikx_o + |k|y - |k|y_o)\right)e^{-ikx}dk\Big|_{x=x_o,y=y_o}$$
$$= -\sum_{n=1}^{\infty}\alpha_n e^{\alpha_n h}\sinh(\alpha_n h) \times \int_0^{\infty}\frac{(-1)^n\sin[k(L-x_o)] + \sin(kx_o)}{\pi(k^2-\alpha_n^2)\sinh(k)/k} \times (A_n\sinh[k(1-y_o)] + B_n e^{\alpha_n}\sinh(ky_o))dk, \quad (15)$$

where use has been made of the formulas given by Gradshteyn and Ryzhik[20] and "·" denotes time differentiation. The continuity of pressure at the porous layer surfaces requires that

$$\frac{\partial}{\partial t}\left(\frac{\partial\psi}{\partial y}\Big|_{y=0}\right) = \eta\frac{\partial}{\partial t}\left(\frac{\partial\psi_{pl}}{\partial y}\Big|_{y=0}\right) + R_f\frac{\partial\psi_{pl}}{\partial y}\Big|_{y=0} \quad \text{and}$$

$$\frac{\partial}{\partial t}\left(\frac{\partial\psi}{\partial y}\Big|_{y=1}\right) = \eta\frac{\partial}{\partial t}\left(\frac{\partial\psi_{pu}}{\partial y}\Big|_{y=1}\right) + R_f\frac{\partial\psi_{pu}}{\partial y}\Big|_{y=1}, \quad (16)$$

where $\eta=\rho_e/\rho$. The application of inverse Fourier transform of Eq. (7) suggests that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|k|(\dot{G}_2-\dot{G}_1)e^{-ikx}dk = \sum_{n=1}^{\infty}(\eta\dot{A}_n + R_f A_n)\alpha_n e^{\alpha_n h}\sin(\alpha_n x)\cosh(\alpha_n h), \quad (17)$$

and

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}|k|(\dot{H}_2 e^{|k|} - \dot{H}_1 e^{-|k|})e^{-ikx}dk = -\sum_{n=1}^{\infty}(\eta\dot{B}_n + R_f B_n)\alpha_n e^{\alpha_n(1+h)}\sin(\alpha_n x)\cosh(\alpha_n h). \quad (18)$$

Using the technique of Lau and Tang,[18] the rates of change in the mode magnitudes and the vortex velocities can be estimated from their instantaneous values by solving the simultaneous equations [Eqs. (14), (15), (17), and (18)] with the initial condition $\dot{A}_n=\dot{B}_n=0$, $\dot{y}_o=0$, and $\dot{x}_o=\cot(\pi y_o)/4$. The position of the vortex as well as the mode magnitudes can then be estimated using standard Runge–Kutta method as in the previous study of the authors.[18]

The flow potential due to the presence of the vortex can be found by applying the Cauchy–Rieman principle,[21] which states that

$$\frac{\partial\psi}{\partial y} = \frac{\partial\phi}{\partial x} \Rightarrow \phi = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{i}{k}\frac{\partial\Psi}{\partial y}e^{-ikx}dk + C, \quad (19)$$

where C is a spatial invariant, but may vary with time.

$$\phi = \frac{1}{2\pi}\tan^{-1}\left[\tan\left(\frac{(1-y_o+y)\pi}{2}\right)\tanh\left(\frac{(x-x_o)\pi}{2}\right)\right] + \frac{1}{2\pi}\tan^{-1}\left[\tan\left(\frac{(1-y_o-y)\pi}{2}\right)\tanh\left(\frac{(x-x_o)\pi}{2}\right)\right] + \frac{1}{\pi}\sum_{n=1}^{\infty}\alpha_n e^{\alpha_n h}\sinh(\alpha_n h) \times \int_0^{\infty}\frac{(-1)^{n+1}\sin[k(x-L)] + \sin(kx)}{k^2-\alpha_n^2} \times \frac{A_n\cosh[k(1-y)] - B_n e^{\alpha_n}\cosh(ky)}{\sinh(k)}dk + C. \quad (20)$$

The summation term in Eq. (20) represents the vortex potential induced by fluid motions at the surface of the porous material, while those with the arctangent the vortex potential for infinitely long rigid duct. It can be observed that

$$(\phi - C)|_{x\to\infty} = -(\phi - C)|_{x\to-\infty} \Rightarrow (\dot{\phi} - \dot{C})|_{x\to\infty} = -(\dot{\phi} - \dot{C})|_{x\to-\infty} \tag{21}$$

and

$$\partial\phi/\partial x|_{x\to\infty} = \partial\phi/\partial x|_{x\to-\infty}. \tag{22}$$

The general solution of the flow potential at $|x|\to\infty$, where the duct walls are rigid, is

$$\phi_i = \phi + \Omega x, \tag{23}$$

where Ω is a function of time. The low Mach number vortex motion will produce a low frequency plane wave at $|x|\to\infty$,[11] and thus in the leading order at large $|x|$:

$$\frac{1}{c}\frac{\partial\phi_i}{\partial t} = -\operatorname{sgn}(x)\frac{\partial\phi_i}{\partial x}, \tag{24}$$

where c denotes the ambient speed of sound normalized by Γ/d. It can then be concluded from Eqs. (21)–(24) that $\dot{C}\to 0$, implying that C is also a or nearly a time invariant and is not important in the sound generation.

It can be shown using the formula depicted by Gradshteyn and Ryzhik[20] that the vortex potential for $x\to\infty$ is

$$\phi_\infty = \frac{1}{2}(1 - y_o) + \sum_{n=1,3,5,\ldots}^{\infty}\left[e^{\alpha_n h}(B_n e^{\alpha_n} - A_n)\frac{\sinh(\alpha_n h)}{\alpha_n}\right] + O(e^{-x}). \tag{25}$$

A plane wave is generated in the far field and thus the far field potential ϕ_0 downstream of the lined duct will take the form of

$$\phi_o = \Lambda\exp(-ikx) \tag{26}$$

in the frequency domain. By applying the technique of matched asymptotic expansion,[22] the low frequency inner solution of the far field potential (at $kx\to 0$) must match with the Fourier transform of ϕ_∞ with respect to time. Therefore

$$\Lambda = \int_{-\infty}^{\infty}\phi_\infty e^{-i\omega t}dt. \tag{27}$$

It follows that the far field pressure as $|x|\to\infty$ is

$$p(x,t) = -\frac{1}{2\pi}\frac{\partial}{\partial t}\int_{-\infty}^{\infty}\operatorname{sgn}(x)\Lambda\exp(-ik|x|)e^{i\omega t}d\omega = -\operatorname{sgn}(x)\frac{\partial}{\partial t}\phi_\infty(t - |x|/c). \tag{28}$$

A planar dipole is produced and the far field pressure is normalized by $\rho\Gamma^2/d^2$.

In the present study, the effect of a low Mach number mean flow inside the duct is not considered. However, it has been shown by Ffowcs Williams and Lovely[23] and more recently by Tang *et al.*[24] that the mean flow tends to strengthen the overall sound power radiation. It should be noted that the results of Howe[25] show that the mean flow will induce "jetting" effects at the apertures of a perforated plate which eventually causes the vortex to move toward the plate. Such phenomenon is also expected when the perforated plate is replaced by a piece of porous material. One can anticipate that the jetting effect is weaker for porous material because of the damping from the flow resistance in the complicated lattice of the porous material. However, the anticipated slightly higher vortex transverse velocity will result in stronger sound radiation. The present case therefore represents the minimum vortex sound radiation in a lined duct. The effect of mean flow on the duct sound generation is much more complicated and is left for further investigation.

III. RESULTS AND DISCUSSIONS

In the present study, the vortex is located far upstream of the porous materials initially. Therefore, only the cases where the initial height of the vortex y_{oi} is smaller than 0.5 will be considered as the vortex is stationary if $y_{oi}=0.5$ and will move further upstream if $y_{oi}>0.5$ in the absence of a mean flow with nearly constant speed and thus radiates no sound. One can also notice that many equations involved infinite summations which have to be truncated in the computation as in many previous studies (for instance, Lau and Tang[18]). It is found that the difference of the far field pressure amplitudes for $y_{oi}=0.2$, $L=2$, $h=0.2$, and $\eta=3$ with various R_f computed with five-term summation and ten-term summation is within 1%–2%. Thus, five-term summation is adopted in the present computation. The density ratio η in practice is less than 5 according to Morse and Ingard,[19] while the practice range of R_f is discussed by Lau and Tang.[17] In fact, the value of R_f can vary over a wide range. For a perfectly inviscid fluid, $R_f=0$, but for a nearly rigid material, $R_f\to\infty$.

A. Perfectly inviscid fluid

For a perfectly inviscid fluid, the flow resistance inside the lattice of the porous material vanishes ($R_f=0$). Figure 2(a) shows the paths of the vortex under different combinations of h and η with L fixed at unity and $y_{oi}=0.2$. The pressure releasing effect from the lower porous material results in the downward movement of the vortex. The smaller the effective density η or the increase in h strengthens the pressure releasing effect of the porous materials and thus gives rise to deeper downward bend of the vortex path. As in Ref. 18, the paths will become independent of h when h becomes large (not shown here). One can also observe from Fig. 2(a) that the influence of the porous materials becomes significant at $x\sim -1$. Unlike the case of an infinite plate in an unbound fluid medium,[18] the vortex paths in the present lined duct case are not symmetrical about the vertical middle plane of the porous materials ($x=0.5L$). The vortex attains its minimum height at a location slightly larger than $0.5L$ so that the vortex cannot return to its original height even under the pressure supporting effects of the downstream rigid duct walls. The difference between the original and the final height of the vortex decreases with weaker pressure releasing effect of the porous materials. The vortex therefore propagates with higher speed after moving across the lined section of the duct. Such increase in vortex speed suggests that a net vertical downward force is exerted onto the fluid during such

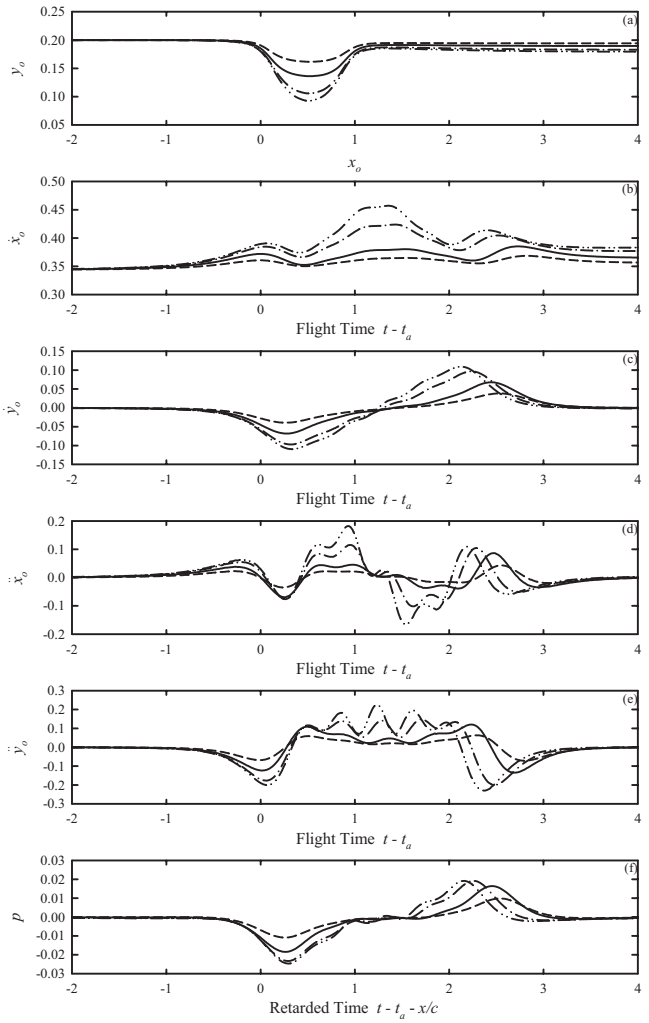


FIG. 2. Effects of pressure-releasing linings on vortex dynamics and sound generation under inviscid condition for $L=1$ and $y_{oi}=0.2$. (a) Vortex path, (b) longitudinal velocity, (c) transverse velocity, (d) longitudinal acceleration, (e) transverse acceleration, and (f) far field sound pressure. ——: $h=0.2$, $\eta=3$; – – –: $h=0.2$, $\eta=5$; –·–: $h=0.4$, $\eta=3$; and –···–: $h=0.8$, $\eta=3$.

maneuver of the vortex, which resulted from the difference in the pressure releasing effects between the upper and lower porous linings.

Figures 2(b) and 2(c) illustrate the time variations in the longitudinal and transverse velocities of the vortex during its interaction with the linings. t_a in the figures and hereinafter denotes the time at which the vortex passes across the upstream edges of the linings ($x=0$ plane). While the transverse vortex velocity turns from negative to positive during the passage of the vortex across the lined portion of the duct, the longitudinal vortex velocity remains higher than the original speed of the vortex throughout the interaction period. Again, the more pressure releasing the lining is, the higher the vortex velocity amplitude resulted. According to Eqs. (25) and (28), such higher vortex velocity, especially its transverse component, will imply stronger sound radiation.

The vortex accelerations contain high frequency components, as shown in Figs. 2(d) and 2(e). The time variations in the far field acoustic pressures shown in Fig. 2(f) follow closely those of the transverse vortex velocities [Fig. 2(c)], suggesting that the unsteady transverse vortex motion is the major mechanism of sound radiation for $L=1$ and $y_{oi}=0.2$. The small amplitude fluctuations embedded in the far field pressure fluctuations are due to the terms $\dot{A}_n$'s and $\dot{B}_n$'s, which reflect the pressure fluctuations on the surfaces of the porous materials and are related to the acceleration of the vortex [cf. Eqs. (14) and (15)]. The results for the cases with $L=1$ and $y_{oi}=0.3$ are very similar to those shown in Fig. 2, but with smaller amplitudes. They are therefore not presented. The small amplitude fluctuations in the far field pressure for $L=1$ and $y_{oi}=0.3$ are too weak to be significant in the overall acoustic radiation. Such kind of small fluctuations is not observed in the flat plate case of Lau and Tang.[18] The presence of the upper duct wall results in non-monotonic longitudinal and transverse variations in the flow field with height in the lined duct section and thus the wrangling observed in Fig. 2.

It can be concluded that the pressure releasing effect of the porous linings, which is uneven on the upper and lower sides of the vortex due to the vortex height, results in a downward force on the fluid which is larger than that experienced by the vortex in the rigid wall section of the duct. This force causes the vortex to move downward and is only partially compensated by the increase in the vortex force[26] as the vortex speed increases. The vortex then accelerates downward until various forces are balanced. The vortex starts going upward under the effect of the downstream rigid duct wall after that. This transverse vortex motion results in a longitudinal vortex force which creates a longitudinal push to the fluid, which propagates into the far field and becomes sound. Such pushing is a plane dipole source. Sound generated from fluctuating vortex forces is also observed in the duct exhaust configuration of Cannell and Ffowcs Williams.[11]

The period of interaction between the vortex and the porous linings increases with L. Figure 3 summarizes the inviscid vortex dynamics and the sound radiation for $L=2$ and $y_{oi}=0.2$. While many of the essential features of Fig. 3 follow those shown in Fig. 2, the longer lengths of the linings result in stronger high frequency fluctuations in the vortex velocities and accelerations. For smaller h and/or larger η, pressure pulses are observed at the instants the vortex moves into and out of the lined duct section (the $x=0$ and $x=L$ planes, respectively). When the porous linings become more pressure releasing, the rates of change in A_n and B_n become dominant in the far field sound radiation. These rates of change are directly related to the vortex accelerations, such that the time fluctuation of far field sound pressure resembles those of the vortex accelerations instead of that of the transverse vortex velocity. Since it is not practical to have $h>1$ and the effect of the thickness of the porous lining on the vortex dynamics has been found to be more or less independent of h for $h>0.8$, the results with $h\geqslant 1$ are not discussed.

B. Combined effects of R_f and η

The flow resistance inside the lattice of the porous material is pressure-supporting. Large flow resistance therefore will produce an effect similar to large η in principle. Figure

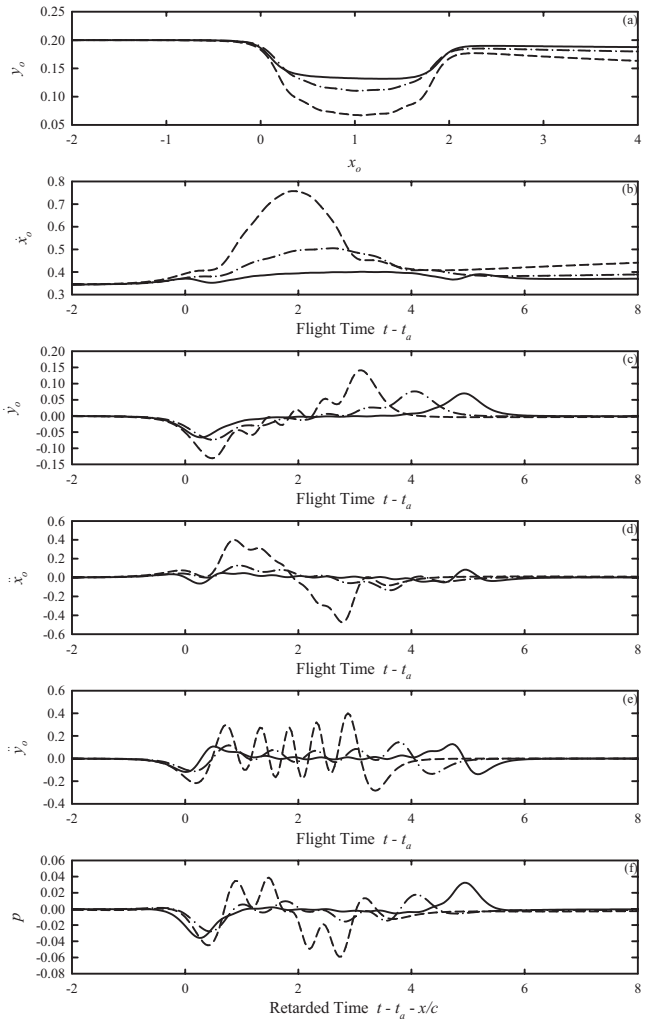


FIG. 3. Effects of pressure-releasing linings on vortex dynamics and sound generation under inviscid condition for $L=2$ and $y_{oi}=0.2$. (a) Vortex path, (b) longitudinal velocity, (c) transverse velocity, (d) longitudinal acceleration, (e) transverse acceleration, and (f) far field sound pressure. ——: $h=0.2$, $\eta=3$; – – –: $h=0.8$, $\eta=3$; and –·–: $h=0.8$, $\eta=5$.

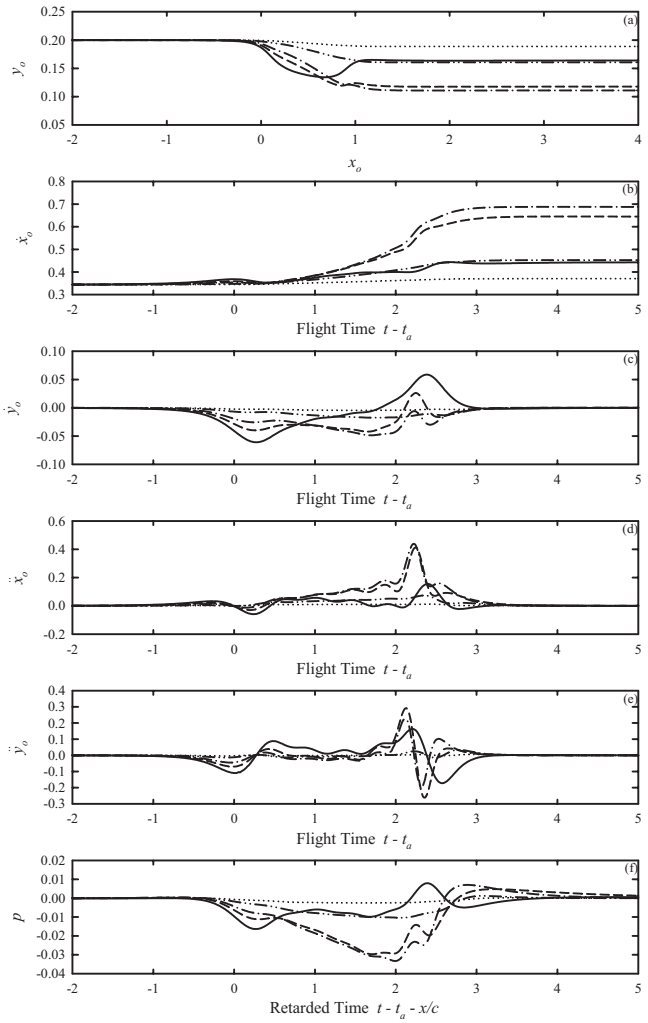


FIG. 4. Effects of R_f on vortex dynamics and sound generation for $L=1$, $y_{oi}=0.2$, $h=0.2$, and $\eta=3$. (a) Vortex path, (b) longitudinal velocity, (c) transverse velocity, (d) longitudinal acceleration, (e) transverse acceleration, and (f) far field sound pressure. ——: $R_f=0.5$; – – –: $R_f=3$; –·–: $R_f=7$; –··–: $R_f=30$; and ···: $R_f=100$.

4 illustrates the vortex dynamics and the far field sound radiation for $y_{oi}=0.2$, $h=0.2$, $L=1$, and $\eta=3$ with increasing R_f. For small R_f, the pressure-releasing effect of the lining dominates and the vortex first bends toward the lower porous lining and then moves upward as it leaves the lined duct section under the pressure-supporting effect of the downstream rigid duct walls [Fig. 4(a)]. However, when R_f increases, the transverse velocity of the vortex is significantly reduced, as shown in Fig. 4(c). The vortex continues to move downward as it accelerates toward the end of the porous linings. The action time of the pressure-supporting downstream rigid duct wall is insufficient to pull the vortex up and can only result in producing a short duration of significant vortex acceleration when the vortex leaves the lined section [Figs. 4(d) and 4(e)]. Further increase in R_f makes the porous linings more pressure-supporting so that the bending of the vortex path, and thus the sound radiation [Fig. 4(f)], becomes less and less significant.

The far field sound time fluctuation patterns illustrated in Fig. 4(f) suggest that the transverse vortex velocity is not the major sound generation mechanism in the presence of a non-vanishing R_f. Strong sound energy radiation is observed within the range $3<R_f<10$ while the strongest radiation takes place when the vortex leaves the lined duct section. The rates of changes in the mode amplitudes A_n and B_n, which control the pressure fluctuations on the porous lining surfaces, play a role at least as significant as the transverse vortex velocity. One can also observe that the amplitude of the sound generated within this R_f range is greater than that for the inviscid cases. The increase in h results in strong pressure-releasing porous linings and thus deeper vortex path bending and stronger sound radiation at a fixed R_f. The opposite is observed when η is increased. The strongest sound radiation is still found within the same R_f range, and the instants of strongest sound radiation remain unchanged. These results are expected and thus are not presented.

The increase in the length of the lined section to $L=2$ gives rise to more severe vortex path bending toward the lower porous surface, as shown in Fig. 5(a). The prolonged interaction between the vortex and the porous linings at increased L results in higher vortex velocities especially when the vortex is close to the end of the lined section [Figs. 5(b)

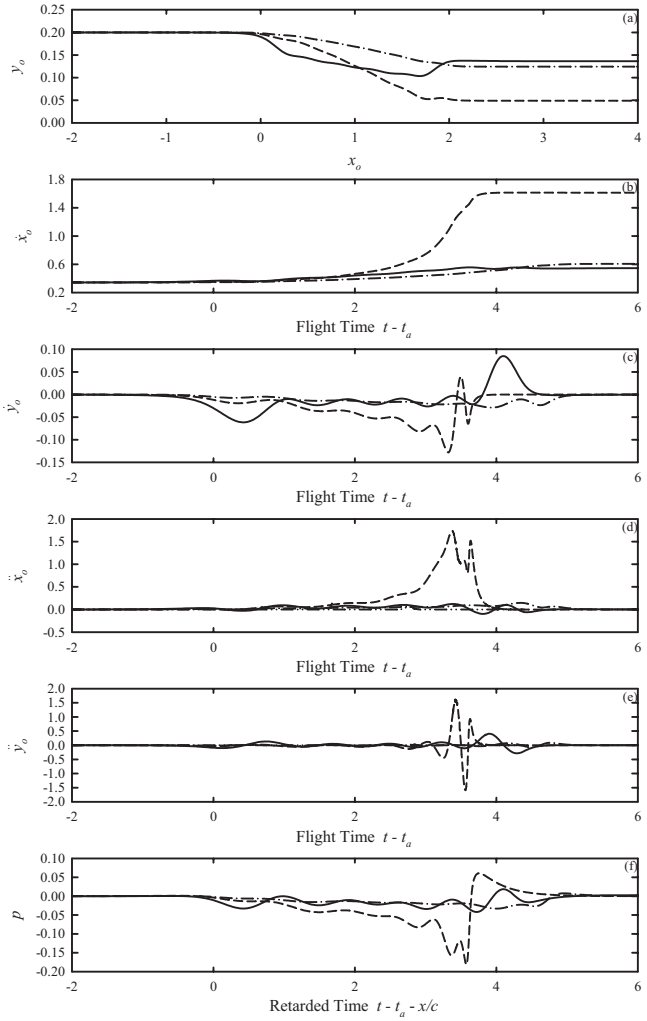


FIG. 5. Effects of R_f on vortex dynamics and sound generation for $L=2$, $y_{oi}=0.2$, and $\eta=3$. (a) Vortex path, (b) longitudinal velocity, (c) transverse velocity, (d) longitudinal acceleration, (e) transverse acceleration, and (f) far field sound pressure. ——: $R_f=0.5$, $h=0.2$; – – –: $R_f=10$, $h=0.2$; and –·–: $R_f=30$, $h=0.2$.

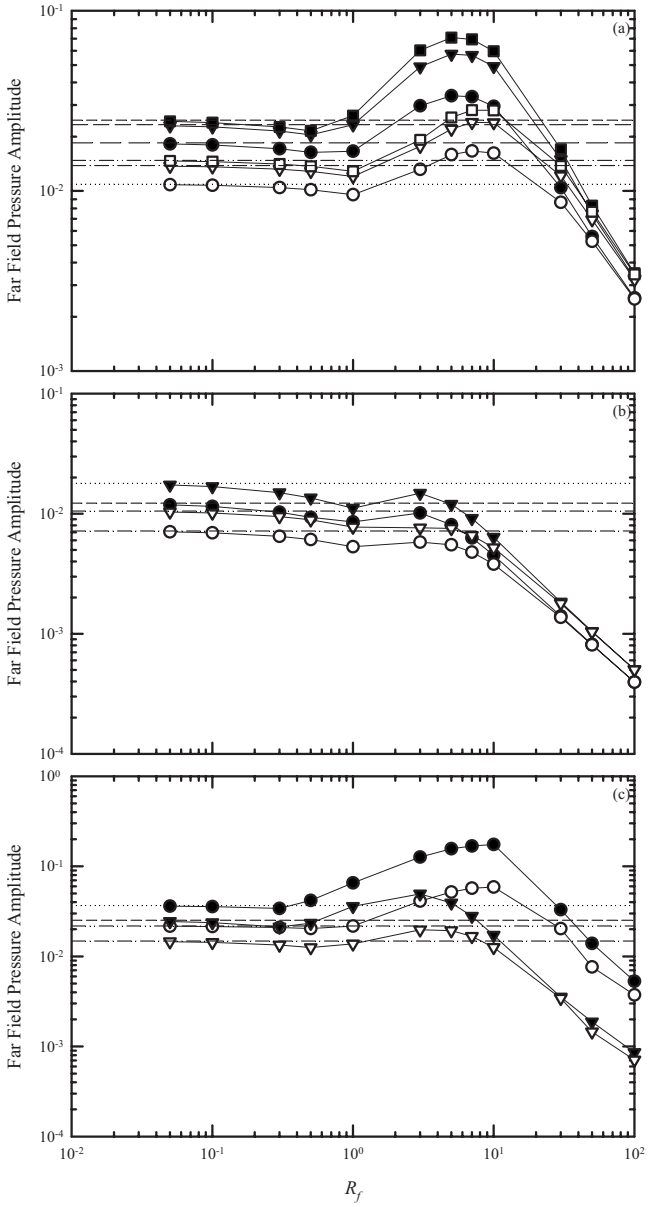


FIG. 6. Far field sound pressure amplitude under different combinations of system parameters. (a) $L=1$, $y_{oi}=0.2$, ●: $h=0.2$; ▼: $h=0.4$; ■: $h=0.8$; ----: $h=0.8$, $\eta=3$, $R_f=0$; – – –: $h=0.4$, $\eta=3$, $R_f=0$; – – –: $h=0.2$, $\eta=3$, $R_f=0$; ·····: $h=0.2$, $\eta=5$, $R_f=0$; –·–: $h=0.4$, $\eta=5$, $R_f=0$; and –·· –: $h=0.8$, $\eta=5$, $R_f=0$. (b) $L=1$, $y_{oi}=0.3$, ●: $h=0.2$; ▼: $h=0.4$; -----: $h=0.2$, $\eta=3$, $R_f=0$; ·····: $h=0.4$, $\eta=3$, $R_f=0$; –·–: $h=0.2$, $\eta=5$, $R_f=0$; and –··–: $h=0.4$, $\eta=5$, $R_f=0$. (c) $L=2$, ●: $y_{oi}=0.2$, $h=0.2$; ▼: $y_{oi}=0.3$, $h=0.2$. ----: $y_{oi}=0.3$, $h=0.2$, $\eta=3$, $R_f=0$; ·····: $y_{oi}=0.2$, $h=0.2$, $\eta=3$, $R_f=0$; –·–: $y_{oi}=0.3$, $h=0.2$, $\eta=5$, $R_f=0$; and –··–: $y_{oi}=0.2$, $h=0.2$, $\eta=5$, $R_f=0$. Closed symbols for $\eta=3$; open symbols for $\eta=5$.

and 5(c)]. Together with the higher vortex acceleration and deceleration near the instant at which the vortex leaves the lined section under the pressure-supporting effect of the rigid wall, stronger sound radiation is observed [Figs. 5(d)–5(f)]. The increase in the thickness of the porous linings results in stronger pressure-releasing effect, leading to higher vortex accelerations. Stronger sound radiation can then be expected from the results discussed previously in relation to Fig. 3. At $h=0.8$, the amplitude of the far field pressure is about seven times that at $h=0.2$ with the same initial vortex conditions for $R_f=10$ (not shown here). Other features observed in Fig. 5 are inline with those associated in Fig. 4 and thus are not discussed further.

Figure 6 summarizes the amplitude of the far field pressure generated under various combinations of the parameters studied. The horizontal lines represent the corresponding values for the perfectly inviscid case ($R_f=0$). The increase in h or a decrease in η will produce a louder sound because of the strong pressure-releasing effect of the porous lining which is expected.

For $y_{oi}=0.2$ and $L=1$, it can be observed that the slight increase in R_f from the vanishing value results in a very small drop of the pressure amplitude below that of the corresponding inviscid case [Fig. 6(a)]. For $1<R_f<25$, louder sound than that created in the inviscid case is observed for $\eta=3$. The maximum sound pressure appears at $R_f\sim 6$. The maximum sound pressure level increases by 5 dB above that in the inviscid case for $h=0.2$ and by as high as 9 dB for $h=0.8$. Further increase in R_f results in more pressure-supporting porous linings. This effect overcomes the pressure-releasing effect of η, resulting in the continuous drop of sound pressure as R_f increases behind 30.

The range of R_f for sound amplification is slightly reduced when η increases from 3 to 5, as shown in Fig. 6(a). The increases in sound pressure levels above the inviscid value are ~4 and ~6 dB for $h=0.2$ and 0.8, respectively. This trend suggests that no sound amplification will be achieved when η is increased further or when the lining becomes more and more pressure-supporting.

When the initial vortex height y_{oi} is increased, the effects of the linings on the vortex dynamics are less severe than those in the case when the vortex moves closer to the lower lining. The increase in R_f then results in weaker sound generation, as shown in Fig. 6(b). One should note that the lower lining in this case is less pressure-releasing as experienced by the vortex and thus the results in Fig. 6(b) are actually following the trend illustrated in Fig. 6(a).

The increase in L to 2 prolongs the active interaction period between the vortex and the linings. The magnitude of the far field sound pressure is increased, as indicated previously in Figs. 4 and 5. The range of R_f for sound amplification is large, as shown in Fig. 6(c). The increase in the far field sound pressure level is also impressive. For $y_{oi}=0.2$, $h=0.2$, and $\eta=3$, the maximum sound pressure level created is ~14 dB higher than both resulted in the inviscid situation and in the corresponding $L=1$ case. The increase in h to 0.8 results in an ~17 dB higher in the maximum sound pressure level than the corresponding $h=0.2$ case (not shown here). The increase in either y_{oi} or η reduces the far field sound amplitude as expected. In addition, one can observe from Fig. 6(c) that the value of R_f for maximum sound amplitude decreases as y_{oi} increases, but does not appear to depend much on η.

The variations in the radiated acoustical energy per unit spanwise length E with the present system parameters follow very closely those of the sound pressure amplitudes, as shown in Fig. 7. This is expected as a plane wave is produced at the far field so that

$$E = 2\int_{-\infty}^{\infty}\int_{0}^{1}(p^2/c)dydt = \frac{2}{c}\int_{-\infty}^{\infty}p^2dt, \tag{29}$$

and E is normalized by $\rho\Gamma^2$. However, the range of R_f for stronger acoustical energy radiation than the inviscid case, if there is, is wider than that for the sound pressure amplification shown in Fig. 6. This is because of the longer durations of active sound radiation than those in the inviscid cases. For $y_{oi}=0.2$, $h=0.8$, and $\eta=3$, the highest increase in the energy radiation is about 18 dB above that of the corresponding inviscid case, which is significantly higher than the 9 dB increase in the sound pressure level discussed before, and it occurs at $R_f \sim 5$ [Fig. 7(a)].

Figure 7(b) illustrates that lower energy radiation than that in the inviscid case is still observed for $y_{oi}=0.3$ when the lower porous lining effect is reduced. However, the percentage energy reduction is less than that for the sound pressure amplitude shown in Fig. 6(b). The longer lining length results in stronger acoustical energy radiation than that in the inviscid case even at small R_f [Fig. 7(c)]. Again, the percentage increase in the acoustical energy radiation is higher than that of the sound pressure amplitude.

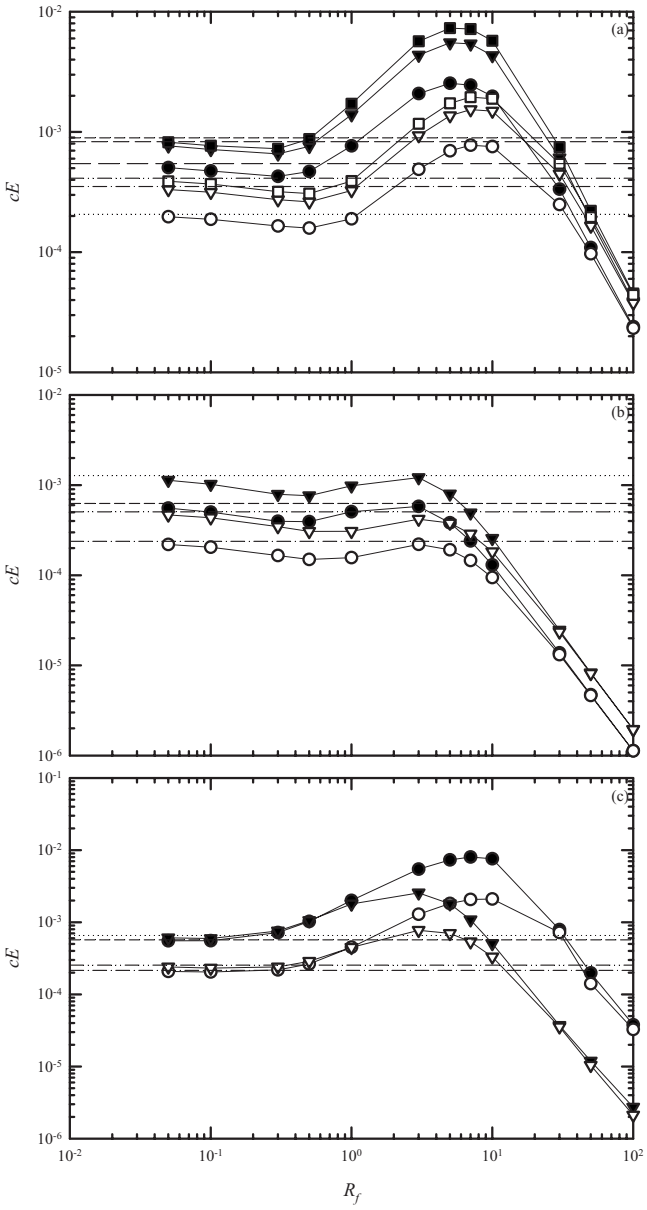


FIG. 7. Acoustical energy radiated per unit spanwise length under different combinations of system parameters. (a) $L=1$, $y_{oi}=0.2$; (b) $L=1$, $y_{oi}=0.3$; and (c) $L=2$. Legends: same as those for Fig. 6.

IV. CONCLUSIONS

The sound generated by the unsteady motion of a vortex moving across a lined duct section is investigated theoretically in the present study. The streamfunctions inside the flow field and inside the porous linings derived in terms of infinite series consist of modes with time varying magnitudes. The standard fourth order Runge–Kutta method was used to solve the coupled vortex dynamics equations numerically. The method of matched asymptotic expansion was applied to determine the time variation in the far field pressure so generated. The vortex was set at large distance upstream of the lined duct section initially and was located below the duct centerline to ensure its downstream propagation.

For a perfectly inviscid fluid, the pressure-releasing effect of the porous linings results in a downward transverse motion of the vortex. The longitudinal speed of the vortex

increases. The transverse vortex motion is the major sound generation mechanism when the lined section is not too long. As this length increases, the effects of the rates of change in the mode magnitudes, which affect directly the pressure on the lining surfaces and the vortex accelerations, become more significant in the sound radiation process. The increase in the effective fluid density inside the linings weakens the sound radiation.

The flow resistance inside the porous linings is pressure-supporting. It tends to increase the duration of active interaction between the vortex and the porous linings, resulting in stronger sound radiation after it has increased to an extent after which it overrides the pressure-releasing effect of the effective fluid density. However, its pressure-supporting property eventually attenuates the unsteady motion of the vortex and thus weakens the sound generation as it increases further. In the presence of a non-vanishing flow resistance, the rates of change in the mode magnitudes, which are directly related to the vortex accelerations, and the transverse vortex motions are of comparable importance in the radiation of sound. The prolonged interaction between the vortex and the linings in the presence of flow resistance results in greater/lower percentage increase/reduction in the acoustical energy radiated than that in the sound pressure amplification/attenuation.

ACKNOWLEDGMENT

This study was supported by a grant from the Research Grant Council, The Hong Kong Special Administration Region, People's Republic of China (Project No. PolyU5266/05E).

APPENDIX: G_1, G_2, H_1, H_2, AND THE RATES OF CHANGE OF MODE MAGNITUDES

The expressions for G_1, G_2, H_1, and H_2 can be obtained by solving Eqs. (8), (11), and (12). It is straight-forward to show that

$$G_1 = -\frac{e^{ikx_o}}{2|k|}\frac{\sinh(|k|(1-y_o))}{\sinh(|k|)} + \sum_{n=1}^{\infty}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}\frac{A_n e^{|k|}-B_n}{2\sinh(|k|)}\alpha_n e^{\alpha_n h}\sinh(\alpha_n h),$$

$$G_2 = \frac{e^{ikx_o}}{2|k|}\frac{\sinh(|k|(1-y_o))}{\sinh(|k|)} - \sum_{n=1}^{\infty}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}\frac{A_n e^{-|k|}-B_n}{2\sinh(|k|)}\alpha_n e^{\alpha_n h}\sinh(\alpha_n h),$$

$$H_1 = \frac{e^{ikx_o}}{2|k|}\frac{\sinh(|k|y_o)}{\sinh(|k|)}e^{|k|} + \sum_{n=1}^{\infty}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}\frac{A_n e^{|k|}-B_n}{2\sinh(|k|)}\alpha_n e^{\alpha_n(1+h)}\sinh(\alpha_n h),$$

$$H_2 = -\frac{e^{ikx_o}}{2|k|}\frac{\sinh(|k|y_o)}{\sinh(|k|)}e^{-|k|} - \sum_{n=1}^{\infty}\frac{(-1)^n e^{ikL}-1}{k^2-\alpha_n^2}\frac{A_n e^{-|k|}-B_n}{2\sinh(|k|)}\alpha_n e^{\alpha_n(1+h)}\sinh(\alpha_n h). \quad \text{(A1)}$$

It can also be shown from Eqs. (17) and (18) that

$$\int_0^L\int_{-\infty}^{\infty}|k|(\dot{G}_2-\dot{G}_1)e^{-ikx}dk\sin(\alpha_m x)dx = \pi(\eta\dot{A}_m + R_f A_m)\alpha_m e^{\alpha_m h}\cosh(\alpha_m h), \quad \text{(A2)}$$

and

$$\int_0^L\int_{-\infty}^{\infty}|k|(\dot{H}_2 e^{|k|}-\dot{H}_1 e^{-|k|})e^{-ikx}dk\sin(\alpha_m x)dx = -\pi(\eta\dot{B}_m + R_f B_m)\alpha_m e^{\alpha_m(1+h)}\cosh(\alpha_m h). \quad \text{(A3)}$$

With the expressions in Eq. (A1), Eqs. (A2) and (A3) can be re-written into the forms

$$-X_{A,m}\dot{x}_o + Y_{A,m}\dot{y}_o + \sum_{n=1}^{\infty}\beta_{mn}\dot{A}_n\alpha_n e^{\alpha_n h}\sinh(\alpha_n h) + \sum_{n=1}^{\infty}\gamma_{mn}\dot{B}_n\alpha_n e^{\alpha_n(1+h)}\sinh(\alpha_n h) = \frac{L}{2}(\eta\dot{A}_m + R_f A_m)\alpha_m e^{\alpha_m h}\cosh(\alpha_m h) \quad \text{(A4)}$$

and

$$-X_{B,m}\dot{x}_o + Y_{B,m}\dot{y}_o - \sum_{n=1}^{\infty}\gamma_{mn}\dot{A}_n\alpha_n e^{\alpha_n h}\sinh(\alpha_n h) + \sum_{n=1}^{\infty}\beta_{mn}\dot{B}_n\alpha_n e^{\alpha_n(1+h)}\sinh(\alpha_n h) = -\frac{L}{2}(\eta\dot{B}_m + R_f B_m)\alpha_m e^{\alpha_m(1+h)}\cosh(\alpha_m h), \quad \text{(A5)}$$

respectively, where

$$\beta_{mn} = -\frac{1}{\pi}\int_0^{\infty}\frac{k\alpha_m\coth(k)}{(k^2-\alpha_n^2)(k^2-\alpha_m^2)}\times\{(-1)^{m+n}-[(-1)^n+(-1)^m]\cos(kL)+1\}dk,$$

$$\gamma_{mn} = \frac{1}{\pi}\int_0^{\infty}\frac{k\alpha_m}{(k^2-\alpha_n^2)(k^2-\alpha_m^2)\sinh(k)}\times\{(-1)^{m+n}-[(-1)^n+(-1)^m]\cos(kL)+1\}dk,$$

$$X_{A,m} = -\frac{1}{\pi}\int_0^{\infty}\frac{k\alpha_m\sinh(k(1-y_o))}{(k^2-\alpha_m^2)\sinh(k)}\times[(-1)^m\sin(k(L-x_o))+\sin(kx_o)]dk,$$

$$Y_{A,m} = -\frac{1}{\pi}\int_0^\infty \frac{k\alpha_m \cosh(k(1-y_o))}{(k^2-\alpha_m^2)\sinh(k)} \times[(-1)^m \cos(k(L-x_o)) - \cos(kx_o)]dk,$$

$$X_{B,m} = \frac{1}{\pi}\int_0^\infty \frac{k\alpha_m \sinh(ky_o)}{(k^2-\alpha_m^2)\sinh(k)} \times[(-1)^m \sin(k(L-x_o)) + \sin(kx_o)]dk,$$

and

$$Y_{B,m} = -\frac{1}{\pi}\int_0^\infty \frac{k\alpha_m \cosh(ky_o)}{(k^2-\alpha_m^2)\sinh(k)} \times[(-1)^m \cos(k(L-x_o)) - \cos(kx_o)]dk.$$

It can be noted that $\beta_{mn}=\gamma_{mn}=0$ if $(m+n)$ is an odd integer. The rates of change in the mode magnitudes can then be obtained by simple matrix operation as in Ref. 18 once A_n, B_n, and the vortex velocity are known.

[1]H. G. Davies and J. E. Ffowcs Williams, "Aerodynamic sound generation in a pipe," J. Fluid Mech. **32**, 765–778 (1968).
[2]N. Curle, "The influence of solid boundaries upon aerodynamic sound," Proc. R. Soc. London, Ser. A **231**, 505–514 (1955).
[3]C. M. Harris, *Handbook of Noise Control* (McGraw-Hill, New York, 1979).
[4]A. Cummings and I.-J. Chang, "Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent," J. Sound Vib. **127**, 1–17 (1988).
[5]K. S. Peat and K. L. Rathi, "A finite element analysis of the convected acoustic wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545 (1995).
[6]A. Selamat, I. J. Lee, and N. T. Huff, "Acoustic attenuation of hybrid silencers," J. Sound Vib. **262**, 509–527 (2003).
[7]J. E. Ffowcs Williams, "The acoustics of turbulence near sound-absorbent liners," J. Fluid Mech. **51**, 737–749 (1972).
[8]M. C. Quinn and M. S. Howe, "On the production and absorption of sound by lossless liners in the presence of mean flow," J. Sound Vib. **97**, 1–9 (1984).
[9]L. M. Milne-Thomson, *Theoretical Hydrodynamics* (The University Press, Glasgow, UK, 1968).
[10]A. Powell, "Theory of vortex sound," J. Acoust. Soc. Am. **36**, 177–195 (1964).
[11]P. Cannell and J. E. Ffowcs Williams, "Radiation from line vortex filaments exhausting from a two-dimensional semi-infinite duct," J. Fluid Mech. **58**, 65–80 (1973).
[12]D. G. Crighton, "Radiation from vortex filament motion near a half plane," J. Fluid Mech. **51**, 357–362 (1972).
[13]F. Obermeier, "The influence of solid bodies on low Mach number vortex sound," J. Sound Vib. **72**, 39–49 (1980).
[14]M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, 2003).
[15]S. K. Tang, "Effects of porous boundaries on the dynamics of an inviscid vortex filament," Q. J. Mech. Appl. Math. **54**, 65–84 (2001).
[16]S. K. Tang and C. K. Lau, "Vortex sound in the presence of a wedge with inhomogeneous surface flow impedance," J. Sound Vib. **281**, 1077–1091 (2005).
[17]C. K. Lau and S. K. Tang, "Sound generated by vortices in the presence of a porous half-cylinder mounted on a rigid plane," J. Acoust. Soc. Am. **119**, 2084–2095 (2006).
[18]C. K. Lau and S. K. Tang, "Vortex sound under the influence of a piecewise porous material on an infinite rigid plane," J. Acoust. Soc. Am. **122**, 2542–2550 (2007).
[19]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[20]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[21]R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (McGraw-Hill, New York, 1990).
[22]M. Van Dyke, *Perturbation Methods in Fluid Mechanics* (Parabolic, Stanford, CA, 1975).
[23]J. E. Ffowcs Williams and D. J. Lovely, "Sound radiation into uniformly flowing fluid by compact surface vibration," J. Sound Vib. **71**, 689–700 (1975).
[24]S. K. Tang, R. C. K. Leung, and R. M. C. So, "Vortex sound due to a flexible boundary backed by a cavity in a low Mach number mean flow," J. Acoust. Soc. Am. **121**, 1345–1352 (2007).
[25]M. S. Howe, "A note on the interaction of unsteady flow with an acoustic liner," J. Sound Vib. **63**, 429–436 (1979).
[26]P. G. Saffman, *Vortex Dynamics* (Cambridge University Press, Cambridge, 1992).

Improved jet noise modeling using a new time-scale

M. Azarpeyvand[a] and R. H. Self
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton SO17 1BJ, United Kingdom

(Received 26 September 2008; revised 28 June 2009; accepted 5 July 2009)

To calculate the noise emanating from a turbulent flow using an acoustic analogy knowledge concerning the unsteady characteristics of the turbulence is required. Specifically, the form of the turbulent correlation tensor together with various time and length-scales are needed. However, if a Reynolds Averaged Navier–Stores calculation is used as the starting point then one can only obtain steady characteristics of the flow and it is necessary to model the unsteady behavior in some way. While there has been considerable attention given to the correct way to model the form of the correlation tensor less attention has been given to the underlying physics that dictate the proper choice of time-scale. In this paper the authors recognize that there are several time dependent processes occurring within a turbulent flow and propose a new way of obtaining the time-scale. Isothermal single-stream flow jets with Mach numbers 0.75 and 0.90 have been chosen for the present study. The Mani–Gliebe–Balsa–Khavaran method has been used for prediction of noise at different angles, and there is good agreement between the noise predictions and observations. Furthermore, the new time-scale has an inherent frequency dependency that arises naturally from the underlying physics, thus avoiding supplementary mathematical enhancements needed in previous modeling. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3192221]

PACS number(s): 43.28.Ra, 43.58.Ta [JWP] Pages: 1015–1025

I. INTRODUCTION

Jet noise is a major contribution to the overall noise produced by modern aero-engines, and it is important that reliable prediction schemes are developed as an aid to design engineers. To this end there has been much in interest recent years in the development of Reynolds Averaged Navier–Stokes (RANS) based jet noise prediction schemes. While the details vary, such schemes always seek to use the output of a RANS calculation (typically with a k-ε turbulence model) as input to a noise calculation. This method is also known as the two-step method. An obvious attraction of such a procedure is the relative speed of calculation compared to an LES or DNS calculation, but this is bought at the expense of a need for increased modeling of the noise production mechanism.

Perhaps one of the earliest, and certainly the best known, two-step RANS-based methods is the one developed by Mani–Gliebe–Balsa in the late 1970s.[1–3] These authors employed an empirically derived steady flow prediction to provide the input to an acoustic analogy model based on Lilley's analogy.[4] This method was improved by Khavaran *et al.*[5] and Khavaran[6] who replaced the empirical basis of the flow prediction with a CFD calculation based on a k-ε turbulence model. This technique is generally referred to as the Mani–Gliebe–Balsa–Khavaran (MGBK) method. One of the first applications of the MGBK method was made by Khavaran and Georgiadis[7] for supersonic elliptic jets. Hamed *et al.*[8] used the method for high bypass coplanar coaxial jets and Barber *et al.* applied the method to axisymmetric multistream[9] and high speed jet flows[10] while the mean and turbulence parameters were obtained using the WIND flow solver. Frendi *et al.*[11,12], applied the MGBK methodology to a supersonic jet flow, while a multiple time-scale approach was used and results were in acceptable agreement with experimental data. More recently, researchers at the NASA Glenn Research Center have been working on further improvement of the MGBK method by including the turbulence anisotropy effect,[6] source compactness effect,[13] alternative source term descriptions,[13,14] and also the refraction effect.[15–17]

In addition to the MGBK method, a number of other RANS-based jet noise prediction methodologies have also been developed and used. Self[18] and Self and Bassetti[19] used a fourth-order space-time velocity correlation model in the definitions of the source terms. A similar methodology was also applied to isothermal coaxial jets by Page *et al.*[20] Another RANS-based model has been developed by Tam and Auriault.[21] According to this method the radiated noise originates from two types of sources, those radiated by fine-scale eddies and those produced by large-scale ones. The application of this method to other geometries and working condition has also been tested.[22–24] More recently, a hybrid prediction methodology for jet noise has been developed by Karabasov *et al.*[25] In this method the sound generation is modeled following Goldstein's acoustic analogy,[26] the far field propagation is modeled via solution of a system of adjoint linear Euler equations, and the constants of proportionality are extracted via comparison with turbulence length- and time-scales obtained from an LES prediction, rather than being determined empirically.

[a]Author to whom correspondence should be addressed. Electronic mail: ma@isvr.soton.ac.uk

Morris and Farassat pointed out that a k-ε model provides only time averaged properties of the flow, whereas a more detailed knowledge of the statistical nature of the turbulence is required for the noise calculation; in particular, the two-point correlation of the Reynolds stresses is needed.[27] Although the model used for the source terms is of great importance, a proper definition of the defining parameters in the source term model, e.g., time-scale, length-scale, and convection velocity, can be even more important. It has been known for some time that the frequency dependency of these parameters is of significant importance.[28] Morris and Boluriaan[29] made use of the frequency dependent length-scale, originally suggested and measured by Harper-Bourne[30] for a low speed (M_J=0.18) jet flow. Self[18] recently implemented this frequency dependence in his statistical noise prediction methodologies and showed that noise prediction improvement is possible if frequency dependent parameters are used. In this paper the authors shall first provide a mathematical model for jet noise prediction based on the MGBK method. We will then consider frequency dependency of different parameters used in the source term definition, such as time-scale and length-scale, and will propose and validate a new time-scale based on the turbulent energy transfer rate.

The layout of this paper is as follows. In Sec. II we shall provide some preliminary formulation for the governing equations. This section discusses the derivation of the MGBK method and how the refraction and source compactness effects are taken into account. Gaussian functions are used for the modeling of the spatial and temporal correlation terms but before they can be used for noise prediction the various parameters indicative of the turbulence must be decided upon, in particular, the correlation length-scale and time-scale are required. This is not a simple matter and Sec. III gives a review of how previous authors have tackled the problem and gives a discussion of the underlying physics involved. By considering time-scales associated with the different physical processes occurring in the turbulent cascade, a new expression for the time-scale is proposed which is based on energy transfer rate. It should be mentioned here that in this paper, we only examine the effect of the choice of time- and length-scales, and it does not evaluate the effects of other assumptions and choices in the model. Section IV considers numerical results and comparisons. Predictions of the far-field noise are made using both old and new time-scales, and the results are presented and compared, as are the axial source distributions.

II. BACKGROUND THEORY

In this section we review the theory required for the noise prediction model and show the expressions that will be used for the far-field acoustic intensity spectrum. The MGBK method has been used for the prediction of the radiated noise from the jet. The starting point of the MGBK method is Lilley's equation,[1,4] where its operator and the source term for an inviscid flow linearized about a unidirectional transversely shear mean flow in a cylindrical coordinates, (x,r,θ), can be expressed as

$$L(p,x) = c^{-2}D^3p - D\Delta p - \frac{d}{dr}(\log c^2)D\frac{\partial p}{\partial r} + 2\frac{dU}{dr}\frac{\partial^2 p}{\partial x \partial r}$$
$$= \rho D\nabla\cdot\nabla\cdot(v_1v_1 - \overline{v_1v_1}) - 2\rho\frac{dU}{dr}\frac{\partial}{\partial x}\nabla\cdot(v_1v_2 - \overline{v_1v_2}), \quad (1)$$

where t denotes time, p is acoustic pressure, c is the sound speed, ρ is the density, v_1 and v_2 are the axial and radial turbulent velocity fluctuations, respectively ($v_2^2 \approx v_3^2$ for an axisymmetric jet flow), and U is the mean velocity. The overbar notation shows an ensemble average, D is the convective derivative, $D=\partial/\partial t + U(\partial/\partial x)$, and

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}. \quad (2)$$

It has been shown that the total radiated noise can be expressed as the summation of the self-noise (turbulence-turbulence interaction) and shear-noise (mean flow and turbulence interaction).[31] The derivation of the noise source terms, their directivity, and also refraction effect can be found in Ref. 1. The final expressions of the power spectral directivity of an axisymmetric jet for an observer located at (R,θ) can be expressed as

$$\overline{p^2_{\text{self}}}(R,\theta,\omega) = \int_y\int_{-\infty}^{+\infty}\{D_{11}^{(M)} + 2\zeta_1 D_{22}^{(M)} + 4(\zeta_2+2\zeta_4)D_{12}^{(M)} + 2(\zeta_3+2\zeta_5)D_{23}^{(M)}\}\Pi_M e^{j\Omega\tau}d\tau d\mathbf{y}, \quad (3)$$

$$\overline{p^2_{\text{shear}}}(R,\theta,\omega) = \int_y\int_{-\infty}^{+\infty}\{\zeta_4 D_{11}^{(M)} + (\zeta_1+\zeta_5)D_{12}^{(M)}\}\times\Pi_D e^{j\Omega\tau}d\tau d\mathbf{y}, \quad (4)$$

where

$$\Pi_M = \frac{1}{(4\pi Rc^2)^2(1-M_s\cos\theta)^2(1-M_c\cos\theta)^2}I_{1111}(\mathbf{y},\tau), \quad (5)$$

$$\Pi_D = \left(\frac{2dU/dr}{\Omega}\frac{1-M_c\cos\theta}{1-M_s\cos\theta}\right)^2\Pi_M. \quad (6)$$

In the above equations M_c and M_s are the convective and local Mach numbers, and the proportionality factor, I_{1111}, in Eq. (5) is proportional to

$$I_{1111}(\mathbf{y},\tau) \propto (1-M_c\cos\theta)^{-1}(1-M_s\cos\theta)^{-2}\times\int\frac{\partial^4}{\partial\tau^4}\overline{v_iv_jv_k'v_l'}d^3\mathbf{r}, \quad (7)$$

where the term $\overline{v_iv_jv_k'v_l'}$ is a fourth-order correlation and its modeling will be dealt with later, and Ω is the observer frequency which relates to the source frequency through[11]

$$\Omega = \omega\sqrt{(1-M_c\cos\theta)^2 + (\alpha_c k^{0.5}/c_0)^2}, \quad (8)$$

where the term $\alpha_c k^{0.5}/c$ accounts for the finite lifetime of the eddy as it is convected downstream. The constant α_c is found

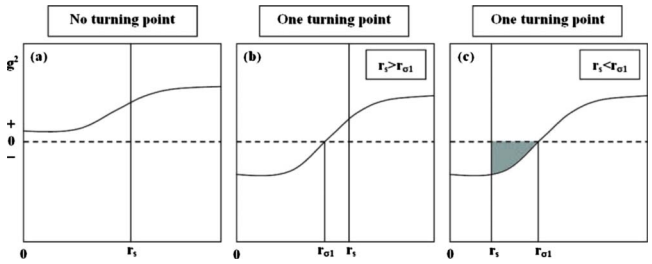


FIG. 1. (Color online) Position of the turning points. Shaded areas denote shielding of source.

from the measured data to be approximately 0.5. The directivity factors can be obtained from

$$D_{11}^{(M)} = \frac{\cos^4\theta}{(1 - M_c\cos\theta)^4}\beta_{xx},$$

$$D_{12}^{(M)} = \frac{g_s^2\cos^2\theta}{2(1 - M_c\cos\theta)^2}\beta_{xy},$$

$$D_{22}^{(M)} = \tfrac{3}{8}g_s^4\beta_{yy},$$

$$D_{23}^{(M)} = \tfrac{1}{8}g_s^4\beta_{yz},$$

where g_s is the value of the shielding function, $g(r)$, given below, at the source location. The shielding function is given by

$$g^2(r) = \frac{(1 - M_s\cos\theta)^2 - \cos^2\theta}{(1 - M_c\cos\theta)^2}. \tag{9}$$

The shielding coefficients β_{xx}, β_{xy}, β_{yy}, and β_{yz} depend on the case encountered in Fig. 1 and Table I. The parameters β_{01} is defined as

$$\beta_{01} = \exp\left\{-2\Omega/c\int_{r_s}^{r_\sigma}|g^2(r)|^{1/2}dr\right\}, \tag{10}$$

where r_s is the radial distance of the source from the jet axis and r_σ is the turning point, i.e., location in the shear layer where g^2 changes sign. When a negative region exists, fluid shielding of the source is possible and the amount of shielding depends on the proximity of the source with respect to the turning point as well as the number of turning points; see Fig. 1. For a single flow cold jet, it is reasonable to assume that only one turning point occurs.

The weight coefficients, ζ_i (i=1 to 5), appearing in Eqs. (3) and (4) are related to the non-isotropic factors $\beta_c = 1 - \overline{v_2^2}/\overline{v_1^2}$ and $\Upsilon = L_2/L_1$ (L_1 and L_2 are length-scales in the axial, radial, and azimuthal directions, respectively) through

TABLE I. Shielding coefficients β_{ij}.

Case	β_{xx}	β_{xy}	β_{yy}	β_{yz}
a	1	1	1	1
b	1	1	1	1
c	β_{01}	0	0	0

$$\zeta_1 = \frac{3}{2}\beta_c^2 + \frac{1}{32}[9(\Upsilon + \Upsilon^{-1})^4 - 48(\Upsilon + \Upsilon^{-1})^2 + 80] - \frac{\beta_c}{4}(6 - \Upsilon^2 + 3\Upsilon^{-2}),$$

$$\zeta_2 = \tfrac{1}{8},$$

$$\zeta_3 = \tfrac{1}{8}\left[\tfrac{3}{4}(\Upsilon + \Upsilon^{-1})^4 - 4(\Upsilon + \Upsilon^{-1})^2 + 7 - 2\Upsilon^2 + 4\beta_c^2\right] + 2\beta_c(\Upsilon^2 - 2 - \Upsilon^{-2}),$$

$$\zeta_4 = \tfrac{1}{16}(5 + 2\Upsilon^{-2} - 8\beta_c),$$

$$\zeta_5 = \tfrac{1}{2}(\zeta_1 - \zeta_3). \tag{11}$$

In the limiting case of isotropic turbulence the anisotropy factors reduce to Υ=1 and β_c=0.

Regarding the fourth-order correlation function in Eq. (7), it has been shown that for a homogeneous and isotropic turbulent flow, it consists of a sum of second-order velocity correlations[32]

$$\overline{v_i v_j v_k' v_l'} = \overline{v_i v_j}\,\overline{v_k' v_l'} + \overline{v_i v_k' v_j v_l'} + \overline{v_i v_l' v_j v_k'}. \tag{12}$$

This formulation was first used and experimentally examined by Uberoi.[33] One of the most convenient forms of correlation was suggested by Ribner[34,35] as

$$\overline{v_i v_j'}(r,\tau) = \Psi_{ij}(r)g(\tau), \tag{13}$$

where for homogeneous isotropic turbulence the spatial part can be written as[32]

$$\Psi_{ij}(r) = \overline{v_1^2}\,[(f(r) + 1/2rf'(r))\delta_{ij} - 1/2f'(r)r_i r_j/r], \tag{14}$$

where f is function of r which can take different forms, and the prime denotes derivation with respect to r. A Gaussian form has been chosen for f function as

$$f(r) = e^{-\pi(r_1^2/L_1^2 + (r_2^2 + r_3^2)/L_2^2)}. \tag{15}$$

The temporal part of the correlation function can also be modeled by

$$g(\tau) = e^{-(\pi/\tau_0)^2}, \tag{16}$$

where τ_0 and L are characteristic time- and length-scales of the turbulence, respectively.

It should be noted here that the Gaussian form of the spatial and temporal correlation functions is not always supported by experiments. But, it is a convenient mathematical form. A review of the available forms of these correlation functions can be found in Ref. 30. Note, however, that the choice of correlation function may have an effect on the eventual noise prediction. This should be borne in mind when considering later comparisons. In this paper we limit ourselves to showing that the choice of an energy transfer derived time-scale improves noise prediction given an assumption of Gaussian statistic. However, in the authors opinion such improvements are likely to be reflected for other correlation functions.

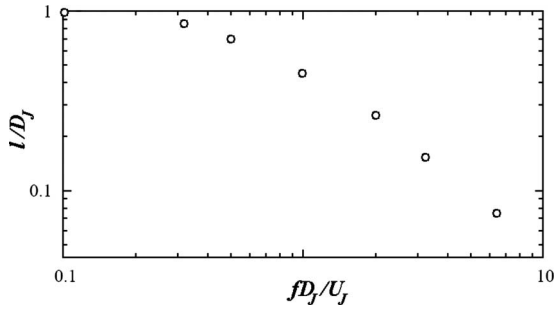


FIG. 2. Frequency dependency of the length-scale as measured by Harper-Bourne (Ref. 30).

III. MODELING THE LENGTH- AND TIME-SCALES

While the functional form chosen to model the turbulent correlations is clearly important, another significant factor is choosing an appropriate relationship between the scales that appear in Eqs. (15) and (16) and their RANS derived counterparts. Sections II A and II B provide a comprehensive review of the possible choices of the definition of the length-scale and time-scale.

A. Length-scale

This section concerns possible definitions of the length-scale. The length-scale that characterizes the energy-containing scales is given by

$$l_d = c_l \frac{\sqrt{k^3}}{\varepsilon}, \tag{17}$$

where c_l is a calibrating constant. Here k is the turbulent kinetic energy and ε is a measure of the dissipation of turbulent energy. This length-scale is the most used model and has been routinely used in the literature.

One of the most important characteristics of the length-scale is its observed frequency dependency. The frequency dependency of the length-scale was first demonstrated by Harper-Bourne.[30] His results indicate that for low frequencies the assumption of a constant length-scale is reasonable, but that for higher frequencies a nearly inverse dependence on Strouhal number is obtained; see Fig. 2. Later, Self[18] proposed a frequency dependent model for the length-scale based on experimental results presented in Ref. 30. He showed that Harper-Bourne's experimental observations can be fitted into an analytic formula of the form

$$l(\omega) = c_1 W/(1 + \omega/\omega_c), \tag{18}$$

where W is the shear layer width and

$$\omega_c = 2\pi c_2 U_1/W, \tag{19}$$

where c_1 and c_2 are calibrating coefficients and U_1 is the center-line jet velocity. An exponential fit to Harper-Bourne's results was given later by Morris and Boluriaan.[29] The longitudinal length-scale was modeled using

$$l(\omega) = cD_J \frac{1 - e^{-c_s \,\mathrm{St}\, l_d/D_J}}{\mathrm{St}}, \tag{20}$$

where l_d is the turbulence length-scale, Eq. (17) (with $c_l=1$), and Strouhal number is defined as $\mathrm{St}=fD_J/U_J$. The numerator of the ratio is chosen in such a way that l is constant at low frequencies but then decreases with increasing frequency. According to the authors' interpretation, c_s is a factor which determines the transition between the low and high frequency behaviors of the spectrum. In fact, this factor can control both the location of the peak frequency (i.e., adjusting the local Strouhal number, i.e., $\omega l_x/U$) and also to some extent the shape of the spectrum. It was also shown that the low frequency part of the spectrum is much more sensitive to c_s than the high frequency part.

B. Time-scale

In defining a time-scale to characterize the acoustic sources, the turbulent dissipation rate is a common choice and an equation analogous to Eq. (17) is used:

$$\tau_d = c_\tau \frac{k}{\varepsilon}. \tag{21}$$

It has been pointed out elsewhere that this definition of time-scale rarely results in good agreement between predicted and measured acoustic spectra over the entire frequency range of interest with under-prediction at both the high and the low ends of the spectrum (see, for example, Ref. 18). A model to take account of the frequency dependency was suggested in Ref. 18 and used to capture the 90° spectrum of a single-stream jet using the Lighthill acoustic analogy. Subsequently the model was applied to a co-axial jet.[20] According to this model, the time-scale varies with frequency as

$$\tau_d(\omega) = c_\tau \frac{2\pi}{\omega_c + \omega}, \tag{22}$$

where c_τ is a calibrating coefficient that must be obtained empirically, and the critical radian frequency ω_c is defined by

$$\omega_c = c_\omega \frac{2\pi U_c}{l_d}, \tag{23}$$

with U_c being the local mean velocity and $l_d=k^{3/2}/\varepsilon$ is the turbulent length-scale (with $c_l=1$). In simple physical terms, Eq. (21) is equivalent to an assumption that the lifetime of a turbulent eddy is proportional to the local mean shear, while Eqs. (22) and (23) essentially allow for an enhanced decay rate of those eddies whose characteristic size is comparable to that of the local shear layer width. However, this is a largely qualitative explanation that offers only a partial insight into the underlying physics that determines the exact nature of the dependence of time-scale on frequency. While superior to the simple model of Eq. (21), the model described by Eqs. (22) and (23) shares the assumption that the time-scale depends on the dissipation of turbulent eddies, i.e., that it should be derived solely from a turbulent time-scale that is based on the dissipation rate of the turbulent kinetic energy. However, it is well known in the literature on turbulent flows that several different processes operate si-

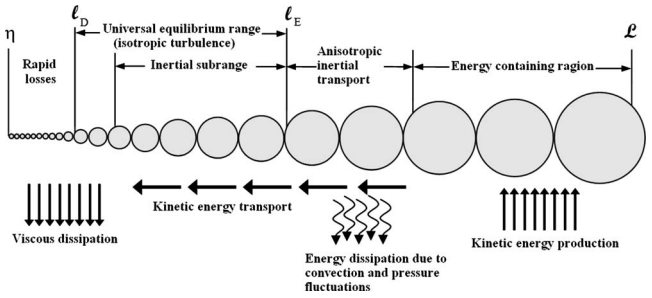


FIG. 3. A schematic diagram of the energy cascade and eddy sizes, showing the various length-scales and the ranges.

multaneously and that a characteristic time-scale can be associated with each of them. There appears to be no *a priori* reason to favor one of these time-scales over the others, but rather it would seem sensible to assume that the time-scale depends foremost on the time-scale associated with the physical process that dominates the behavior of the turbulent flow in any particular region of the jet. Thus we are led to consider time-scales associated with two physical processes:

- production of turbulent energy and
- dissipation of turbulent energy.

The time-scale associated with the characteristic turbulent time-scale for each of these processes will be denoted by τ_p and τ_d, respectively. Assuming a simple proportional dependence of the time-scale on the corresponding turbulent time-scale,[11,12] these will be given by

$$\tau_p = \alpha_p \frac{k}{\mathrm{Pr}}, \quad \tau_d = \alpha_d \frac{k}{\varepsilon}, \tag{24}$$

where Pr denotes production of turbulent kinetic energy (see, for example, Ref. 12 for a definition of Pr), and as before, k/ε gives a measure of the dissipation rate, and α_p and α_d are two empirical coefficients. This set of time-scales will be referred to as the combined time-scale (CTS) henceforward.

Each of the two different processes (and hence the corresponding time-scales) can be associated with a region of the turbulent jet where they are the dominant physical process.[11,12] This is illustrated schematically in Fig. 3 (although it should be borne in mind that in reality both two physical processes are present everywhere in the jet flow). Consequently, we may expect that predictions using the theory developed in Sec. II will give, for each of the time-scales defined in Eq. (24), good agreement for those frequencies emanating from the portion of the jet with which the time-scale is associated. We shall show later in Sec. IV that this is indeed the case, but that none of the time-scales alone give good agreement throughout the entire frequency range.

This suggests that in order to accurately predict the noise over the entire frequency range some appropriate combination of the time-scales defined in Eq. (24) should be used. A model of this type was first proposed by Frendi *et al.*[11,12] in their "dual time-scale" model. Thus,

$$I_T = \sum_j w_j I_j, \quad j = \{p,d\}, \tag{25}$$

where I_j refers to sound intensity calculated using each of the time-scales in Eq. (24) and w_j are weight parameters that must be determined empirically. An optimization has to be performed to determine the two calibration coefficients α_p and α_d. While such a procedure might suggest the essential correctness of this approach we now have the difficulty of calibrating two different time-scale parameters and corresponding length-scale parameters, and the number of empirical coefficients is raised still further if the weightings in Eq. (25) are included. However, one can readily realize that specifying regions for each time-scale is a very difficult task and for this reason we follow the method used by Frendi *et al.*[11,12] in comparison below. According to this method the far-field mean-square-pressure is determined as the average of the mean-square-pressure given by the individual time-scales (i.e., $w_p = w_d = 1/2$).

Another way of resolving the problem is by introducing a time-scale that is based on the transfer of energy between different wave numbers of the turbulent fluctuations and which naturally reduces to each of the time-scales in Eq. (24) in the differing regimes of the jet. The energy transfer rate can be estimated via $\varepsilon/E(\kappa)$, which shows the rate at which the energy travels through the cascade $(d\kappa/dt)$, and $E(\kappa)$ is the energy spectrum of the turbulence. So, the time scale can be found through

$$\tau_T \propto \int \frac{E(\kappa)}{\varepsilon} d\kappa. \tag{26}$$

In order to proceed with the integration, Pao's energy spectrum model[36] is used:

$$E(\kappa) \propto \varepsilon^{2/3} \kappa^{-5/3} e^{-3/2C(\kappa\eta)^{4/3}}, \tag{27}$$

where η in this equation is the Kolmogorov length-scale which can be computed from $\eta = (\nu^3/\varepsilon)^{3/2}$, with ν being the kinematic viscosity, and C is set to 1.5. Pao's energy spectrum is only valid over the inertial spectrum and the dissipation region, which contain most of the acoustic sources.[37] The corresponding time-scale is now written as

$$\tau_T = \alpha_T \tau_d \cdot \left(\frac{\Lambda}{l_d}\right)^{2/3} \cdot e^{3/2C(2\pi(\eta/l_d))^{4/3}} - \alpha^T \sqrt{\pi C}\, \tau_\eta \cdot \mathrm{erf}\, i\left(\frac{3}{2} C \left(2\pi \frac{\eta}{l_d}\right)^{2/3}\right), \tag{28}$$

where α_T is now the sole calibrating parameter, Λ denotes the size of the eddy which can be either found from the experimental results or can be estimated from the shear layer thickness, l_d is again the length-scale, given by Eq. (17), $\tau_\eta = \sqrt{\nu/\varepsilon}$ is the Kolmogorov time-scale, and erf i the imaginary error function defined by

$$\mathrm{erf}\, i(z) = \frac{2}{i\sqrt{\pi}} \int_0^{iz} e^{-t^2} dt. \tag{29}$$

It should be noted here that the exponential factors in the first and second terms of Eq. (28), which depend on the Kolmog-

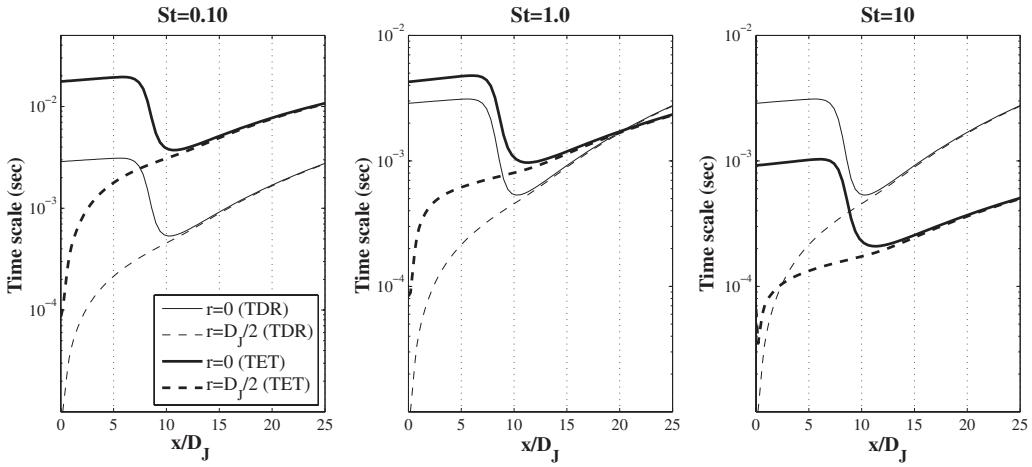


FIG. 4. Comparison of the TET and TDR time-scales at different radial locations at three Strouhal numbers (St==0.1,1,10).

orov scales, can be neglected in comparison with the other terms and allows simplification of Eq. (28) to the following form:

$$\tau_T \approx \alpha_T \tau_d \left(\frac{\Lambda}{l_d}\right)^{2/3}. \tag{30}$$

A comparison of the turbulent energy transfer time-scale (30) (will be called TET time-scale hereafter) and the traditional turbulent dissipation rate (24) (referred to as TDR in the following) is presented in Fig. 4. The eddy size Λ can be estimated using either the shear layer thickness, Eq. (18), or the frequency dependent length-scale, Eq. (20) (these two are, however, equivalent). In this comparison the latter model is used. Comparisons are provided for three Strouhal numbers, St=0.1, 1, and 10, at two radial distances ($r=0$ and $r=D_J/2$). It can be seen that the time-scale associated with $r=0$ has a "step jump" around $x=6D_J$ which is because it crosses the surface of the potential core. The most informative curves are the ones associated with $r=D_J/2$ (across the nozzle lip-line). It can be seen from the figures that the TET time-scale provides smaller values than the TDR time-scale at high frequencies (St=10). Conversely, at lower frequencies the TDR time-scale is smaller than the TDR time-scale. In other words, the TET time-scale provides smaller values for small eddies and greater values for large eddies. The two curves cross when $\Lambda=l$. This position moves toward the jet exit as the frequency increases. However, as mentioned earlier one can also use a shear layer thickness model in order to estimate the eddy size Λ. In this case the TET time-scale depends on the axial position of source instead of the source frequency.

The effect of the Kolmogorov factors in Eq. (28) and comparisons with its simplified version, Eq. (30), are presented in Fig. 5. For this case a source with a turbulent dissipation rate $\varepsilon=2.0\times10^8$ m^2 s^{-3} is assumed. The new time-scale clearly manifests a faster decay as we approach the smaller scales. It is seen that the Kolmogorov scale effect becomes dominant for very small eddies, at which the original new time-scale, Eq. (28), falls exponentially and much faster than the simplified version without the viscous dissipation factor. However, it is known that the eddies very close to the Kolmogorov size are not significant contributors to the noise production and radiation mechanism. So, it is reasonable to use the simple form of Eq. (30) in our noise prediction codes.

IV. NUMERICAL RESULTS AND DISCUSSIONS

A single-stream isothermal jet working at $M_J=0.75$ and 0.90 has been considered in this study, and a RANS scheme using a k-ε turbulence model was used to achieve the required input for the acoustic source model. Such a solution provides an estimate of the amplitude of the turbulent velocity fluctuations as well as a local turbulent length-scale. The following coefficients are chosen in our k-ε model:

$$C_{1\varepsilon}=1.44, \quad C_{2\varepsilon}=1.83, \quad C_\mu=0.09. \tag{31}$$

In this model $C_{2\varepsilon}$ is changed from the default value of 1.92–1.83 to reduce the spreading rate from 0.12 to 0.10 and get a better self-similarity agreement in the fully developed range.

Various ways of defining a time-scale for jet noise prediction based on information obtained via a RANS calculation have been considered in Sec. III. In Secs. IV A–IV C we shall make use of these time-scales to predict noise radiation

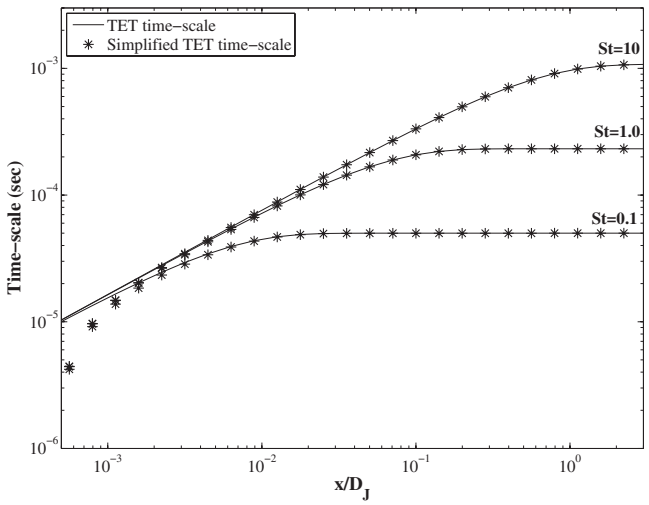


FIG. 5. Effects of the Kolmogorov scales on the TET time-scale.

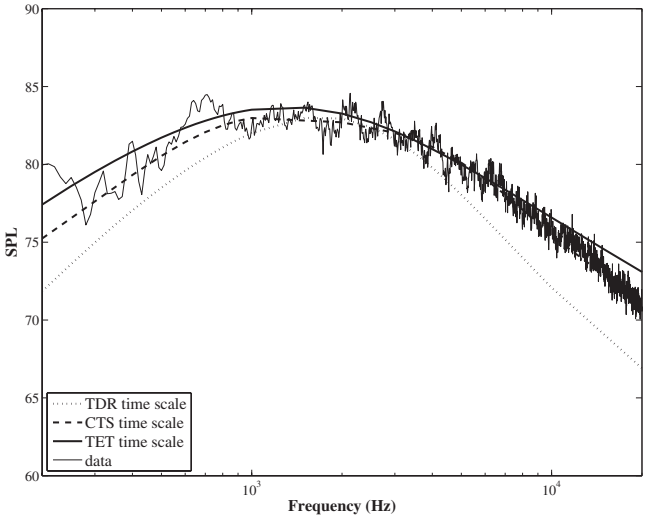


FIG. 6. Comparison of narrowband data with predicted spectral density at 90° to the jet axis using three different time-scales and based on the MGBK method: $R=50D_J$, $M_J=0.75$, and $D_J=0.05$ m.

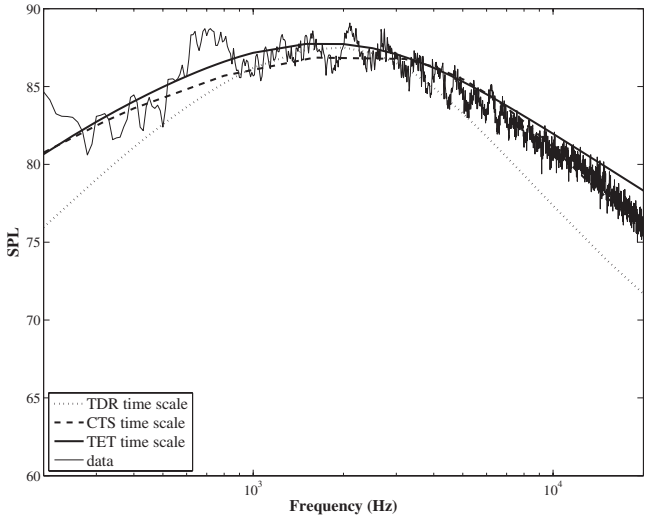


FIG. 7. Comparison of narrowband data with predicted spectral density at 90°to the jet axis using three different time-scales and based on the MGBK method: $R=50D_J$, $M_J=0.90$, and $D_J=0.05$ m.

to an observer located at 90°, to examine the directivity effect, and finally to assess the jet source distribution.

A. Noise prediction comparisons at 90°

Calculations of the far-field noise for an $M_J=0.75$ jet (at $R=50D_J$, and 90° to the jet axis) based on the model given in Sec. II are presented in Fig. 6. Three sets of calibrating coefficients have been found using three types of time-scales (TDR, CTS, and TET time-scales). In each case the coefficients have been varied to obtain the best fit of predicted noise with experimental data at 90°. The following coefficients have been found:

$$c_\tau = 0.40, \quad c_1 = 0.69 \quad \text{(TDR)},$$

$$\left.\begin{aligned} w_p &= 1/2, \quad \alpha_p = 1.50, \quad c_l = 1.71, \\ w_d &= 1/2, \quad \alpha_d = 1.35, \quad c_l = 1.71, \end{aligned}\right\} \quad \text{(CTS)},$$

$$\alpha_T = 0.26, \quad c_l = 0.49, \quad c_s/2\pi = 3.58 \quad \text{(TET)},$$

w_j are weighting coefficients defined in Eq. (25). The anisotropy parameters Υ and β_c are chosen as 0.5 and 0.4, respectively, following Ref. 38.

Figure 6 compares predictions made using the simple dissipation time-scale (21) with those of the two physical time-scales based on turbulent energy production rate and turbulent energy dissipation rate, Eq. (24), and finally the new time-scale presented in Eq. (28). These results are compared with narrowband data obtained within the JEAN research program.[39] Assuming that the CTSs, as given in Eq. (24), are the correct ones to use in the different regimes of the jet where each physical process dominates, it can be seen why the time-scale given by Eq. (21) leads to agreement with measured spectra around the vicinity of the peak noise. This noise is generally associated with the region of the jet just downstream of the end of the potential core where the dissipation of turbulent energy is the dominant process. Moving away from this region leads to progressively poorer agreement as other physical processes begin to dominate. In these regions an improved prediction of the noise can be obtained by using Eqs. (22) and (23); see Ref. 18. This is because such a model effectively mimics the strain rate time-scale and (to some extent) the production time-scale which are dominant in regions of the jet away from the end of the potential core. Additionally, the composite time-scale (TET) has been proposed based on the energy transfer rate in the cascade and which matches the two separate time-scales in the different jet regimes. The new time-scale leads to good agreement with experimental data, as is evident in Fig. 6. Significantly, in this CTS approach the relationship of acoustic and turbulent time-scales is one of simple proportionality. The effective dependence on frequency now arises naturally as the relative importance of different physical processes changes with different regions of the jet flow. Simulation has also been performed for a single-flow isothermal $M_J=0.90$ jet. Comparison of the results obtained using the TDR, CTS, and TET time-scales is summarized in Fig. 7. A similar behavior and trend to the $M_J=0.75$ jet can be observed from this high speed subsonic jet flow. Use of the TDR time-scale can result in up to 5 dB discrepancy at 200 Hz and 10 dB difference at 10 kHz, while the results obtained using the TET time-scale follow the experimental data over the entire frequency range of interest (200 Hz–20 kHz).

B. Directivity effects

The prediction of the noise at other angles is a more challenging problem than that of the 90° as it involves more physical effects. Equations (4) and (5) along with the TET time-scale, Eq. (30), are used for prediction of the far-field noise at various angles to the jet axis. Results are presented in Fig. 8. It can be seen that the results show an acceptable agreement over all observer angles between 50° and 130° to the downstream axis. Although there is some discrepancy of the noise at high frequencies for small and large angles these results are still better than those obtained using the time-scale defined by Eq. (21) or even Eqs. (22) and (23); see Refs. 19 and 20. Some mismatch between peak frequency of

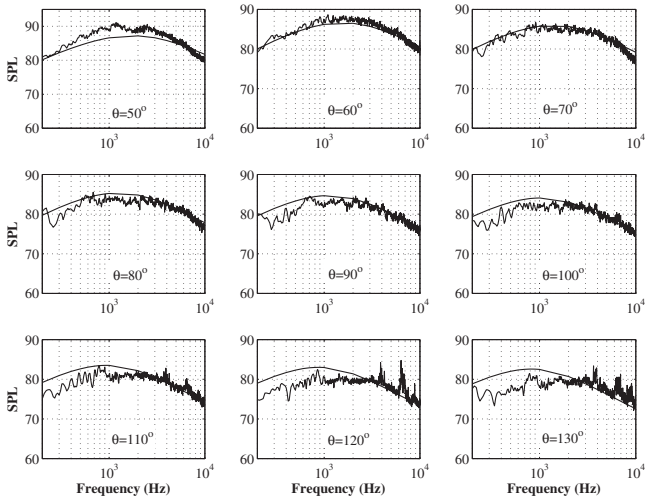


FIG. 8. Comparison of experimental narrowband data with predicted spectral density at different angles to jet axis using energy transfer rate time-scale; based on the MGBK formulations for directivity, $R=50D_J$, $M_J=0.75$, and $D_J=0.05$ m.

prediction and experimental results can also be seen for the observers near both the jet axis rear and forward arcs. Regarding the rear arc (i.e., small angles) this might have originated from one or more causes. First, the radiated noise from large eddies is not properly captured, which leads to poor agreement at the low frequency range for observers located nearby, and accordingly peak frequency mismatch. The failure to accurately compute the low frequency noises actually stems from the failure of the retarded-time assumption, which is questionable for large eddies since their distance to the observer is sometimes less than what is considered as the acoustic far-field. The second possible reason is the use of the high frequency approximated solution of the refraction effect for all frequencies. The third possibility has to do with the definition of the convection velocity. It has been shown before that the peak frequency depends on the large eddy sweeping velocity.[37] Thus, the convection velocity value at some particular place has a direct relation to the peak frequency value. Furthermore, it is well understood from the literature that the large eddies are convected to the downstream at a speed different from the small eddies. This shows that value of the convection speed of the large eddies can play an important role in the accuracy of the noise prediction when the observer is positioned close to the jet axis. Regarding the forward arc (i.e., large angles) the mismatch is more worrying as the trend of the peak frequency is opposite to those shown by data. A different noise production mechanism may be at play here, but this is not certain and this problem requires further exploration.

C. Source distribution

Location of jet noise sources is a far from trivial problem that is of great importance for both the understanding of the noise production and the radiation mechanisms and also for finding new jet noise reduction strategies. According to the nature of jet turbulence it can be readily realized that the high frequency noise sources are mostly aggregated in the vicinity of shear sub-layer, especially where the shear layer is thinner (usually between the nozzle tip and end of the potential core). In contrast, the low frequency sources are associated with larger eddies which are mostly formed in the fully developed region and close to the jet axis. One of the earliest works on this subject was published by Ribner[31] in 1958. This work was very short and its most important result was that the overwhelming bulk of the jet noise is emitted from first eight to ten diameters, which is regarded as the mixing region. It was found that the sound power distribution in the mixing region is constant, while it is proportional to the reciprocal seventh power of axial distance for the fully developed region. Later, Dyer[40] investigated a similar problem and derived a simple procedure for obtaining axial source distributions in a turbulent jet flow. Upon using this procedure one can find the frequency of the sources as a function of location along the jet axis.

In this section we shall obtain a mathematical model for jet noise source distribution which makes use of the CFD-turbulence results as an input for source modeling. The basis is quite similar to the sound intensity calculation. Taking the overall intensity as an integral over the axial extent of the jet,

$$I(R,\omega)=\int I_x(\omega)dx, \tag{32}$$

where $I_x(\omega)$ defines the axial source distribution at each frequency. Since we are only interested in the source distribution from the standpoint of the 90° observer, the following relation can be readily found from Lighthill's equation:

$$I_x(\omega)\propto\int_{\phi=0}^{2\pi}\int_{r=0}^{r_\infty} l_s^3\tau_s u_0^4 e^{-(\omega l_s/2c_0)^2-(\omega\tau_s/2)^2}drd\phi. \tag{33}$$

Source distributions for an isothermal $M_J=0.90$ jet are presented in Fig. 9. The left hand side contour shows the source distribution results found from the JEAN experiment using a polar array technique.[41] The right hand side figure illustrates the axial position at which the source location, I_x, peaks at a particular frequency. In other words, maxima of I_x at each frequency are found. The measured data, shown by " *," are extracted from the left hand side contour. The figure shows a comparison of the results obtained using the analytical solution, given by Eq. (33), when the TDR and TET time-scales are used with the experimental data. It can be seen that the high frequency sources are aggregated in the vicinity of the jet exit, while the low frequency ones are located further downstream after the potential core. In addition, it can be seen from the figure that using the TDR time-scale results in a broader acoustic source regime, while the source domain captured using the TET time-scale shows a better agreement with the experimental data. Although some differences between the experimental data and TET results at low frequency range can be observed, the curve slopes show that they are sharing the same underlying physical phenomena.

The assessment of source distribution for an isothermal $M_J=0.90$ jet flow using different time-scales is also considered here. Source distributions, I_x, for models based on the TDR time-scale, the CTS and the TET time-scale have been

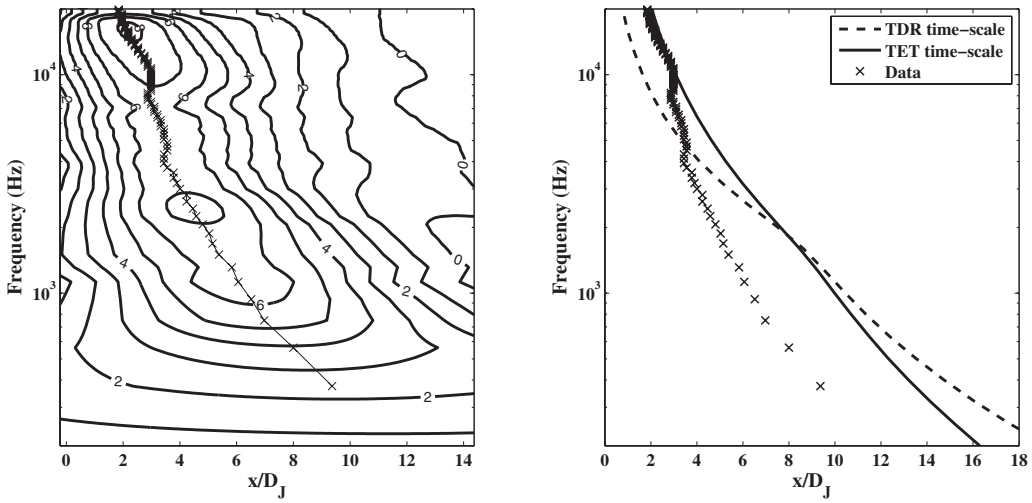


FIG. 9. Source location based on the MGBK method at 90°; $R=50D_J$, $M_J=0.90$, and $D_J=0.05$ m.

calculated using Eq. (33) and are shown in Figs. 10–12. Results are normalized to their maximum values. From studying these figures it can be seen that using the dissipation time-scale, τ_d, results in a source distribution that is roughly of Gaussian shape for all frequencies. Although the Gaussian shape was previously reported in other experimental works,[41,42] the predicted source domain using the TDR time-scale is much longer than those observed in experiment (Fig. 9). As seen in Fig. 6, using a combination of time-scales provides a more realistic results, but Fig. 11 shows source distribution curves that are of markedly different shapes to those observed in real jets.[41,42] However, using the TET time-scale, τ_T, not only leads to a good agreement in the spectra but also reproduces a more physically realistic shape of source distribution, Fig. 12.

V. CONCLUSION

It has been shown that choosing a time-scale that matches the underlying physics of the unsteady turbulent flow significantly improves the prediction of the radiated noise. Three different time-scales have been considered in this paper: the traditional time-scale based on dissipation rate, a CTS (dissipation and production rate), and a new

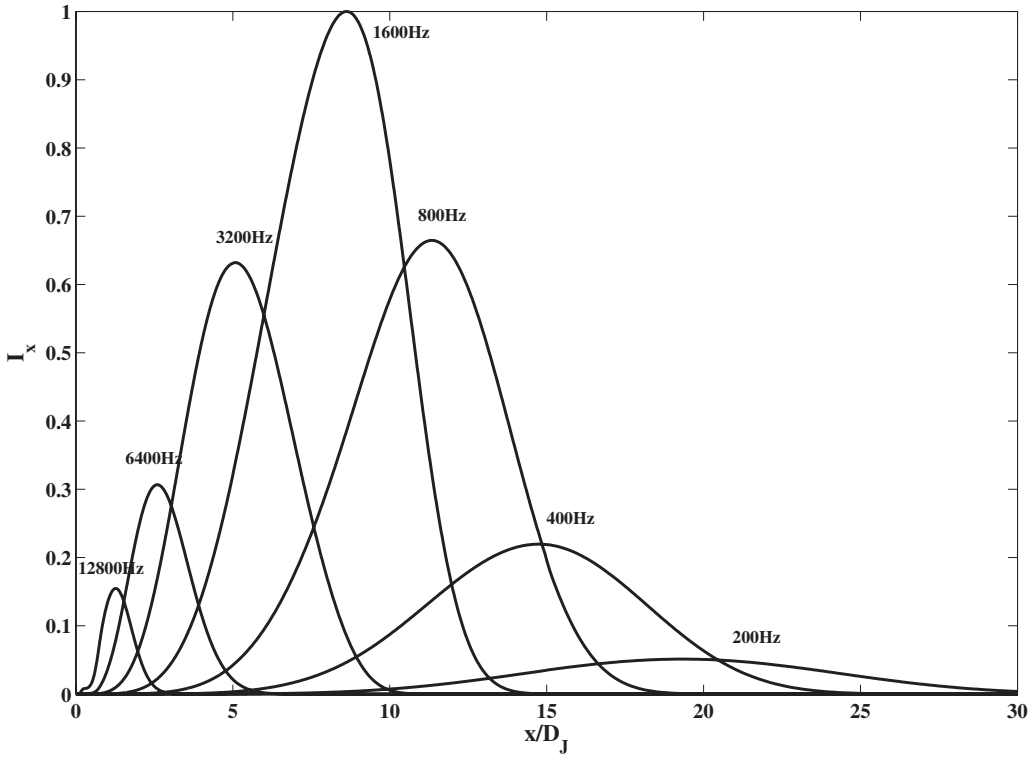


FIG. 10. Predicted noise distribution at each frequency using dissipation rate time-scale and based on the MGBK method for 90°; $R=50D_J$, $M_J=0.90$, and $D_J=0.05$ m.

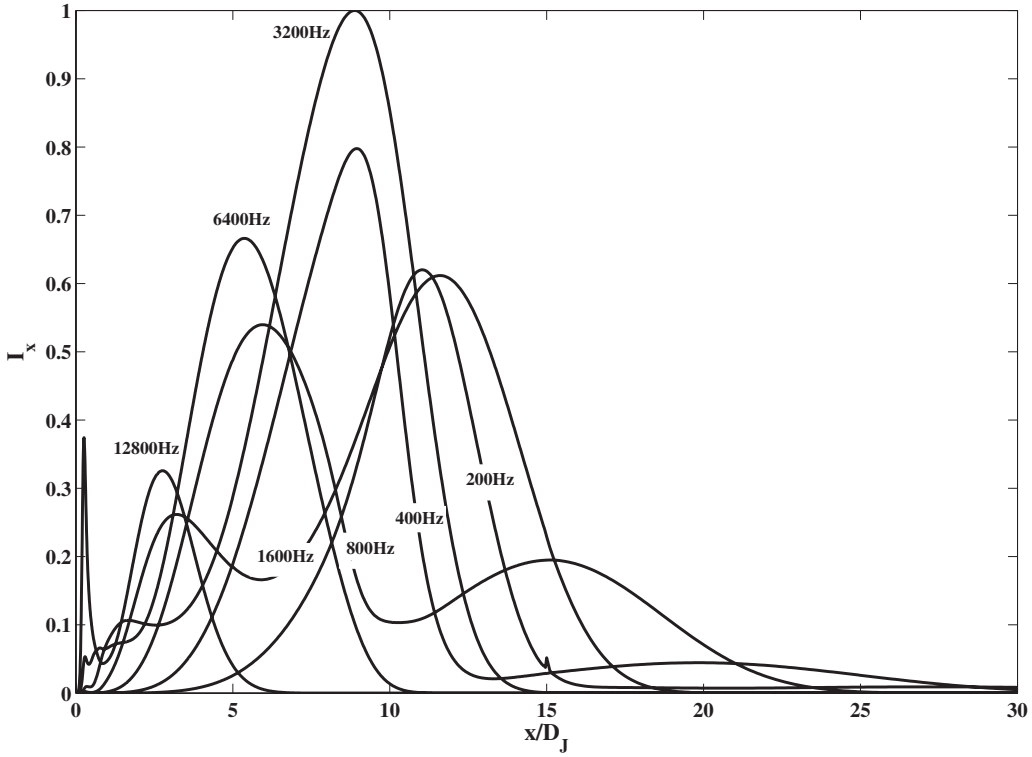


FIG. 11. Predicted noise distribution at each frequency using combined time-scale and based on the MGBK method for 90°; $R=50D_J$, $M_J=0.90$, and $D_J=0.05$ m.

time-scale based on energy transfer rate. It has been shown that the last two time-scales give very good agreement with measured data. More importantly, by using the energy transfer time-scale, the known frequency dependence of the acoustic time-scale no longer needs to be modeled explicitly as it arises as a natural consequence of the underlying physics. It is the authors' contention that acoustic source models using the present approach are likely to prove far more ro-

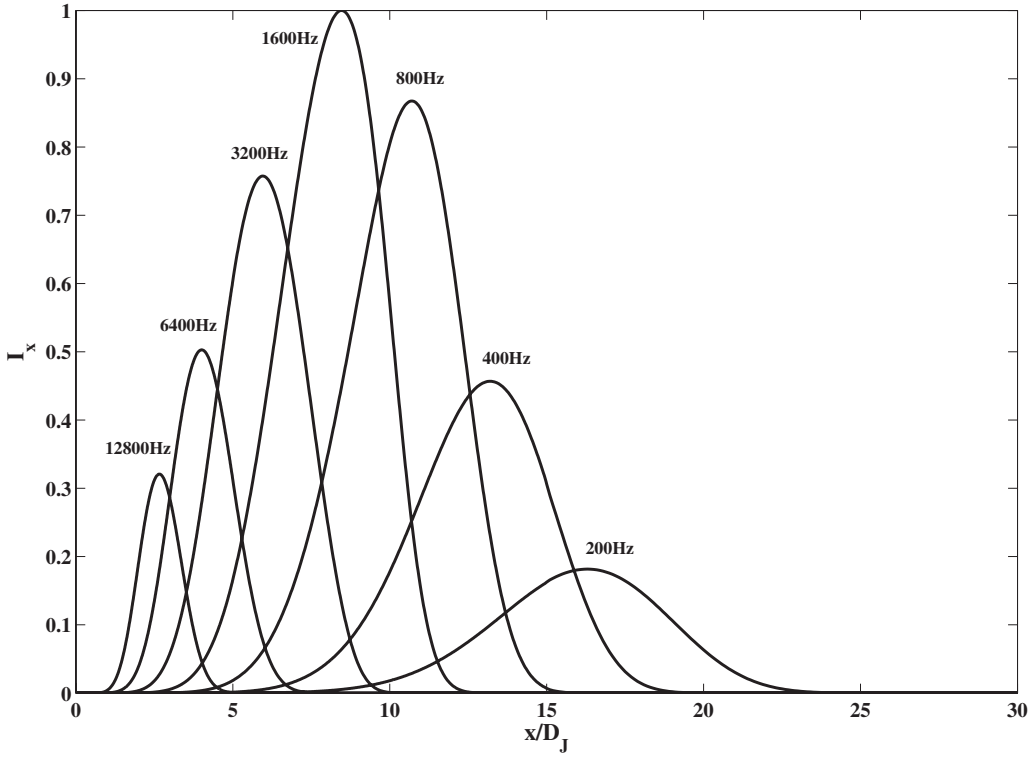


FIG. 12. Predicted noise distribution at each frequency, using energy transfer rate time-scale and based on the MGBK method for 90°; $R=50D_J$, $M_J=0.90$, and $D_J=0.05$ m.

bust than other approaches and may also find application in the prediction of radiated noise from other applications of turbulent flows such as cavity noise, airframe noise, etc.

ACKNOWLEDGMENTS

The first author (M.A.) is grateful for the financial support provided by TNO organization (Delft, The Netherlands). The authors would also like to thank members of Flow-Acoustics group at TNO for their technical discussions over the present paper.

[1]R. Mani, T. F. Balsa, and P. R. Gliebe, "High-velocity jet noise source location and reduction," Federal Aviation Administration Report No. FAA-RD-76-II, 1978.
[2]T. F. Balsa and P. R. Gliebe, "Aerodynamics and noise of coaxial jets," AIAA J. **15**, 1550–1558 (1977).
[3]T. F. Balsa, "The acoustic field of sources in shear flow with application to jet noise: Convective amplification," J. Fluid Mech. **79**, 33–47 (1977).
[4]B. Tester and C. Morfey, "Developments in jet noise modeling-theoretical predictions and comparisons with measured data," J. Sound Vib. **46**, 79–103 (1976).
[5]A. Khavaran, E. A. Krejsa, and C. M. Kim, "Computation of an axisymmetric convergent-divergent nozzle," J. Aircr. **31**, 603–609 (1994).
[6]A. Khavaran, "Role of anisotropy in turbulent mixing noise," AIAA J. **37**, 832–841 (1999).
[7]A. Khavaran and N. Georgiadis, "Aeroacoustics of supersonic elliptic jets," AIAA Paper No. 96-0641 (1996).
[8]A. Hamed, A. Khavaran, and S. Lee, "The flow and acoustic predictions for a high by-pass ratio exhaust nozzle," AIAA Paper No. 98-3257 (1998).
[9]T. J. Barber, L. M. Chiappetta, and S. H. Zysman, "An assessment of jet noise analysis codes for multistream axisymmetric and forced mixer nozzles," AIAA Paper No. 96-0750 (1996).
[10]T. J. Barber, A. Nedungadi, and A. Khavaran, "Predicting the jet noise from high-speed round jets," AIAA Paper No. 2001-0819 (2001).
[11]A. Frendi, W. D. Dorland, T. Nesman, and T. S. Wang, "A jet engine noise measurement and prediction tool," J. Acoust. Soc. Am. **112**, 2036–2042 (2002).
[12]A. Frendi, T. Nesman, and T. S. Wang, "On the effect of time scaling on the noise radiated by an engine plume," J. Sound Vib. **256**, 969–979 (2002).
[13]A. Khavaran and J. Bridges, "Modelling of fine-scale turbulence mixing noise," J. Sound Vib. **279**, 1131–1154 (2005).
[14]A. Khavaran, J. Bridges, and J. B. Freund, "A parametric study of fine-scale turbulence mixing noise," AIAA Paper No. 2002-2419 (2002).
[15]D. W. Wundrow and A. Khavaran, "On the applicability of high-frequency approximation to Lilley's equation," J. Sound Vib. **272**, 793–830 (2004).
[16]M. E. Goldstein, A. Khavaran, and R. E. Musafir, "Jet noise predictions based on two different forms of Lilley's equation, Part 1: Basic theory," NASA Report No. TM-2005-213829-Part 1, National Aeronautics and Space Administration, Washington, DC, 2005.
[17]M. E. Goldstein, A. Khavaran, and R. E. Musafir, "Jet noise predictions based on two different forms of Lilley's equation, Part 2: Acoustic predictions and comparison with data," NASA Report No. TM-2005-213829-Part 2, National Aeronautics and Space Administration, Washington, DC, 2005.
[18]R. H. Self, "Jet noise prediction using the Lighthill acoustic analogy," J. Sound Vib. **275**, 757–768 (2004).
[19]R. H. Self and A. Bassetti, "A RANS based jet noise prediction scheme," AIAA Paper No. 2003-3325 (2003).
[20]G. J. Page, J. J. McGuirk, M. Hossain, R. H. Self, and A. Bassetti, "A CFD coupled acoustics approach for coaxial jet noise," AIAA Paper No. 2003-3286 (2003).
[21]C. K. W. Tam and L. Auriault, "Jet mixing noise from fine-scale turbulence," AIAA J. **37**, 145–153 (1999).
[22]C. K. W. Tam, N. N. Pastouchenko, and L. Auriault, "Effect of forward flight on jet mixing noise from fine-scale turbulence," AIAA J. **39**, 1261–1269 (2001).
[23]C. K. W. Tam and N. N. Pastouchenko, "Noise from fine-scale turbulence of nonaxisymmetric jets," AIAA J. **40**, 456–464 (2002).
[24]C. K. W. Tam, N. N. Pastouchenko, and K. Viswanathan, "Fine-scale turbulence noise from hot jets," AIAA J. **43**, 1675–1683 (2005).
[25]S. Karabasov, M. Afsar, T. Hynes, A. Dowling, W. McMullan, C. Pokora, G. Page, and J. McGuirk, "Using large eddy simulation within an acoustic analogy approach for jet noise modelling," AIAA Paper No. 2008-2985.
[26]M. Goldstein, "A generalized acoustic analogy," J. Fluid Mech. **488**, 315–333 (2003).
[27]P. J. Morris and F. Farassat, "Acoustic analogy and alternative theories for jet noise prediction," AIAA J. **40**, 671–680 (2002).
[28]M. J. Fisher and P. O. A. L. Davies, "Correlation measurements in a non-frozen pattern of turbulence," J. Fluid Mech. **18**, 97–116 (1964).
[29]P. J. Morris and S. Boluriaan, "The prediction of jet noise from CFD data," AIAA Paper No. 2004-2977 (2004).
[30]M. Harper-Bourne, "Jet near field noise prediction," AIAA Paper No. 99-1838 (1999).
[31]H. S. Ribner, "Strength distribution of noise source along a jet," J. Acoust. Soc. Am. **30**, 876 (1958).
[32]G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge University Press, Cambridge, 1999).
[33]M. S. Uberoi, "Quadruple velocity correlations and pressure fluctuations in isotropic turbulence," J. Aeronaut. Sci. **20**, 197–204 (1953).
[34]H. S. Ribner, "Theory of two-point correlations of jet noise," NASA Report No. TND-8330, National Aeronautics and Space Administration, Washington, DC, 1976.
[35]H. S. Ribner, "Quadrupole correlation governing the pattern of jet noise," J. Fluid Mech. **38**, 1–24 (1969).
[36]J. O. Hinze, *Turbulence* (DCW Industries, New York, 1975).
[37]R. Rubinstein and Y. Zhou, "The frequency spectrum of sound radiated by isotropic turbulence," Phys. Lett. A **267**, 379–383 (2000).
[38]A. Khavaran and E. A. Krejsa, "On the role of anisotropy in turbulent mixing noise," AIAA Paper No. 98-2289 (1998).
[39]P. Jordan and Y. Gervias, "Modeling self and shear noise mechanisms in inhomogeneous, anisotropic turbulence," J. Sound Vib. **279**, 529–555 (2005).
[40]I. Dyer, "Distribution of sound sources in a jet stream," J. Acoust. Soc. Am. **31**, 1016–1022 (1959).
[41]J. Battaner-Moro, "Report on automated source breakdown for coaxial and single jet noise measurements," ISVR Internal Report No. 03/10, Institute of Sound and Vibration Research, Southampton, UK, 2003.
[42]C. K. W. Tam, N. N. Pastouchenko, and R. H. Schlinker, "Noise source distribution in supersonic jets," J. Sound Vib. **291**, 192–201 (2006).

Statistics of normal mode amplitudes in an ocean with random sound-speed perturbations: Cross-mode coherence and mean intensity

John A. Colosi
Department of Oceanography, Naval Postgraduate School, Monterey, California 93943

Andrey K. Morozov
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 20 November 2008; revised 13 May 2009; accepted 14 May 2009)

In this paper Creamer's [(1996). J. Acoust. Soc. Am. **99**, 2825–2838] transport equation for the mode amplitude coherence matrix resulting from coupled mode propagation through random fields of internal waves is examined in more detail. It is shown that the mode energy equations are approximately independent of the cross mode coherences, and that cross mode coherences and mode energy can evolve over very similar range scales. The decay of cross mode coherence depends on the relative mode phase randomization caused by coupling and adiabatic effects, each of which can be quantified by the theory. This behavior has a dramatic effect on the acoustic field second moments like mean intensity. Comparing estimates of the coherence matrix and mean intensity from Monte Carlo simulation, and the transport equations, good agreement is demonstrated for a 100-Hz deep-water example. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158818]

PACS number(s): 43.30.Bp, 43.30.Re, 43.60.Cg [WLS] Pages: 1026–1035

I. INTRODUCTION

Numerous observations and numerical studies have shown that random internal-wave-induced sound-speed perturbations in both shallow- and deep-water environments can cause significant changes in the mean acoustical intensity relative to the unperturbed intensity. Examples in deep-water problems are the depth broadening of the acoustic finale (Worcester *et al.*, 1994; Colosi *et al.*, 1994; Colosi and Flatté, 1996; Worcester *et al.*, 1999) and the so-called deep shadow zone arrivals (Dushaw *et al.*, 1999; Flatté and Colosi, 2008; Van Uffelen *et al.*, 2009). For shallow-water problems, on the other hand, the acoustic arrivals are seen to have significant time spreading (Tielburger *et al.*, 1997; Fredricks *et al.*, 2005) which could be due to both random linear internal waves and nonlinear internal solitary waves. Lacking, however, has been a theoretical understanding of the dominant acoustic scattering physics leading to the mean redistribution of the acoustical energy. This theoretical understanding is prerequisite to formulating a useful reduced physics model to predict such effects without having to resort to time consuming Monte Carlo simulations. Available theoretical models to understand this behavior have fallen short: Path integral results have been shown to break down at long range due to the instability of the unperturbed ray path (Colosi *et al.*, 1999; Beron-Vera *et al.*, 2003), and ray chaos methods cannot accurately predict intensity. Coupled mode approaches are thus promising. The seminal work of Creamer (1996) first introduced a transport equation for the cross mode coherence matrix, a necessary ingredient for the calculation of second moments like mean intensity. However in that work, the coherence terms were neglected in order to study the asymptotic evolution of mode energy, the diagonal of the coherence matrix. Later work by Voronovich and Ostashev (2006, 2009) has refocused on the coherence matrix and has included out-of-plane coupling effects to predict horizontal coherence.

As a first step toward gaining further physical understanding of the aforementioned observed intensity, theoretical results for the mean intensity at a single frequency are presented in this paper utilizing the simple two-dimensional (2D) results of Creamer (1996). The normal mode framework developed by Dozier and Tappert (1978a, 1978b) and extended by Dozier (1983) and Creamer (1996) takes advantage of the facts that (1) the coupling is weak (i.e., there is small angle forward scattering) and (2) that the Markov approximation is valid. Within this framework, this paper delves somewhat more deeply into the transport equation for the cross mode coherence matrix to underscore some points not previously appreciated. In particular, it is shown that the decay of cross mode coherences is controlled by terms in the equations associated with adiabatic and mode coupling induced phase randomization. In the deep-water cases addressed here, the coupling effects dominate, and thus mode energy and cross mode coherence evolve over similar range scales; a result contrary to conventional speculation in the literature (Dozier and Tappert, 1978a, 1978b; Creamer, 1996). Second it is shown that the evolution equation for the modal energies is insensitive to the actual values of the off-diagonal terms or cross mode coherences. This result explains the successful predictions of modal energies using theory which neglects cross mode coherence terms (Dozier and Tappert, 1978a, 1978b; Creamer, 1996). Finally using Monte Carlo simulation techniques for a deep-water example

the cross mode coherence matrix evolution equations are shown to produce accurate predictions of mean intensity out to very long range.

The outline of this paper is as follows. Section II describes the evolution equation for the cross mode coherence matrix. Section III addresses the mode energy observable and its connection to the off-diagonal cross mode coherences. Section IV gives a two mode example which helps understand the nature of the solutions to the evolution equation, while Sec. V demonstrates the accuracy of the method for predicting mean intensity. Section VI gives brief summary and conclusions.

II. 2D COUPLED MODE THEORY

The acoustic pressure at frequency ω is expressed in terms of the normal mode expansion (Creamer, 1996)

$$p(r,z;\omega)=\sum_{n=1}^{N}\frac{a_n(r)\phi_n(z)}{\sqrt{k_n r}}, \tag{1}$$

where the unperturbed normal mode equation, $\rho_0(z)\partial/\partial z(\rho_0^{-1}(z)\partial\phi_n/\partial z)+(\bar{k}^2(z)-k_n^2)\phi_n=0$, gives the eigenmodes ϕ_n and eigenwavenumbers k_n, and all the variability is contained in the mode amplitude a_n. Here the background density is $\rho_0(z)$, and sound speed is $c(r,z)=\bar{c}(z)+\delta c(r,z)$, with $\bar{k}(z)=\omega/\bar{c}(z)$. Without any loss of generality the modal wavenumber can be considered to have a small complex component from attenuation such that $l_n=k_n+i\alpha_n$. In addition it is useful to define a reduced modal amplitude quantity that removes the rapid oscillations in range so that $\psi_n(r)=a_n(r)e^{-il_n r}$ with the result that the mean intensity for weak attenuation is given by

$$\langle I(r,z)\rangle=\langle|p(r,z)|^2\rangle=\sum_{n=1}^{N}\sum_{p=1}^{N}\langle\psi_n\psi_p^*(r)\rangle\frac{e^{i(l_n-l_p^*)r}}{r}\frac{\phi_n(z)\phi_p(z)}{\sqrt{k_n k_p}}, \tag{2}$$

where $\langle\psi_n\psi_p^*(r)\rangle$ is the cross mode coherence matrix. Here the importance of the cross mode coherences to the mean intensity observable is clearly evident. Figure 1 shows example calculations of deep-water 100-Hz directed beam propagation with and without ocean internal-wave-induced sound-speed perturbations (for simulation details see the Appendix). The spatial coherence of the perturbed and mean intensity beams in depth and range for the first 500–1000 km shows visually that the cross mode coherence is not decaying rapidly with range. After 1000-km range a smooth nearly featureless mean intensity pattern is seen to be due to an absence of cross mode coherence. Morozov and Colosi (2005, 2007) showed other numerical examples of 100-, 125-, and 250-Hz directed beam propagation through ocean internal waves in a deep-water environment which have a similar character. It should be noted that some simple analytic estimates describing cross mode coherences have been derived using the ray-mode duality (Virovlyanskii, 1989; Virovlyanskii *et al.*, 1989).

Dozier (1983) and Creamer (1996) showed that the coupled mode equations for small attenuation are

$$\frac{d\psi_n}{dr}=-i\sum_{m=1}^{N}\rho_{mn}(r)e^{il_{mn}r}\psi_m(r), \tag{3}$$

where $l_{mn}=l_m-l_n=k_{mn}+i\alpha_{mn}$, and the symmetric coupling matrix $\rho_{mn}(r)$ is given by

$$\rho_{mn}(r)=\frac{k_0^2}{\sqrt{k_n k_m}}\int_0^D\frac{\phi_n(z)\phi_m(z)}{\rho_0(z)}\mu(r,z)dz. \tag{4}$$

Here k_0 is a reference wavenumber, D is the water depth, and $\mu(r,z)=\delta c(r,z)/c_0$ is the random fractional sound-speed fluctuation, assumed zero in the seabed. This equation accurately models wide angle propagation in the sound channel.

A. Correlation function of the coupling matrix

The correlation function of the coupling matrix will be central to this analysis. Writing $\mu(z,r)$ in terms of a linear superposition of internal waves with mode number j and Cartesian horizontal wavenumber k_r, the coupling matrix becomes

$$\rho_{mn}(r)=\frac{\mu_0 k_0^2}{\sqrt{k_m k_n}}\int_0^D dz\left(\frac{N(z)}{N_0}\right)^{3/2}\frac{\phi_n(z)\phi_m(z)}{\rho_0(z)}\times\sum_{j=1}^{\infty}h_j\sin[\pi j\hat{z}(z)]\int_{-\infty}^{\infty}dk_r b_j(k_r)e^{ik_r r}, \tag{5}$$

where $N(z)$ is the buoyancy frequency profile, μ_0 is a reference rms fractional sound speed, h_j and $b_j(k_r)$ are complex Gaussian random variables, and $\hat{z}(z)$ is the Wentzel-Kramers-Brillouin-Jeffreys (WKBJ) stretched vertical coordinate (Colosi and Brown, 1998). Multiplying by $\rho_{qp}^*(r-\xi)$ and taking the expectation value the result is

$$\Delta_{mn,qp}(\xi)=\langle\rho_{mn}(r)\rho_{qp}^*(r-\xi)\rangle=\sum_{j=1}^{\infty}\langle|h_j|^2\rangle G_{mn}(j)G_{qp}(j)\int_{-\infty}^{\infty}dk_r\langle|b_j(k_r)|^2\rangle e^{-ik_r\xi}, \tag{6}$$

where

$$G_{mn}(j)=\mu_0 k_0^2\sqrt{\frac{2}{k_n k_m}}\int_0^D dz\left(\frac{N(z)}{N_0}\right)^{3/2}\times\sin[\pi j\hat{z}(z)]\frac{\phi_n(z)\phi_m(z)}{\rho_0(z)}, \tag{7}$$

and from the Garrett-Munk (GM) spectrum the expressions are

$$\langle|h_j|^2\rangle=\frac{1}{M_j}\frac{1}{j^2+j_*^2},$$

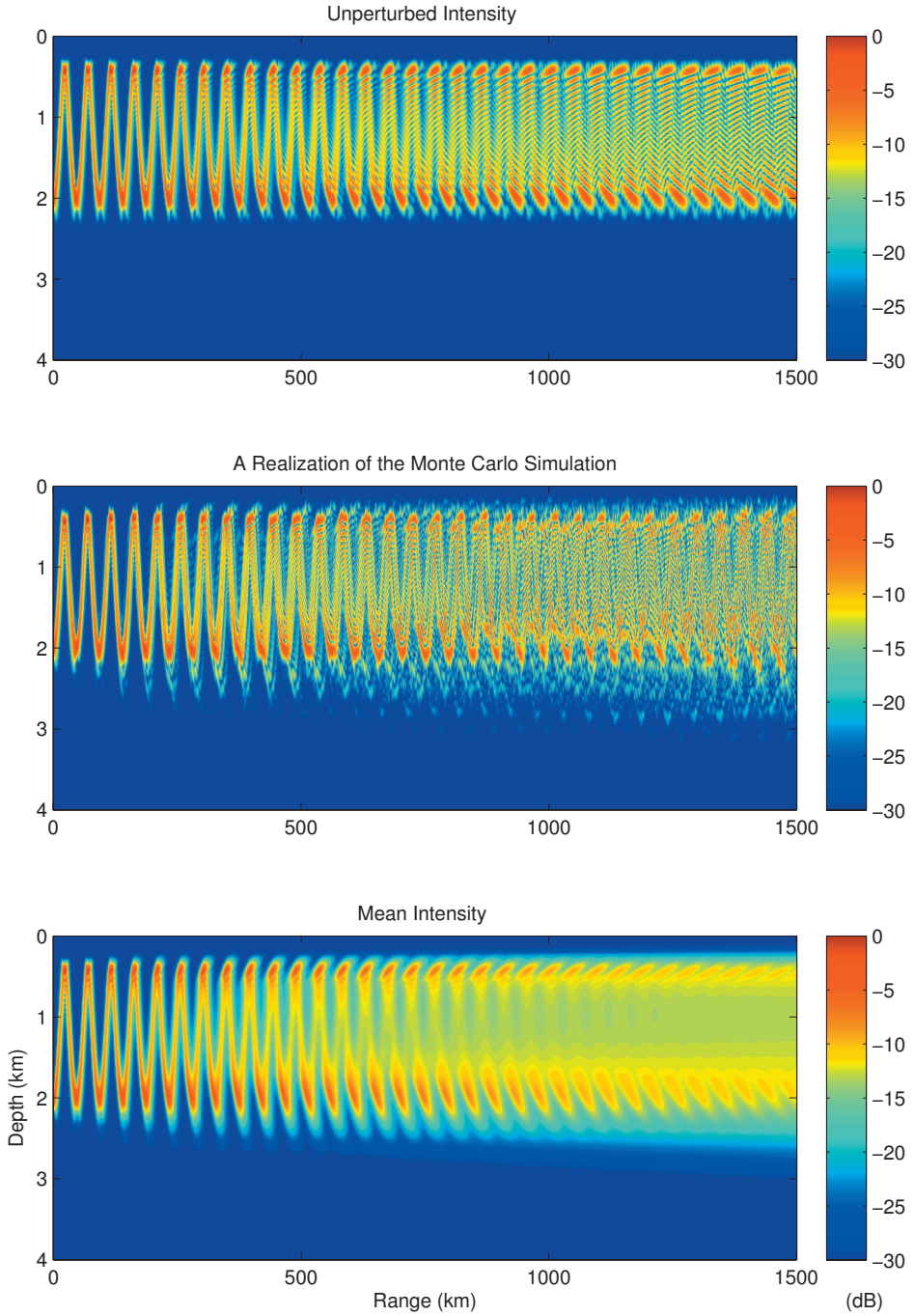


FIG. 1. Acoustic beams from 100-Hz numerical simulations using the canonical ocean described in the Appendix. The upper panel displays the unperturbed acoustical beam while the middle panel shows a realization of the beam propagation through a random realization of internal-wave-induced sound-speed perturbations. The lower panel shows the mean intensity averaged over 500 realizations of the internal wave field. Cylindrical spreading is not included in the mean intensity. The intensity scale is decibel referenced to the maximum value.

$$\langle |b_j(k_r)|^2 \rangle = \frac{2}{\pi^2}\left[\frac{k_j}{k_r^2+k_j^2} + \frac{1}{2}\frac{k_r^2}{(k_r^2+k_j^2)^{3/2}}\log\left(\frac{(k_r^2+k_j^2)^{1/2}+k_j}{(k_r^2+k_j^2)^{1/2}-k_j}\right)\right] \simeq \frac{1}{\pi}\frac{\hat{k}_j}{\hat{k}_j^2+k_r^2}, \tag{8}$$

with $M_j^{-1}=\Sigma_{j=1}^{\infty}(j^2+j_*^2)^{-1}$, $k_j=\pi f j/N_0 B$, $\hat{k}_j^2=2k_j^2$, $N_0 B=\int_0^D N(z)dz$, and f is the Coriolis parameter (Morozov and Colosi, 2007). Using the approximation in the last line of Eq. (8) a useful and accurate approximate form of the correlation function is obtained, namely,

$$\Delta_{mn,qp}(\xi) = \sum_{j=1}^{\infty} \langle |h_j|^2 \rangle G_{mn}(j) G_{qp}(j) e^{-\hat{k}_j|\xi|}. \tag{9}$$

The matrices involved in the cross-mode coherence transport equation involve integrals over the correlation function $\Delta_{mn,qp}$ of the form

$$I_{mn,qp} = \int_0^{\infty} d\xi \Delta_{mn,qp}(\xi) e^{il_{pq}\xi}. \tag{10}$$

Some useful symmetry properties of Eq. (10) are $I_{mn,qp}=I_{nm,qp}$ and $I_{mn,pq}=I^*_{mn,qp}$. For zero attenuation, the real part of this function is closely related to the wavenumber spectrum of the coupling matrix, that is,

$$\frac{1}{\pi}\,\mathrm{Re}(I_{mn,qp}) = \langle \hat{\rho}_{mn}\hat{\rho}^{*}_{qp}(k_{pq})\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\xi \Delta_{mn,qp}(\xi)e^{ik_{pq}\xi}. \tag{11}$$

This function then is associated with resonance conditions that pick out specific internal wave wavenumbers that contribute to mode coupling. In addition for $q=p$ this function has the important physical interpretation in terms of a correlation length in the horizontal, that is, $L_H(mn,p) = I_{mn,pp}/\Delta_{mn,pp}(0)$. When attenuation is added to the picture the resonance is shifted so that different wavenumbers contribute to the coupling (Colosi, 2008). Using the approximate correlation function in Eq. (9), the transport matrix has a useful analytic form

$$I_{mn,qp} = \sum_{j=1}^{\infty} \langle |h_j|^2\rangle G_{mn}(j)G_{qp}(j)\frac{\hat{k}_j + \alpha_{pq} + ik_{pq}}{(\hat{k}_j + \alpha_{pq})^2 + k_{pq}^2}, \quad |\alpha_{pq}| < \hat{k}_j, \tag{12}$$

where the attenuation shifting of the resonance is evident.

B. Cross-mode coherence transport equation: A heuristic derivation

Creamer's (1996) transport equation for cross-mode coherence was derived using the assumptions of small angle scattering and the Markov approximation, yet while the derivation is formally rigorous, relying on operator methods, it is somewhat opaque with regard to the physical assumptions. Thus we provide here a heuristic derivation of the same equation that, for some, may better elucidate the physical concepts.

Using Eq. (3) the cross-mode coherence equation is

$$\frac{d\psi_n\psi_p^*}{dr} = \psi_p^*\frac{d\psi_n}{dr} + \psi_n\frac{d\psi_p^*}{dr} = -i\sum_{m=1}^{N}(\rho_{mn}(r)e^{il_{mn}r}\psi_m\psi_p^* - \rho_{mp}(r)e^{-il_{mp}^{*}r}\psi_n\psi_m^*). \tag{13}$$

For small angle scattering the coupling matrix itself is small, so it is useful to solve Eq. (13) by iteration (Sakurai, 1985). Each order of the iterated solution is interpreted physically in terms of multiple scattering, so to second order (second order scattering) the solution is

$$\begin{aligned}\psi_n\psi_p^*(r) &= \psi_n\psi_p^*(0) - \sum_{m=1}^{N}\sum_{q=1}^{N}\int_0^r dr'\int_0^{r'} dr''(\psi_q\psi_p^*(0)\rho_{mn}(r')\rho_{qm}(r'')\\ &\times e^{i(l_{mn}r'+l_{qm}r'')} - \psi_m\psi_q^*(0)\rho_{mn}(r')\rho_{qp}(r'')e^{i(l_{mn}r'-l_{qp}^{*}r'')}\\ &- \psi_q\psi_m^*(0)\rho_{mp}(r')\rho_{qn}(r'')e^{-i(l_{mp}^{*}r'-l_{qn}r'')}\\ &+ \psi_n\psi_q^*(0)\rho_{mp}(r')\rho_{qm}(r'')e^{-i(l_{mp}^{*}r'+l_{qm}^{*}r'')}),\end{aligned} \tag{14}$$

where r is the range. The first order terms have been left out of Eq. (14) because they will drop out when the expectation value is taken since $\langle\rho_{mn}\rangle=0$. Here the objective is to derive an evolution equation for the cross-mode coherence matrix as a function of range which may include many correlation lengths of the random sound-speed structure. We make use of Eq. (14) and conceptually consider the change in the coherence matrix over a distance r when some initial field $\psi_n\psi_p^*(0)$ is incident upon a section of sound-speed fluctuations. The range r is loosely defined such that $r \gg r_c$ with r_c the largest correlation length of any of the coupling matrix terms. Taking the expectation value of Eq. (14) and making the reasonable assumption that the initial coherence matrix is uncorrelated with the subsequent coupling matrices, the result is

$$\begin{aligned}\gamma_{np}(r) &= \gamma_{np}(0) - \sum_{m=1}^{N}\sum_{q=1}^{N}\int_0^r dr'\int_0^{r'} dr''(\gamma_{qp}(0)\Delta_{mn,qm}(\xi)\\ &\times e^{i(l_{mn}r'+l_{qm}r'')} - \gamma_{mq}(0)\Delta_{mn,qp}(\xi)e^{i(l_{mn}r'-l_{qp}^{*}r'')}\\ &- \gamma_{qm}(0)\Delta_{mp,qn}(\xi)e^{-i(l_{mp}^{*}r'-l_{qn}r'')}\\ &+ \gamma_{nq}(0)\Delta_{mp,qm}(\xi)e^{-i(l_{mp}^{*}r'+l_{qm}^{*}r'')}),\end{aligned} \tag{15}$$

where $\gamma_{np}=\langle\psi_n(r)\psi_p^*(r)\rangle$, $\Delta_{mn,qp}$ are the correlation functions of the coupling matrices previously discussed, and $\xi=r'-r''$. Changing the r'' integration variable to ξ, assuming that $r \gg r_c$ so that the ξ integration limit can go to infinity, and differentiating with respect to r the final transport equation is

$$\begin{aligned}\frac{d\gamma_{np}(r)}{dr} &= -\sum_{m=1}^{N}\sum_{q=1}^{N}(\gamma_{qp}(r)I_{mn,qm}e^{il_{qn}r}\\ &- \gamma_{mq}(r)I_{mn,qp}^{*}e^{i(l_{mn}r-l_{qp}^{*}r)}\\ &- \gamma_{qm}(r)I_{mp,qn}e^{-i(l_{mp}^{*}r-l_{qn}r)} + \gamma_{nq}(r)I_{mp,qm}^{*}e^{-il_{qp}^{*}r}),\end{aligned} \tag{16}$$

where the $I_{mn,qm}$ matrices are given by Eq. (10),[1] Importantly on the right-hand side of Eq. (16) the initial coherence matrices $\gamma_{np}(0)$ have been replaced with the value at range r. This approximation is justified for two reasons. First, considering propagation through multiple correlation lengths the initial field must change with range, and second $\gamma_{np}(0) \simeq \gamma_{np}(r)$ because the first correction is at second order. The derivation of this transport equation involves the Markov approximation (Van Kampen, 1981; Creamer, 1996; Henyey and Ewart, 2006) since it is assumed that $r \gg r_c$ but in the end r is taken to be infinitesimally small; the coupling matrices are thus assumed to be delta correlated. Also, as previously discussed, the present derivation assumes small-angle scattering because only terms to second order are retained in Eq. (13). The conceptual derivation just described can be obtained more rigorously using operator methods (see Van Kampen, 1981 and Creamer, 1996), and Eq. (16) can also be derived by assuming Gaussian statistics for ρ_{mn} (Morozov and Colosi, 2007).

The right-hand side of this equation has factors that depend on r, making numerical or further theoretical progress difficult. The r dependence can be removed by undoing the transformation $\psi_n=a_ne^{il_nr}$; the result is[2]

$$\frac{d\langle a_n a_p^* \rangle(r)}{dr} = i(l_n - l_p^*)\langle a_n a_p^* \rangle - \sum_{m=1}^{N}\sum_{q=1}^{N}(\langle a_q a_p^* \rangle I_{mn,qm} - \langle a_m a_q^* \rangle I_{mn,qp}^* - \langle a_q a_m^* \rangle I_{mp,qn} + \langle a_n a_q^* \rangle I_{mp,qm}^*). \quad (17)$$

Equation (17) has been shown to be consistent with the cross-mode coherence equation obtained by Voronovich and Ostashev (2009) when they neglect azimuthal coupling (Voronovich and Ostashev, personal communication). The determination of regimes in which the present 2D approximation breaks down and a fully three-dimensional 3D treatment is needed is an interesting area of future research.

III. MODE ENERGY

An acoustic observable that has received much attention since the seminal paper by Dozier and Tappert (1978a, 1978b) is the mode energy or $\langle |a_n|^2 \rangle$ (Creamer, 1996; Colosi and Flatte, 1996; Tielburger *et al.*, 1997; Wage *et al.*, 2005). For the case of weak attenuation a state of equipartitioning of modal energy is understood to be a modal manifestation of full saturation (Flatté *et al.*, 1979). However, theoretical work to date has not been able to address the influences of the cross-mode coherences on the modal energy evolution. From Eq. (17) the evolution of the mode energy is given by

$$\frac{d\langle |a_n|^2 \rangle}{dr} + 2\alpha_n \langle |a_n|^2 \rangle = \sum_{m=1}^{N}(\langle |a_m|^2 \rangle f_{mn} - \langle |a_n|^2 \rangle g_{mn}) - \sum_{m=1}^{N}\sum_{q=1,q\neq n}^{N} 2\,\mathrm{Re}(\langle a_q a_n^* \rangle I_{mn,qm}) + \sum_{m=1}^{N}\sum_{q=1,q\neq m}^{N} 2\,\mathrm{Re}(\langle a_q a_m^* \rangle I_{mn,qn}), \quad (18)$$

where the diagonal contributions (single sum terms) have been separated from the cross-mode coherence contributions (double sums). In this equation the important matrices are

$$f_{mn} = 2\,\mathrm{Re}(I_{mn,nm}) = 2\int_0^\infty d\xi \Delta_{mn,mn}(\xi)\cos(k_{mn}\xi)e^{+\alpha_{mn}\xi} \quad (19)$$

and $g_{mn}=2\,\mathrm{Re}(I_{mn,mn})$. Because the matrices f_{mn} and g_{mn} are positive definite quantities, whereas the terms $\mathrm{Re}(\langle a_q a_n^* \rangle I_{mn,qm})$ and $\mathrm{Re}(\langle a_q a_m^* \rangle I_{mn,qn})$ are highly oscillatory functions of the indices, the mode energy evolution equations are insensitive to the cross-mode coherences; this fact will be exemplified shortly with a numerical example. Therefore, ignoring the cross-mode coherence terms and assuming that the attenuation is weak enough that the variation in the exponentials in f_{mn} and g_{mn} are weak over the short correlation length of the coupling matrix[3] (Creamer, 1996), the mode energy equations become

$$\frac{\partial \langle |a_n|^2 \rangle}{\partial r} + 2\alpha_n \langle |a_n|^2 \rangle = \sum_{m=1}^{N} f_{mn}(\langle |a_m|^2 \rangle - \langle |a_n|^2 \rangle). \quad (20)$$

In this approximation the matrix $f_{mn}=g_{mn}=2\pi\langle |\hat{\rho}_{mn}(k=k_{mn})|^2 \rangle$ reveals the important resonance condition described by Dozier and Tappert (1978a, 1978b) in which only internal waves whose horizontal wavenumber matches the beat wavenumber k_{mn} contribute to the coupling between modes m and n. Here $\langle |\hat{\rho}_{mn}(k=k_{mn})|^2 \rangle$ is the wavenumber spectrum of the coupling matrix between modes n and m. Equation (20) was obtained by Creamer (1996) and with $\alpha_n=0$, Dozier and Tappert's (1978a, 1978b) "master equations" are recovered.

For the deep-water example demonstrated in Fig. 1, a comparison of the mode energies computed by the Dozier Tappert's (1978a, 1978b) equations and the full evolution equations including the coherences is shown in Fig. 2 (see the Appendix for computational and environmental parameters). Here it is seen that aside from some small oscillations the Dozier Tappert's (1978a, 1978b) results closely models the solution to the full equations. In Fig. 2 mode energies in the neighborhood of mode 20 are considered since our initial condition gives mode 20 the largest initial energy. Other calculations that have been done at different frequencies, source depths, and point source initial conditions show the same behavior as in Fig. 2, and therefore the result appears to be quite robust.

IV. A TWO-MODE EXAMPLE

A two-mode example is useful toward gaining an understanding of the structure of the evolution equations and the relative rates of change in mode energy and coherence as a function of range. Using Eq. (17) and the approximation from Sec. III that the attenuation is weak enough to be ignored in the $I_{mn,pq}$ terms, it is found that several terms simplify to yield the three evolution equations:

$$\frac{d\langle |a_1|^2 \rangle}{dr} + 2\alpha_1 \langle |a_1|^2 \rangle = 2\,\mathrm{Re}(I_{12,12})(\langle |a_2|^2 \rangle - \langle |a_1|^2 \rangle) + 2(I_{21,11} - I_{21,22})\mathrm{Re}(\langle a_1 a_2^* \rangle), \quad (21)$$

$$\frac{d\langle |a_2|^2 \rangle}{dr} + 2\alpha_2 \langle |a_2|^2 \rangle = 2\,\mathrm{Re}(I_{12,12})(\langle |a_1|^2 \rangle - \langle |a_2|^2 \rangle) + 2(I_{21,22} - I_{21,11})\mathrm{Re}(\langle a_1 a_2^* \rangle), \quad (22)$$

$$\frac{d\langle a_1 a_2^* \rangle}{dr} + i(l_2^* - l_1)\langle a_1 a_2^* \rangle = -(I_{11,11} + I_{22,22} - 2I_{11,22} + 2I_{12,12})\langle a_1 a_2^* \rangle + (I_{11,21} - I_{22,21})\langle |a_1|^2 \rangle + (I_{22,21} - I_{11,21})\langle |a_2|^2 \rangle. \quad (23)$$

In Eqs. (21) and (22) as previously noted the effect of the coherence should be small because $\mathrm{Re}(\langle a_1 a_2^* \rangle)$ oscillates.

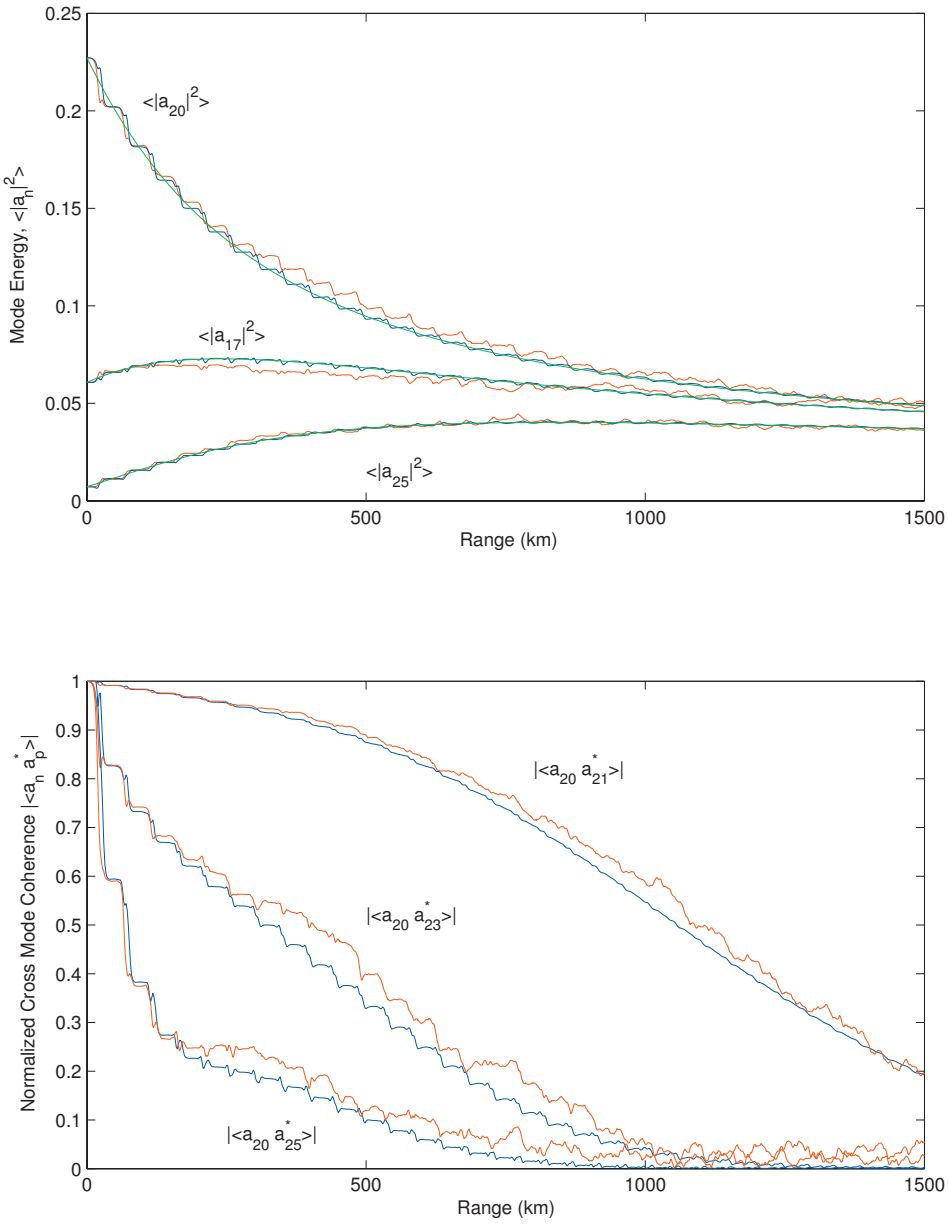


FIG. 2. Simulation and theoretical predictions for mode energy (upper panel) and cross-mode coherence (lower panel) for a few modes. In the upper panel energy for modes 17, 20, and 25 are plotted for the Monte Carlo simulation (red), Dozier–Tappert model [Eq. (20); green], and the full coherence matrix calculation [Eq. (17); blue]. In the lower panel cross-mode coherences between mode 20 and modes 21, 23, and 25 are shown. Monte Carlo simulation results are shown in red, while model results from Eq. (17) are shown in blue.

Further it can be noted in this two-mode example that the terms $I_{21,11}-I_{21,22}$ and $I_{11,21}-I_{22,21}$ can also be small if the projection of the sound-speed fluctuations on the two-mode functions [quantified through the functions $G_{mn}(j)$] is not a strong function of mode number. If this is the case the last terms in Eqs. (21)–(23) can be ignored, yielding coupled equations for the mode energies and a single equation for the coherence,

$$\frac{d\langle|a_1|^2\rangle}{dr} + 2\alpha_1\langle|a_1|^2\rangle = 2\,\mathrm{Re}(I_{12,12})(\langle|a_2|^2\rangle - \langle|a_1|^2\rangle), \quad (24)$$

$$\frac{d\langle|a_2|^2\rangle}{dr} + 2\alpha_2\langle|a_2|^2\rangle = 2\,\mathrm{Re}(I_{12,12})(\langle|a_1|^2\rangle - \langle|a_2|^2\rangle), \quad (25)$$

$$\frac{d\langle a_1 a_2^*\rangle}{dr} + i(l_2^* - l_1)\langle a_1 a_2^*\rangle = -(I_{11,11} + I_{22,22} - 2I_{11,22} + 2I_{12,12})\langle a_1 a_2^*\rangle. \quad (26)$$

Here the transfer of energy between modes 1 and 2 is controlled by the familiar matrix element $2\,\mathrm{Re}(I_{12,12}) = f_{12}$. Numerical tests not shown here for lack of space reveal that Eqs. (24)–(26) are an excellent approximation to the full equations. Solution to equations like Eqs. (21) and (22) with constant coefficients are easily obtained by eigenvector analysis. Here we are primarily interested in the exponential decay rates given by the eigenvalues which are, $\lambda_\pm = -(f_{12} + \alpha_1 + \alpha_2) \pm \sqrt{f_{12}^2 + (\alpha_1 - \alpha_2)^2}$. In the absence of attenuation the eigenvalues are $2f_{12}$ and zero, so the approach to modal energy equipartition is dictated by $2f_{12}$. When small attenuation is added to the picture such that $f_{12} \gg \alpha_1, \alpha_2$, the eigenvalues are slightly modified giving $2f_{12} + \alpha_1 + \alpha_2$ and $\alpha_1 + \alpha_2$; that is to say, we have the coupling induced rate to equipartition superimposed on the slow overall attenuation decay of the mode amplitudes. This might be the case when the acoustic frequency is high or for deep-water propagation where α_n is small. If the attenuation is larger such that $f_{12} \ll \alpha_1, \alpha_2$ the approach to equipartition is dramatically changed. In this case the eigenvalues are $f_{12} + 2\alpha_1$ and $f_{12} + 2\alpha_2$, and the at-

tenuation dominates the exponential decay of the modes; this is the case in shallow water.

For the cross-mode coherence the solution of Eq. (26) is

$$\langle a_1 a_2^* \rangle(r) = \langle a_1 a_2^* \rangle(0) e^{i(l_1 - l_2^*)r} e^{-(I_{11,11} - 2I_{11,22} + I_{22,22})r} e^{-2I_{12,12}r}. \tag{27}$$

In the coherence solution attenuation only enters through the first exponential term and results in a slow decay of each of the initial mode amplitudes. The remaining terms are to be physically interpreted as phase randomization terms caused by adiabatic effects and coupling: the adiabatic terms are addressed first. In the adiabatic approximation the coupling matrices are diagonal and thus the cross-mode coherence can be easily written as

$$\langle a_n a_p^* \rangle(r) = \langle a_n a_p^* \rangle(0) e^{i(l_n - l_p^*)r} \exp\left(-\frac{1}{2} \left\langle \left(\int_0^r \rho_{nn}(r) dr - \int_0^r \rho_{pp}(r) dr \right)^2 \right\rangle \right), \tag{28}$$

where it has been assumed here that the coupling matrices are Gaussian random variables. The term in the exponent of Eq. (28) is one-half the adiabatic phase structure function: The phase structure function has a deep connection to coherence (Flatté *et al.*, 1979). In terms of the coherence matrices used in this paper the result is

$$\frac{1}{2} \left\langle \left(\int_0^r \rho_{nn}(r) dr - \int_0^r \rho_{pp}(r) dr \right)^2 \right\rangle = (I_{nn,nn} + I_{pp,pp} - 2I_{nn,pp}) r, \tag{29}$$

where it has been assumed that $r \gg r_c$. Hence the second exponential term in Eq. (27) is immediately recognized to be the adiabatic contribution resulting from the adiabatic phase structure function.

The last exponential term in Eq. (27) comes from mode coupling and has both real and imaginary parts. Here the coherence decay from mode coupling goes with the rate $2\,\mathrm{Re}(I_{12,12}) = f_{12}$, which is exactly half the rate driving the modes to energy equipartition. If adiabatic effects are weak, then the coherence decay is dominated by coupling and thus mode energy and cross-mode coherence decay at similar rates. In deep-water environments and for frequencies of order tens to hundreds of hertz it has been found that adiabatic effects are indeed weak and coupling is dominant (Dozier and Tappert, 1978a, 1978b; Colosi and Flatté, 1996). A multi-mode numerical example, demonstrating the relative decay rates of mode energy and coherence in a deep-water setting, will be presented in Sec. V. This example is significant since previous work has suggested that generally the coherence decays more rapidly than the modal energies [Dozier and Tappert (1978a, 1978b); Creamer (1996)]: Now this is understood to be true only in a coupling dominated regime. It is important to note that in shallow-water environments it has been shown by Monte Carlo simulation that cross-mode coherences do decay more rapidly than the mode energies (Creamer, 1996). The physical interpretation of this result [confirmed by theoretical evaluation of Eq. (16) for shallow water] is that shallow-water environments have a strong adiabatic component to mode propagation through random linear internal waves. A full treatment of shallow-water issues is beyond the scope of the present paper and will be addressed in subsequent work.

V. COMPARISONS TO MONTE CARLO SIMULATION

To exemplify the accuracy of the cross-mode coherence transport equation and to further illustrate some of the concepts previously put forth, comparisons between Monte Carlo simulations of mode propagation through random realizations of internal waves and results from the theory are presented. To not broaden the scope of this paper too much we only present a deep-water setting where attenuation is absent (see the Appendix for details). Following Morozov and Colosi (2005, 2007) and as demonstrated in Fig. 1, a 100-Hz directed acoustical beam is considered. For this case the initial mode excitation distribution is centered on mode 20 with a width of roughly ± 5; a total of $N=60$ modes is considered. Figure 2 shows a comparison between mode energy and cross -mode coherence results for the Monte Carlo simulation based on 500 realizations and predictions based on Eq. (17). In Fig. 2 for clarity of presentation only a few mode combinations are displayed in the neighborhood of the mode number 20, but the results are very similar for other combinations. It must be noted that in the lower panel of Fig. 2 results are presented for the normalized cross-mode coherence, that is, $|\langle a_n a_p^* \rangle|(r) / \sqrt{\langle |a_n|^2 \rangle(r) \langle |a_p|^2 \rangle(r)}$. The theory is seen to very accurately predict the range evolution of the cross-mode coherence matrix. It should be noted that the decay of the cross-mode coherences by roughly 1500-km range in this example is not inconsistent with basin-scale deep-water observations showing time resolved acoustic wavefronts up to 5000-km range (Worcester *et al.*, 1999).[4] These wavefronts are due to coherence of the modes across frequency, a problem that will be treated in future work.

The previous comparison is compelling but the most important result is for the real acoustic observable of mean intensity. Figure 3 shows calculations of mean intensity based on the Monte Carlo simulation, and the theory, as well as a calculation of the unperturbed intensity. Example receiver depths are 1.0 km, which is the sound channel axis and 2.2 km, which is in the deep shadow zone of the unperturbed beam. Cylindrical spreading effects are removed from Fig. 3. The Monte Carlo simulation and the theory prediction are seen to be quite different from the unperturbed intensity, demonstrating the well known effect that internal wave can cause a significant mean change in intensity as well as a fluctuation. In particular, for the receiver in the deep shadow zone a difference of 10 dB or more is evident at long range. Most importantly however is the accuracy of the theory compared to the Monte Carlo calculation; rms differences are less than a decibel, and perhaps primarily due to sampling uncertainty in the Monte Carlo calculation.

VI. SUMMARY AND CONCLUSIONS

We have presented some further analysis of Creamer's transport equation for the cross-mode coherence matrix

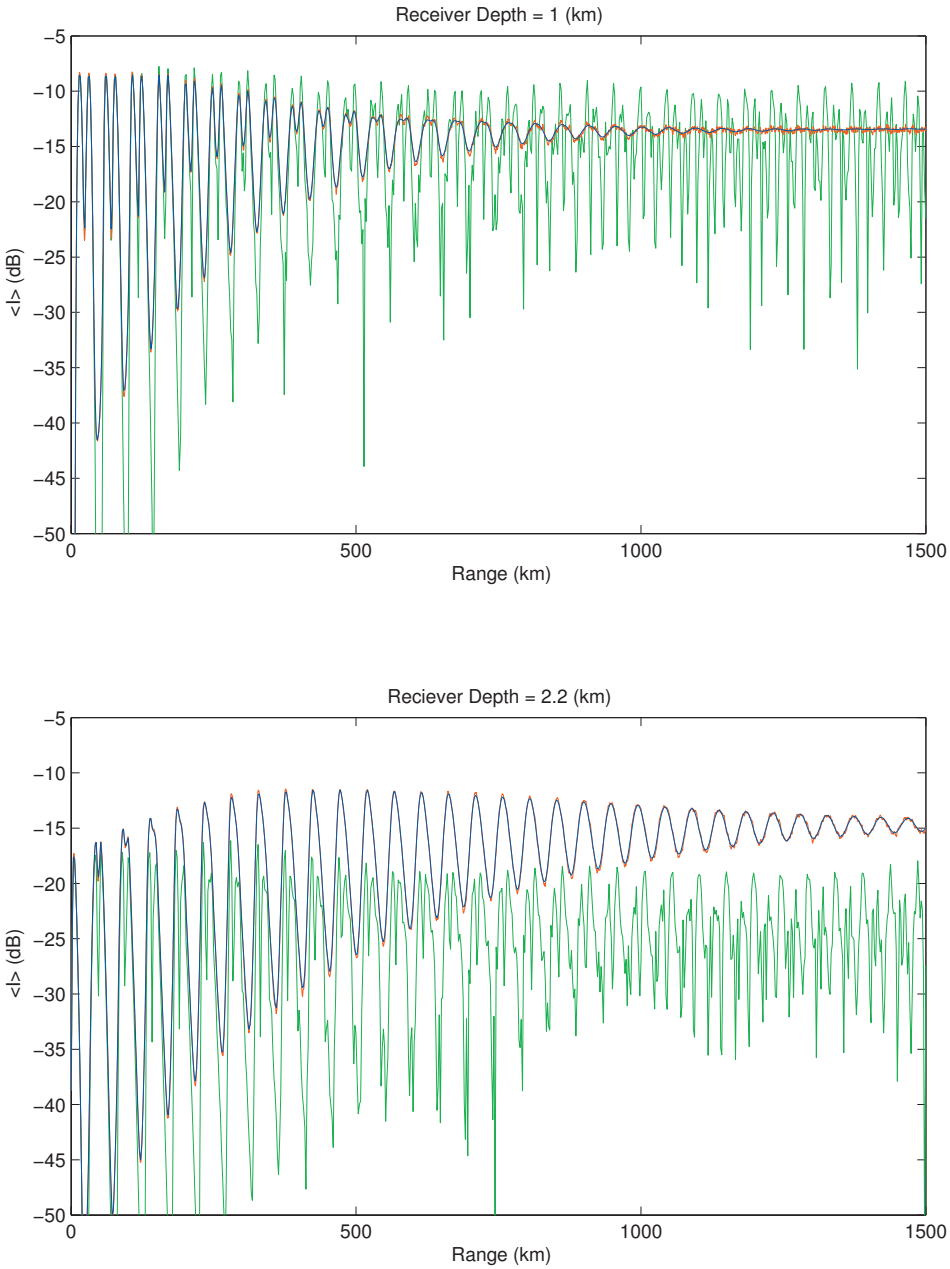


FIG. 3. Mean intensity from the Monte Carlo simulation (red), the model results from Eq. (17) (blue), and the unperturbed intensity (μ_0=0; green). In each calculation Eq. (2) is utilized but the cylindrical spreading factor is ignored. The upper/lower panels show cases for receiver depths at 1.0/2.2 km.

(Creamer, 1996). Three main results are found. First, the diagonal elements of the transport equations which dictate the range evolution of the mode energy are seen to be roughly independent of the off-diagonal elements of the cross-mode coherence matrix. Second, in the deep-water cases addressed here the off-diagonal elements of the cross coherence matrix are seen to evolve in range at similar rates to the diagonal elements or the mode energy. This result comes from a dominance of mode coupling phase randomization effects over those from adiabatic randomization. In shallow-water cases, not treated in detail in this paper but in others (Creamer, 1996) it is found that the opposite is true; cross-mode coherences decay much more rapidly than mode energy because the mode coupling rates are so slow and the adiabatic effect is so strong. Finally, it is shown that Creamer's (1996) transport equation can be used to accurately predict low frequency mean intensity for sound transmission through internal waves in a deep-water environment.

Challenges now are to understand the breakdown of the theory at higher frequency where the coupling gets stronger and to extend the theory to include cross-mode coherence between two different frequencies and thereby treat pulse propagation. The limitations of this 2D approach compared to the 3D treatment of Voronovich and Ostashev (2009) are also of fundamental interest. In addition the accuracy of the model in shallow-water environments where attenuation effects are important needs to be established. Finally comparisons to observations will provide the ultimate test of the utility of the theory and assumed ocean model as an acoustic prediction tool.

ACKNOWLEDGMENTS

The authors appreciate many useful discussions with Frank Henyey, Alex Voronovich, Vladimir Ostashev, and members of the North Pacific Acoustic Laboratory group.

This work was supported by the Office of Naval Research and the Naval Undersea Warfare Center's (NUWC) Under-Sea Warfare (USW) chair at the Naval Postgraduate School.

APPENDIX: MONTE CARLO SIMULATION AND MODEL CALCULATIONS

In this paper, example calculations are presented utilizing a simple deep-water environment, in which attenuation is neglected. A 2D sound-speed field of the form $c(r,z)=\bar{c}(z)+\delta c(r,z)$ is considered, where δc is the internal wave perturbation which is small compared to the mean $\bar{c}(z)$. The mean sound-speed profile is modeled using the Munk canonical form (Munk, 1974)

$$\bar{c}(z) = c_0[1 + \epsilon(e^{-2(z-z_a)/B} + 2(z-z_a)/B - 1)], \quad \text{(A1)}$$

with parameters $c_0=1500$ m/s, the sound channel axial depth $z_a=1000$ m, $B=1000$ m, and $\epsilon=0.005\,515$. The total water depth is chosen to be $D=4000$ m, and the background density is $\rho_0=1025$ kg/m^3. Internal-wave-induced sound-speed perturbations, δc, are modeled using the method of Colosi and Brown (1998); however, instead of using the Garrett–Munk spectrum the approximation in Eq. (8) is utilized. In these calculations the buoyancy frequency profile has an exponential form $N(z)=N_0e^{-z/B}$, where $B=1000$ m and $N_0=5$ cph. In our numerical calculations a maximum internal wave mode number of 100 is used, and internal waves with horizontal scales from 0.5 to 1600 km are simulated. Finally, the fractional sound-speed variance is modeled to be

$$\langle \mu^2(z) \rangle = \langle \mu_0^2 \rangle (N(z)/N_{\text{ref}})^3, \quad \text{(A2)}$$

where $\langle \mu_0^2 \rangle = 6.26 \times 10^{-8}$ and $N_{\text{ref}}=3$ cph. It should be noted, however, that the actual simulation profile of $\langle \mu^2(z) \rangle$ will be modified from Eq. (A2) near the ocean surface and bottom since the internal-wave vertical modes have a zero displacement boundary condition (see Flatte and Colosi, 2008).

Monte Carlo numerical simulations were carried out with the aforementioned environmental parameters. A 100-Hz acoustical beam is considered [Morozov and Colosi (2005, 2007)], where the initial condition is

$$a_n(0) = N_a^{-1}\sqrt{k_0/\delta}\int_0^D (\phi_n(z)/\sqrt{\rho_0(z)})e^{-k_0^2(z-z_s)^2/\delta^2}dz, \quad \text{(A3)}$$

with $\delta=40$, $z_s=2000$ m, and N_a is a normalization factor such that $\Sigma_{n=1}^N \langle |a_n(0)|^2 \rangle = 1$. The maximum mode number N is 60, and the mode number with the largest initial energy is mode 20. The coupled mode equations are solved for random realizations of the internal-wave-induced sound-speed perturbations by transforming Eq. (3) to be expressed in terms of the mode amplitude a_n instead of ψ_n; the coupled equations are then solved using eigenvector techniques, as described in Dozier and Tappert (1978a, 1978b) and Creamer (1996). A total of 500 realizations of $a_n(r)$ was computed to obtain ensemble averages of the cross-mode coherence matrix and the mean intensity.

[1]Equation (16) is equivalent to Eq. (17) in Creamer (1996).

[2]Importantly if attenuation is neglected the resulting symmetry of the matrices means that Eq. (17) conserves energy, that, is $d/dr\Sigma_{n=1}^N \langle |a_n|^2 \rangle (r) = 0$.

[3]Ignoring the exponentials is a small attenuation, high frequency approximation. In the absence of attenuation there is a coupling resonance condition that selects only the internal-wave wavenumber that matches the beat wavenumber k_{mn}. Physically, attenuation broadens the resonance condition so that more internal-wave wavenumbers contribute to the coupling (Colosi, 2008).

[4]The initial condition in this example is clearly different from the point source initial condition in the observations.

Beron-Vera, F. J., Brown, M. G., Colosi, J. A., Virovlyansky, A. L., Zaslavsky, G. M., Tomsovic, S., and Wolfson, M. A. (**2003**). "Ray dynamics in a long range acoustic propagation experiment," J. Acoust. Soc. Am. **114**, 1226–1242.

Colosi, J. A. (**2008**). "Acoustic mode coupling induced by shallow water nonlinear internal waves: Sensitivity to environmental conditions and space-time scales of internal waves," J. Acoust. Soc. Am. **124**, 1452–1464.

Colosi, J. A., and Brown, M. G. (**1998**). "Efficient numerical simulation of stochastic internal wave induced sound speed perturbation fields," J. Acoust. Soc. Am. **103**, 2232–2235.

Colosi, J. A., and Flatté, S. M. (**1996**). "Mode coupling by internal waves for multimegameter acoustic propagation in the ocean," J. Acoust. Soc. Am. **100**, 3607–3620.

Colosi, J. A., Flatté, S. M., and Bracher, C. (**1994**). "Internal wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison to experiment," J. Acoust. Soc. Am. **96**, 452–468.

Colosi, J. A., Scheer, E. K., Flatte, S. M., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999**). "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218.

Creamer, D. (**1996**). "Scintillating shallow water waveguides," J. Acoust. Soc. Am. **99**, 2825–2838.

Dozier, L. B. (**1983**). "A coupled mode model for spatial coherence of bottom-interacting energy," in *Proceedings of the Stochastic Modeling Workshop*, edited by C. W. Spofford and J. M. Haynes (ARL-University of Texas, Austin TX).

Dozier, L. B., and Tappert, F. D. (**1978a**). "Statistics of normal-mode amplitudes in a random ocean. I. Theory," J. Acoust. Soc. Am. **63**, 353–365.

Dozier, L. B., and Tappert, F. D. (**1978b**). "Statistics of normal-mode amplitudes in a random ocean. II. Computations," J. Acoust. Soc. Am. **64**, 533–547.

Dushaw, B. D., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1999**). "Multi-megameter range acoustic data obtained by bottom mounted hydrophone arrays for measurement of ocean temperature," IEEE J. Ocean. Eng. **24**, 203–215.

Flatté, S. M., and Colosi, J. A. (**2008**). "Anisotropy of the wavefront distortion for acoustic pulse propagation through ocean sound speed fluctuations: A ray perspective," IEEE J. Ocean. Eng. **33**, 477–488.

Flatté, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, Cambridge, UK).

Fredricks, A., Colosi, J. A., Lynch, J. F., Gawarkeiwicz, G., Chiu, C. S., and Abbot, P. (**2005**). "Analysis of multipath scintillations observed during the summer 1996 New England shelfbreak PRIMER study," J. Acoust. Soc. Am. **117**, 1038–1057.

Henyey, F., and Ewart, T. E. (**2006**). "Validity of the Markov approximation in ocean acoustics," J. Acoust. Soc. Am. **119**, 220–231.

Morozov, A. K., and Colosi, J. A. (**2005**). "Entropy and scintillation analysis for acoustical beam propagation through ocean internal waves," J. Acoust. Soc. Am. **117**, 1611–1623.

Morozov, A. K., and Colosi, J. A. (**2007**). "Stochastic differential equation analysis for sound scattering by random internal waves in the ocean," Acoust. Phys. **53**, 335–347.

Munk, W. H. (**1974**). "Sound channel in an exponentially stratified ocean, with application to SFOAR," J. Acoust. Soc. Am. **55**, 220–226.

Sakurai, J. J. (**1985**). *Modern Quantum Mechanics* (Addison-Westley, Redwood City, CA).

Tielburger, D., Finette, S., and Wolf, S. (**1997**). "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 789–808.

Van Kampen, N. G. (**1981**). *Stochastic Processes in Physics and Chemistry* (North-Holland, New York).
Van Uffelen, L. J., Worcester, P. F., Dzieciuch, M. A., and Rudnick, D. L. (**2009**). "The vertical structure of shadow-zone arrivals at long range in the ocean," J. Acoust. Soc. Am. **125**, 3569–3588.
Virovlyanskii, A. L. (**1989**). "Correlation of modes in a waveguide with large scale random inhomogeneities," Radiophys. Quantum Electron. **32**, 619–624.
Virovlyanskii, A. L., Kosternin, A. G., and Malakhov, A. N. (**1989**). "Mode fluctuations in a canonical underwater channel," Sov. Phys. Acoust. **35**, 138–142.
Voronovich, A. G., and Ostashev, V. E. (**2006**). "Low frequency sound scattering by internal waves in the ocean," J. Acoust. Soc. Am. **119**, 1406–1419.
Voronovich, A. G., and Ostashev, V. E. (**2009**). "Coherence function of a sound field in an oceanic waveguide with horizontally isotropic statistics," J. Acoust. Soc. Am. **125**, 99–110.
Wage, K. E., Dzieciuch, M. A., Worcester, P. F., Howe, B. M., and Mercer, J. A., (**2005**). "Mode coherence at megameter ranges in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1565–1581.
Worcester, P. F., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Colosi, J. A., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T., and Baggeroer, A. (**1999**). "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201.
Worcester, P. F., Cornuelle, B. D., Hildebrand, J., Hodgekiss, W., Duda, T., Boyd, J., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1994**). "A comparison between measured and predicted broadband acoustic arrival patterns in travel-time depth coordinates at 1000-km range," J. Acoust. Soc. Am. **95**, 3118–3128.

A normal mode projection technique for array response synthesis in range-dependent environments

Kevin D. Heaney[a)]
Ocean Acoustical Services and Instrumentation Systems, Inc., 11006 Clara Barton Drive, Fairfax Station, Virginia 22039

(Received 10 December 2008; revised 25 June 2009; accepted 26 June 2009)

A method is presented to examine the problem of simulating the beam time series response for an array of arbitrary shape in a range-dependent environment using a combination of parabolic equation (PE) forward modeling and local normal mode analysis. The procedure involves computing the acoustic pressure field as a function of depth, using the PE, at the nominal center location of the array. The field is then decomposed into local complex normal mode amplitudes. These mode amplitudes are used to compute the field response on each array element, via range-independent normal mode theory. Conventional plane-wave beamforming is then applied. It is shown that a single matrix computation can be used to map the field as a function of depth to the beam response as a function of angle. The method is applied to two broadband range-independent examples to demonstrate its accuracy. It is then applied to a shallow-water range-dependent experiment from off the Florida coast and a deep-water range-dependent experiment from sound scattering off a seamount in the open ocean. For both range-dependent examples, the model simulation results reproduce the qualitative features observed in the data.
 [DOI: 10.1121/1.3180698]

PACS number(s): 43.30.Dr, 43.30.Bp, 43.30.Zk [EJS] Pages: 1036–1045

I. INTRODUCTION

The development of high fidelity ocean acoustic models, such as the parabolic equation[1] (PE), normal mode, and ray models, has led to the use of these models as tools to interpret underwater acoustic measurements. In terms of output pressure fields, the PE is ideally suited to vertical line array (VLA) data synthesis due to the computational approach of forward marching the vertical pressure field with range. Due to the ease of deployment and the ubiquity of naval towed array receivers, many experiments utilize acoustic arrays with horizontal aperture. For cases where the array is oriented at endfire to the source, array signal synthesis with the PE is trivial. For volumetric or horizontal line arrays (HLAs) which are not oriented at endfire relative to the source, however, the synthesis of the array response is less straightforward. Adiabatic normal mode solutions work well for arrays of arbitrary geometry, but the adiabatic approximation is not valid in many ocean environments. In this paper an efficient matrix multiplication projection algorithm is presented which transforms the field from a densely sampled vertical slice, such as that output by the PE, to the beam response of an array of arbitrary geometry centered at the vertical slice location. The two primary approximations are that the environment is range-independent over the horizontal extent of the array and that the environment from the source to each hydrophone is identical. This approach is commonly used, but the author is not aware of its appearance in the literature.

For experiments which combine broadband transmissions and horizontal aperture, structure of the received signals can be resolved in arrival angle and arrival time. The addition of spatial and temporal resolutions is very useful for ocean acoustic tomography and geo-acoustic inversions. The efficient synthesis of beam-time series is the emphasis of this paper. The basic numerical approach, whose theory is outlined in Sec. II, is to use the PE to generate the single frequency field on a densely sampled vertical slice. This field is then decomposed into local normal mode amplitudes. The modes are then propagated, via normal mode theory, to each element of the horizontal (or volumetric) array. The final step of plane-wave beamforming is then applied. These operations can be combined into a single matrix transformation to go from vertical pressure field to complex beam output. The inverse fast Fourier transform (FFT) of the complex beam output yields the desired beam-time series.

The outline of the paper is as follows. The mode extrapolation method is presented in Sec. II. In Sec. III the technique is applied to range-independent environments in shallow and deep water. Model-data comparison for strongly range-dependent environments for shallow and deep water are presented in Sec. IV. Section V is the summary and conclusion.

II. MODE EXTRAPOLATION METHOD

A. Theoretical derivation

The goal of this paper is to present and interpret a technique for mapping the computed pressure field at a vertical slice, as is outputted by the range-dependent PE model, to the field received on an arbitrary array, which is subsequently conventionally beamformed using plane-wave steering vectors. A standard solution of the acoustic wave equation in a range-independent environment is to apply the technique of separation of variables and compose the field in terms of a

[a)]Electronic mail: oceansound04@yahoo.com

summation of orthogonal acoustic normal modes in the vertical.[2] The far-field solution of the acoustic pressure field (P) to the wave equation, referred to as the normal mode theory solution, follows the form

$$P(\omega,r,z,z_s) \approx \sqrt{\frac{2}{\pi r}} e^{-j\pi/4} \sum_n \frac{\varphi_n(\omega,z)\varphi_n(\omega,z_s)}{\sqrt{k_n(\omega)}} e^{ik_n(\omega)r}, \tag{1}$$

where the source frequency and depth are ω and z_s, the receiver range and depth are r and z, and the mode function and horizontal wavenumber for mode n are $\varphi_n(z)$ and k_n, respectively. The far-field approximation has been applied. For mildly range-dependent environments (where mode coupling is small), the adiabatic mode approximation is obtained by applying the WKB approximation via using the mode functions computed at the source location (for z_s) and receiver location (z) as well as using the range-averaged wavenumber in Eq. (1).

The normal mode functions, which are solutions to the depth-separated wave equation (a Sturm–Louville equation), are orthonormal and complete. The orthonormality condition of the acoustic normal modes, assuming infinite half-space boundaries at 0 and H, can be expressed as

$$\int_0^H \frac{\varphi_n(z)\varphi_m(z)}{\rho(z)} dz = \delta_{mn}, \tag{2}$$

where $\rho(z)$ is the fluid density as a function of depth. The completeness property ensures that any arbitrary acoustic field can be written as the complex superposition of normal modes. Combining the completeness and the orthonormality relations [Eq. (2)] leads to a method for decomposing an arbitrary pressure field $P(\omega,r,z)$ at range r, frequency ω, and depth z, into the corresponding mode amplitude for the mth mode $b_m(\omega,r)$

$$b_m(\omega,r) = \int \frac{\varphi_m(\omega,z)P(\omega,r,z)}{\rho(z)} dz. \tag{3}$$

If propagation is range-independent, the normal mode solution for the vertical field given in Eq. (1) can be substituted into Eq. (3) to compute the complex mode amplitude. Rearranging the order of depth integrals and mode sums, as well as applying the orthonormality relation, [Eq. (2)] yields

$$\begin{aligned} b_m(\omega,r) &= \sqrt{\frac{2}{\pi r}} e^{i\pi/4} \int \varphi_m(z) \sum_n \frac{\varphi_n(\omega,z)\varphi_n(\omega,z_s)}{\rho(z)\sqrt{k_n(\omega)}} e^{ik_n(\omega)r} \cdot dz \\ &= \sqrt{\frac{2}{\pi r}} e^{i\pi/4} \sum_n \frac{\varphi_n(\omega,z_s)}{\sqrt{k_n(\omega)}} e^{ik_n(\omega)r} \int dz \varphi_m(z) \frac{\varphi_n(\omega,z)}{\rho(z)} \end{aligned}$$

$$b_m(\omega,r) = \sqrt{\frac{2}{\pi}} e^{i\pi/4} \frac{\varphi_m(\omega,z_s)}{\sqrt{k_m(\omega)r}} e^{ik_m(\omega)r-\pi/4}, \tag{4}$$

where the orthonormality relation [Eq. (2)] has been used and n is a modal index in the sum and z is depth. The authors see from this that for the normal mode solution the complex mode amplitudes are simply the combination of mode amplitude at the source, a cylindrical spreading term $(1/\sqrt{r})$, and the phase progression of the wave with range ($e^{ik_m r}$). For a fully sampled vertical field, the decomposition of the field into complex mode amplitudes [Eq. (3)] can be performed as a simple matrix computation:

$$b_m = \int \frac{\varphi_m(z)}{\rho(z)} P(z) dz = U\vec{P}, \tag{5}$$

where the mode function matrix U (with dimensions number of modes × number of depths) is defined by

$$U_{m,iz} = \frac{\varphi_m(z_{iz})}{\rho(z_{iz})} dz. \tag{6}$$

The authors now seek to map the field to an arbitrary array at a relative azimuthal angle of θ_s from the source. For illustrative purposes, the authors will consider a uniformly spaced linear horizontal array. This technique does not require such an array, but it provides a simplification of the mathematical interpretation and is the geometry of the arrays used in the model-data comparisons below. For a uniformly spaced (separation Δx) array at a uniform depth, with an element at the origin, the vector position of each element is defined as

$$\vec{r} = (n\Delta x \cos\theta_s, 0, z_r), \tag{7}$$

where n is the element index.

The authors make three assumptions before continuing. The first is that the environment is range-independent along the length of the array. The other two assumptions are that the receiver is in the acoustic far field of the source and that the source is in the far field of the array beampattern. By assuming a range-independent environment over the length of the array and that the source is in the far field, the authors can use normal mode theory to project the field from the origin to each element. By neglecting wavefront curvature, they can represent the incremental range increment for each element by Eq. (7), rather than using the full expression $\vec{r}\cdot\vec{x}$. This approximation simplifies the math, but is not required. For the computed vector of complex mode amplitudes b_m, the pressure field for element n, $\tilde{P}(r_n)$ is then

$$\tilde{P}(r_n) = \sum_m b_m \varphi_m(z_n) e^{ik_n n\Delta x \cos\theta_s} \sqrt{\frac{r}{r+\Delta r_n}},$$

$$\tilde{P}_{\text{array}} = \Phi\vec{b} = \Phi U \vec{P}_{\text{PE}}, \tag{8}$$

where the square root term scales the field, via cylindrical spreading, from range r to the new range $r+\Delta r_n$. The propagation matrix Φ (with dimension number of phones × number of modes) is defined by

$$\Phi_{n,m} = \varphi_m(z_n) e^{jk_m n\Delta x \cos\theta_s} \sqrt{\frac{r}{r+\Delta r_n}}. \tag{9}$$

The authors now have a technique to transform from a vertical field to the field on an arbitrary array. The final step is to apply plane-wave beamforming on the received field where the authors define the plane-wave horizontal steering vector, w, for the nth element as

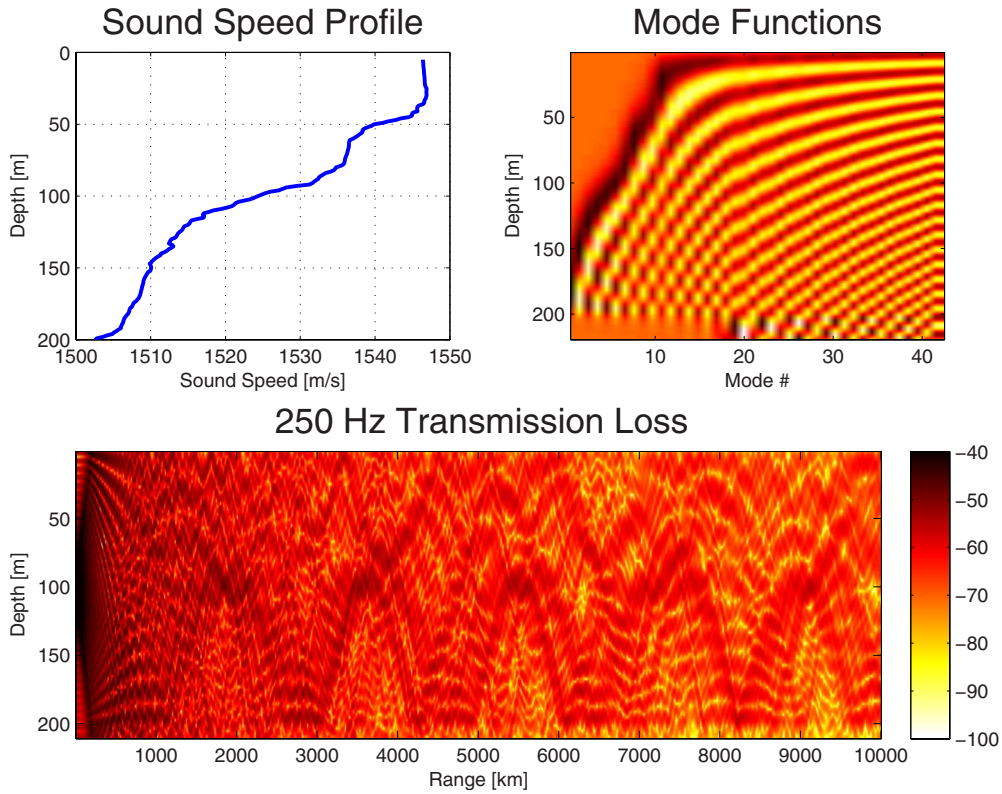


FIG. 1. (Color online) Range-independent shallow-water example (Upper left) Measured CALOPS08 sound speed profile. (Upper Right) Normal mode functions at 250 Hz. (Lower) TL (dB) for a 100 m source.

$$w_n(\phi) = \exp(jkn\Delta x \cos \phi)/\sqrt{NP}, \tag{10}$$

where ϕ is the beam steering angle, NP is the number of elements, k is the reference wave speed (ω/c_0), and w is normalized such that $w'w=1$. Writing the beamformer as a matrix product produces

$$B(\phi) = W\tilde{P}_{\text{array}}, \tag{11}$$

where the beamforming matrix W (dimensions number of steering directions by number of elements) is defined by

$$W_{iphi,n} = \frac{1}{\sqrt{NP}} \exp\left\{ i\frac{\omega}{c_0} n\Delta x \cos \phi_{iphi} \right\}, \tag{12}$$

where *iphi* is the beam index, and n is the element number. Combining the steps of mode decomposition [Eq. (5)], mode projection [Eq. (8)], and beamforming [Eq. (12)] yields a single matrix computation, $W\Phi U$, that transforms the narrowband field from pressure/depth space to beam response:

$$\vec{B}(\omega,\phi) = W\Phi U\vec{P}_{\text{PE}}(\omega,r,z). \tag{13}$$

This transformation matrix depends entirely on the local environment and array geometry and is independent of the field P_{PE}. To generate beam-time series the entire computation is made for each frequency within the source band and the field is inverse Fourier transformed.

B. Interpretation

In order to provide some physical insight into the projection process, two pairs of matrix operations are examined. The first pair is the matrix pair ΦU, corresponding to the sequential combination of the mode projection and mode propagation computations. Using Eqs. (4) and (6) to write out the matrix product explicitly yields

$$G = \Phi U,$$

$$G_{n,iz} = \sum_m \Phi_{n,m} U_{m,iz}$$

$$= \sum_m \varphi_m(z_n) e^{ik_m n\Delta x \cos \theta_s} \sqrt{\frac{r}{r+\Delta r_n}} \frac{\varphi_m(z_{iz})}{\rho(z_{iz})} dz,$$

$$G(z_{iz},z_n,\Delta r) \approx \sum_m \frac{\varphi_m(z_{iz})\varphi_m(z_n)}{\rho(z_{iz})} e^{ik_m\Delta r_n} \sqrt{\frac{r}{r+\Delta r_n}} dz. \tag{14}$$

The authors see that the ΦU matrix operation is effectively the far-field Green's function for a line segment source (length dz) at a source depth z_{iz} on the vertical slice propagating to a horizontal element position at z_n with a range offset of Δr_n.

The second pair of matrix operations $W\Phi$ can be interpreted by explicitly combining Eqs. (6) and (12):

$$H = W\Phi,$$

$$H_{iphi,m} = W_{iphi,n}\Phi_{n,m}$$

$$= \sum_n \frac{1}{\sqrt{NP}} \exp\{-ik_0 n\Delta x \cos \phi_{iphi}\} \varphi_m(z_n)$$

$$\times \exp\{ik_m n\Delta x \cos \theta_s\}, \tag{15}$$

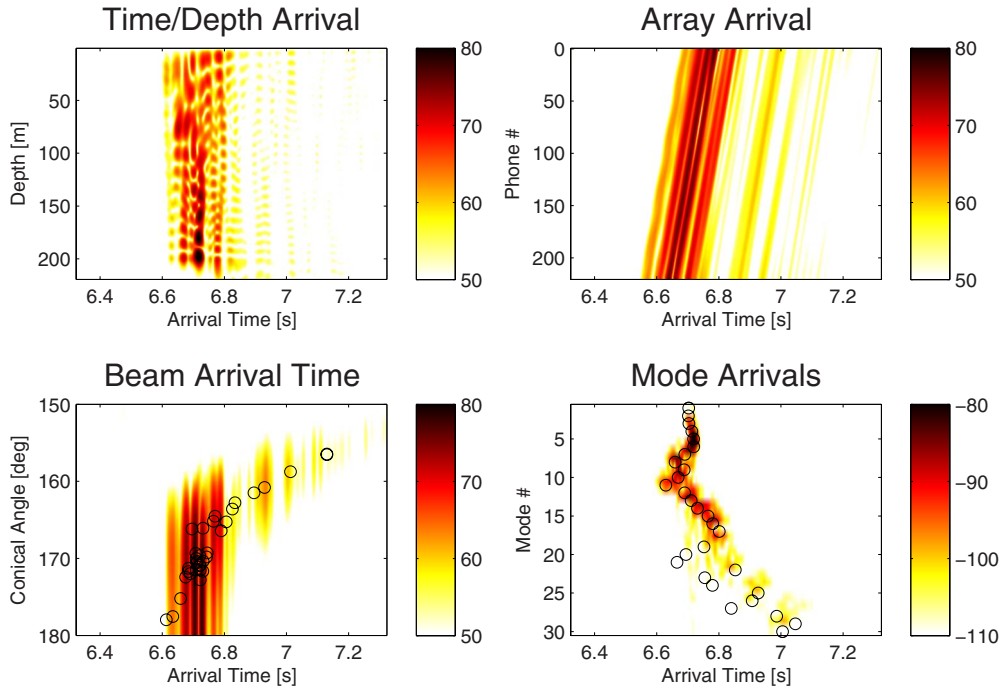


FIG. 2. (Color online) Broadband response for RI-shallow-water case (color online). All colormaps have a 30 dB range from the peak. (Upper left) VLA response from PE at a range of 10 km. (Upper right) HLA response vs element for mode-projection (endfire aspect). (Lower left) Beam-arrival time series from beamformed HLA mode-projection data and the ray predictions (○). (Lower right) Mode arrival time series and 250 Hz group velocity mode arrivals (○).

where *iphi* is the beam index, n is the array element index, and m is the normal mode index. Re-arranging terms and taking the mode function out of the sum yields

$$H_{iphi,m} = W_{iphi,n}\Phi_{n,m} = \frac{\varphi_m(z_n)}{\sqrt{NP}}\sum_n \exp\{in\Delta x(k_m \cos\theta_s - ik_0 \cos\phi_{iphi})\}. \tag{16}$$

If the array is well sampled, the sum over elements, n, can be replaced by an integral. In the case of a long array ($r\to\infty$) yields

$$H_m(\phi) = \frac{\varphi_m(z)}{\sqrt{NP}}\int_0^\infty dr \exp\{ir(k_m \cos\theta_s - k_0 \cos\phi)\},$$

$$H_m(\phi) = \frac{\varphi_m(z)}{\sqrt{NP}}\delta\left(\phi - \cos^{-1}\frac{k_m}{k_0}\cos\theta_s\right). \tag{17}$$

This matrix operator is simply a horizontal wavenumber filter, equivalent to applying a Hankel transform to a pressure field consisting of only one mode. The delta function result in Eq. (17) acts as a resonance condition, selecting only modes with horizontal wavenumber projections matching the beam steering direction. For a source location directly broadside to the array, all the modes arrive on the same broadside beam. For an endfire source (θ_s=0), the authors see that each mode arrives at the cosine of the ratio of the steering wavenumber to the mode wavenumber (higher modes coming in at higher angles as expected).

III. RANGE INDEPENDENT EXAMPLES

The mode projection technique, developed in Sec. II, is now applied to two range-independent environments, one in shallow water and one in deep water. The range-independent numerical examples are presented to provide both a validation of the numerical approach and to facilitate in the interpretation. The emphasis is on beamforming broadband signals. For each example, the source transmits a broadband signal from 200 to 300 Hz.

A. Range-independent shallow-water case

The shallow-water example is taken from an experiment conducted in September off the east coast of Florida.[3] The sound speed profile, measured with a conductivity-temperature-density profiler at the array site, is shown in the upper left panel of Fig. 1. The acoustic environment consists of a 20 m sediment of sand (compressional speed 1560 m/s, density 1.5 gm/cc, and attenuation 0.3 dB km/kHz) overlying a hard basement (compressional speed 1800 m/s, density 1.9 gm/cc, and attenuation 0.05 dB km/kHz). These parameters were considered relevant to the area where a sand layer overlies a limestone basement. The water depth is taken to be 200 m. The source range is 10 km with the source and receiver both at 100 m depth. The normal mode functions at 250 Hz (center of the band) and the transmission loss (TL) using the PE are shown in Fig. 1.

From the sound speed profile, which is typical of warm temperate climates in the summer, the authors see that sound is strongly refracted downwards. There is a surface mixed layer, which will generate trapped energy (rays or modes) but this energy will not be excited by a source at 100 m. From

the mode functions, the authors see that a 100 m source will excite modes above mode 4 at 250 Hz. Modes 1–10 do not interact with the surface and are referred to as RBR modes. Beyond mode 11 all modes are surface and bottom interacting and are labeled surface reflecting, bottom reflecting (SRBR) modes. These mode groups $(1-10, >11)$ will have different dispersion characteristics.

The mode extrapolation technique is applied to the PE field at 10 km. For each frequency, the complex mode amplitudes are generated and the field is marched forward to the positions of a HLA (endfire) to the source at a depth of 100 m. The field is then beamformed and Fourier transformed to generate beam-time series. The horizontal array is 96 elements with a 3 m spacing (half-wavelength at 300 Hz). For this case (endfire array) it is trivial to generate the field on the array by saving the PE field output every range step corresponding to an element position. The broadband impulse response is computed for the vertical field (from the PE), the individual mode arrivals, the phone arrivals on the HLA (after mode projection), and the beam-time series after narrowband beamforming. The results are shown below in Fig. 2.

The upper left panel of Fig. 2 shows the impulse response as would be observed on a VLA. The strongest arrival is for the lower order modes at 6.7 s. The earlier arrival of higher order modes (with turning depths around 75 m) is visible as is the later arrival of SRBR rays/modes. The phone arrival time series as a function of element number and time, computed by mode extrapolation (decomposition and propagation) to the HLA at a depth of 100 m, is shown in the upper right panel. The beam-time series is shown in the lower left, with a ray-trace computation (angle/time-delay) plot overlaid. It is clear that the HLA beam response accurately reflects the arrival of the refracted and surface reflected paths. The lower right panel shows the Fourier transform of the mode decomposition results, which is the mode arrival time series (as sampled by the dense VLA). Overlaid on this are the modal arrival times from the computed mode slowness $(1/vg)$ at 250 Hz. The agreement is excellent (as it should be). The modes that do not agree well are numerically unstable (in the mode computation) and have very little energy. The mode arrival time plot highlights that the source excites modes 4–7 with the most energy. Modes 7–11 are surface refracted and each higher mode arrives earlier because it samples faster water at its WKB turning depth. Beyond mode 11, surface interaction occurs and then the higher modes arrive later, consistent with fact that the sound samples the same sound speed field in the water column, but travels at a higher angle, reducing its group velocity. This example demonstrates that the mode projection algorithm has successfully computed the beam response on an endfire HLA by computing the field densely sampled in the vertical using the PE.

B. Range-independent deep-water case

For the range-independent deep-water case, the authors consider a 50 km transmission in 5000 m of water in the central North Pacific. This is the site of the Basin Acoustic Seamount Scattering Experiment-2004 (BASSEX04) test[4] where sound from an axial broadband source was scattered off the Kermit–Roosevelt seamount system. The sound speed profile, taken in September 2004, is shown in the upper left panel of Fig. 3. For this typical central pacific profile, there is a deep sound channel with a sound speed minimum near 900 m. The source is placed at the axis and the authors seek to simulate the beam response on a HLA deployed at 300 m. Figure 3 shows the sound speed, mode group velocities, and the TL computed using the PE.

From the profile and mode functions, it can be seen that there is a surface thermocline that extends sharply down to 200 m. Below this there is a change in the sound speed gradient as the depth extends to the sound channel axis. From the group velocity function plot, it can be seen that the lower modes (1–70) are trapped in the primary duct, modes 70–100 turn in the thermocline, and modes 100–350 are surface reflecting, bottom refracting (SRR) paths. Beyond mode 350, paths are SRBR. The TL plot indicates several caustics, which show hints of the convergence zone at 50 km.

The mode decomposition, mode propagation, and beamforming algorithms were applied to the vertical PE field at 50 km, just as for the shallow-water case. The PE field, normal modes, and beam-projection matrices of Eq. (13) are computed at each frequency and then Fourier transformed to generate time domain responses. The HLA contains 96 elements, has a length of 200 m, and is deployed at a depth of 300 m. The time series response of the field on a VLA, HLA, beam, and mode arrivals are computed and shown in Fig. 4. For this simulation the source bandwidth was 200–300 Hz, as above. A time-window of 8 s was used for the FFT (df =0.125 Hz) due to the dispersion of the signal.

The upper left panel of Fig. 4 shows the impulse response in time-depth space. Although computed using a PE, a wavefront interpretation using ray-theory is most intuitive in interpreting this range-independent deep-water case. The axial rays that have not completed a full cycle arrive around 33.9 s. The earlier arriving wavefronts correspond to higher angle fully refracted rays that have gone through one turning point. Each doublet, or pair of arrivals, corresponds to a surface reflection pair, where one ray arrives from the surface and the other arrives from below. The separation in arrival time is exactly the array-surface two way travel time. The pair of doublets arriving at 36 and 38 s is caused by the time delay between upward going and downward going rays at the source. Both sets of arrivals have interacted with the seafloor. The weak arriving doublet just after 34 s is interpreted to be a sub-basement reflection (recall that there is 20 m of soft sediment overlying a hard basement). The upper right panel shows the phone time series as seen across the endfire HLA. The weak arrivals coming in beyond 38 s are second order bottom bounces. The beam-arrival time series is plotted in the lower-leftpanel of Fig. 4. The array is oriented with the source at endfire. The low angle energy arrives first with the subsequent higher angle bottom bounces arriving later. For this plot, the ray-theory arrivals are plotted on top of the beam responses. The agreement between the ray theory predictions and the beam-time series is excellent, although the sub-basement reflections are not evident in the ray predic-

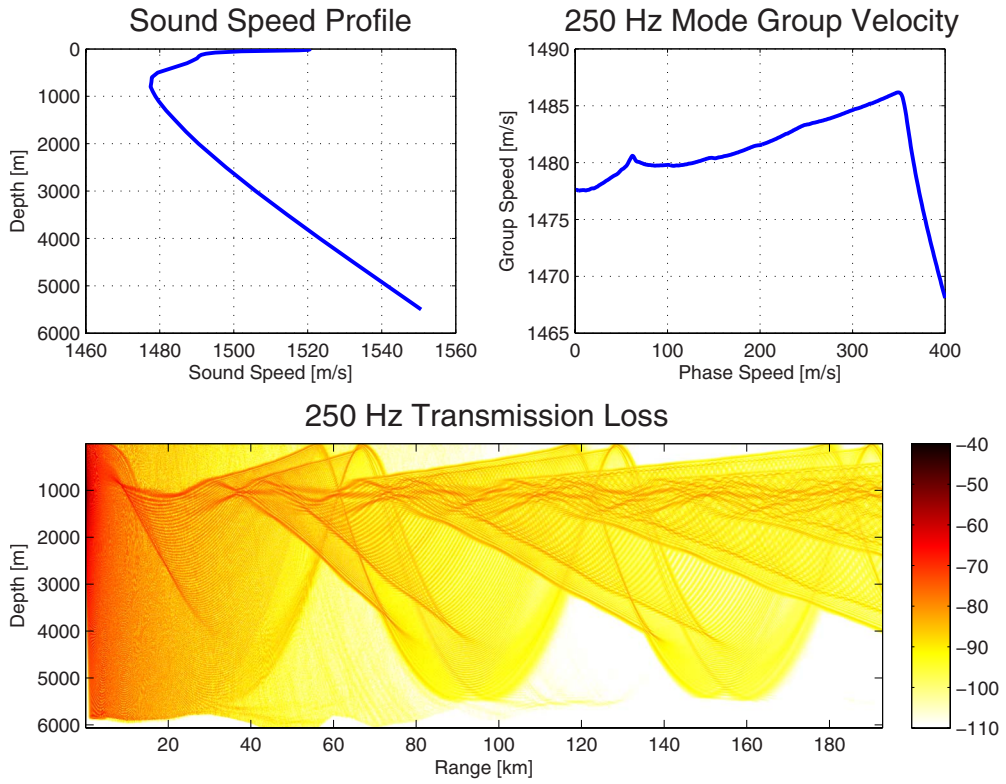


FIG. 3. (Color online) Range-independent deep-water example. (Upper left) Measured BASSEX04 sound speed profile. (Upper Right) Normal mode group velocities at 250 Hz. (Lower) TL (dB) for an 800 m source.

tion. The authors do not expect these arrivals because sub-bottom interacting energy is not modeled (except for a phase delay) in the ray-theory prediction. The mode arrival times from the 250 Hz group velocities are overlaid on the full mode arrival time computation in Fig. 4. The 250 Hz mode arrival times, computed from the normal mode group veloci-

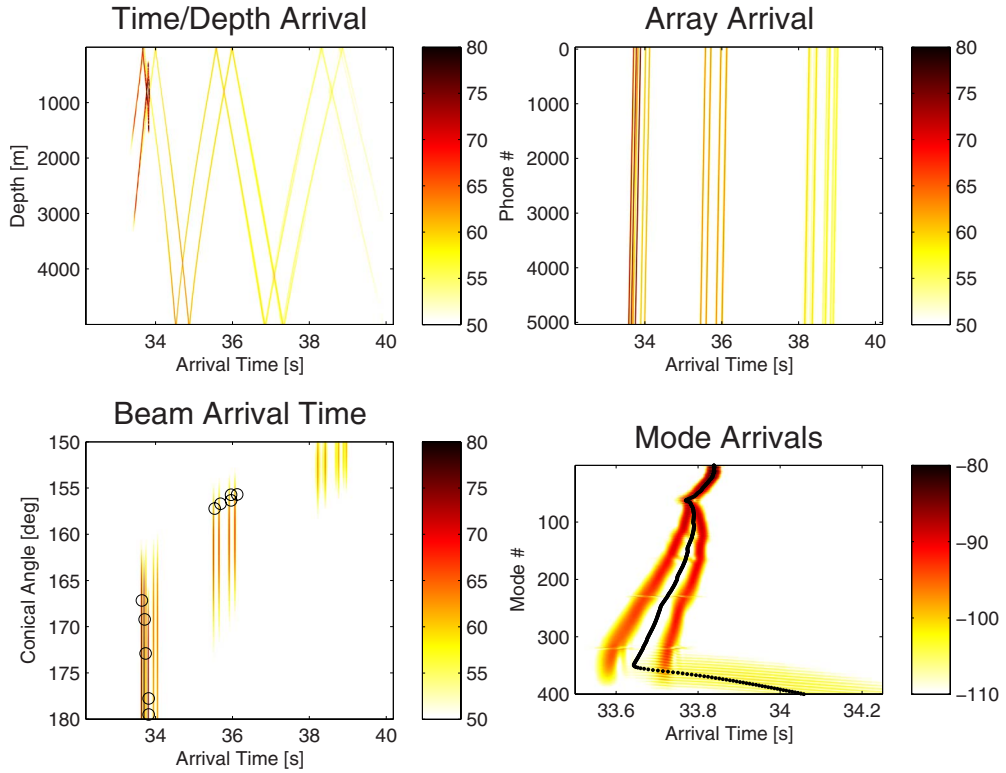


FIG. 4. (Color online) Broadband response for RI-deep-water case (color online). (Upper left) VLA response from PE at a range of 50 km. (Upper right) HLA response vs element for mode-projection (endfire aspect). (Lower left) Beam-arrival time series from beamformed HLA mode-projection data and the ray predictions (○). (Lower right) Mode arrival time series and 250 Hz group velocity mode arrivals (○).

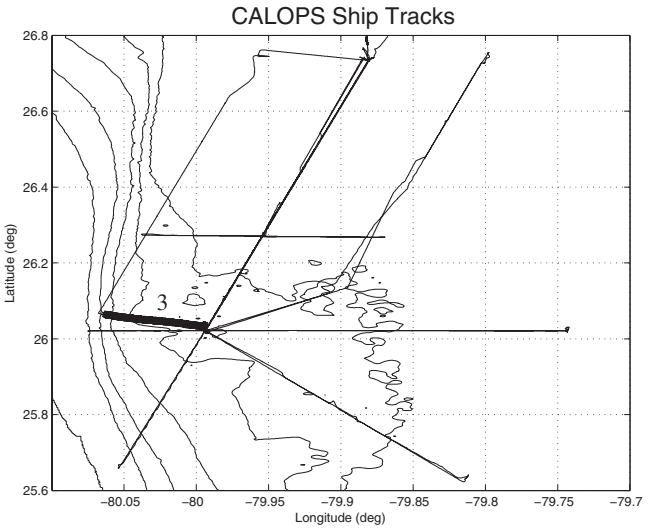


FIG. 5. Bathymetry and source tracks for CALOPS 2007 experiment off Florida. The range-dependent path considered in this paper is path upslope north-west of the receiver (located at the center). Contours are at 50 m with the receiver in the center at 250 m.

ties, are overlaid on the mode arrival time series. The authors see from the mode dispersion that modes 1∼70 are fully refracting, modes 70–370 are SRR and modes beyond 370 are SRBR.

IV. MODEL-DATA COMPARISONS

A. Range-dependent shallow-water case (CALOPS-07)

In the fall of 2007 an experiment (Summer-CALOPS-07) was conducted off the east coast of Florida on the continental shelf.[3] A 200 m HLA was deployed in 250 m of water. The measured sound speed profile at the array site is shown in Fig. 1. The run-geometry and topography of the region are shown in Fig. 5. An acoustic source was towed in various geometries to measure signal evolution and TL as a function of source/receiver position. Many of the geometries were along the 250 m iso-bath, which was ostensibly range independent and broadside to the array. One particular transmission was a linear frequency modulation (LFM) signal transmitted during run 3N (along the solid line in Fig. 5) at a range of 8 km from the receiver. The source was in-shore (upslope) of the array at a relative bearing of 55°. For this set of transmissions the source was deployed at a depth of 20 m. The water depth at the source is 160 m and it rapidly drops (within 2 km) to the continental shelf depth of 250 m.

A set of LFMs and narrowband comb lines were transmitted. For this paper, the matched-filter results from a 7 s wideband LFM from 20 to 420 Hz are presented. The array data was narrowband beamformed [using windowed normalized conventional beamforming as in Eq. (11)] and then inverse Fourier transformed to generate beam-time series. These beam-time series were then matched filtered with the LFM signal to produce beam matched-filter responses. For the model, the PE was computed, across the 200 Hz band, to a range of 8 km and the field was projected onto a 200 m line array on the bottom with source angle relative to the array of 55°. The mode-extrapolated signal was beamformed and the FFT was taken to compute beam matched-filter responses. The 250 Hz narrowband PE, narrow-band mode decomposition vs range, data and simulated beam matched-filter responses are shown in Fig. 6.

From the bathymetry in the TL plot shown in the upper left panel of Fig. 6, it can be seen that the range dependence only occurs over the first 2 km of the transmission path. The strongly downward refracting profile leads to focusing of energy near the bottom. The mode decomposition results (lower left panel) also indicate strong mode coupling out to a range of 2 km. From a ray interpretation the authors expect specular reflections from a downward slope to lead to lower angle rays (and they expect, lower modes). They see, however, that mode energy marches to higher modes. This is due to the effect increased water depth has on the number of propagating modes. As the ocean deepens, there are more waterborne modes within a fixed propagation angle. As the acoustic propagation angles remain unchanged, and the number of propagating modes increases, there must be energy converted to higher modes to accurately represent the same field. In addition to mode coupling in the first 2 km, the authors also see significant mode stripping. Once beyond 2 km, where the bathymetric changes are slight, they see that the mode amplitudes are nearly constant (mode propagation is considered adiabatic). The brief spike just before 4 km is due to the presence of a focal point (evident in the TL plot above). At this location, the vertical field is not highly sampled enough in depth to accurately perform the mode decomposition. Both the modeled and measured beam responses show four to five arrivals, some in doublet form, with later arrivals (presumably higher order bottom bounces) arriving at angles closer to broadside (0° in this case). The measurement and mode-extrapolated synthesis show qualitative agreement in time-spread, arrival angle behavior, and levels. The precise arrival times of individual rays and the energy of later arriving paths is very sensitive to the local sound speed and the local geo-acoustic parameters, respectively. The inversion for the sound speed field (tomography) or the sediment (geo-acoustic inversion) is beyond the scope of this paper.

B. Range-dependent deep-water case (BASSEX-04)

For the range-dependent deep-water case, the authors examine a transmission from the, BASSEX 2004, conducted by Baggeroer, Heaney *et al.*, Worcester, and Dzieciuch. The source was moored in deep water in the central pacific at a depth of 900 m. A towed HLA received broadband *m*-sequences (200–300 Hz) at a location 600 km northeast above the Kermit–Roosevelt seamount, which extends to a minimum water depth of approximately 1200 m. The source-receiver path and the severely range-dependent propagation path around the seamount are shown in Fig. 7.

The challenge here is to efficiently compute the matched-filtered beam-time series for this environment and then compare it with the data. Although the true sound speed field is range-dependent, only measurements taken at the receiver location are available for use in the simulation. The

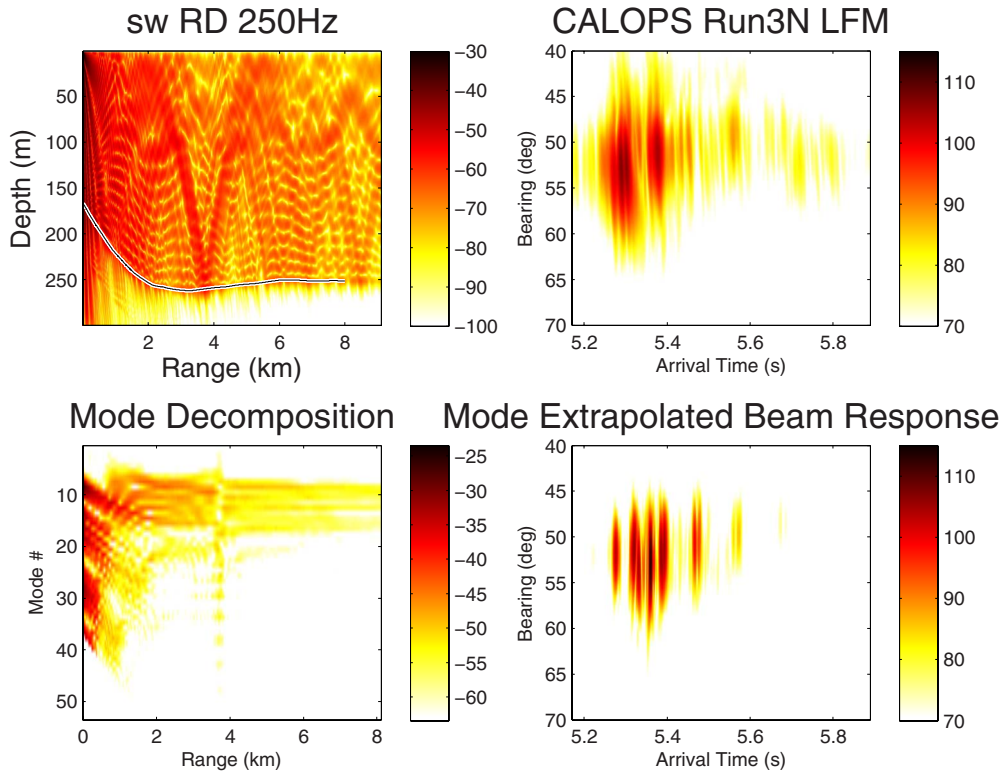


FIG. 6. (Color online) Model-data comparison for shallow-water range-dependent case CALOPS08. (Upper left) 250 Hz TL from PE vs range/depth. (Upper right) Measured beam matched-filter response. (Lower left) 250 Hz mode decomposition (dB). (Lower right) Simulated beam impulse response from mode extrapolation and beamforming.

bathymetry in deep water is taken from the Smith–Sandwell[5] database and near the seamount is taken from a ship-mounted multi-beam system collected during the test. The narrowband PE solution (TL) is plotted in Fig. 8, along with the range-dependent results of mode filtering [Eq. (5)]. For this environment, a deep-sea soft sediment model is assumed, with a thickness of 20 m, overlying a hard basalt basement (neglecting shear). The TL field, shown in the upper panel of Fig. 8 shows deep-water axial propagation until the seafloor rises at a range of 600 km. Then sound is scattered off the seamount to higher angles and energy is stripped from the water column. Beyond the seamount there is evidence of an acoustic shadow as sound re-radiates in more of a convergence-zone type propagation.

The nearly constant mode amplitudes indicate adiabatic propagation for the ranges before 480 km. As the sound interacts with the rising seafloor there is strong coupling to

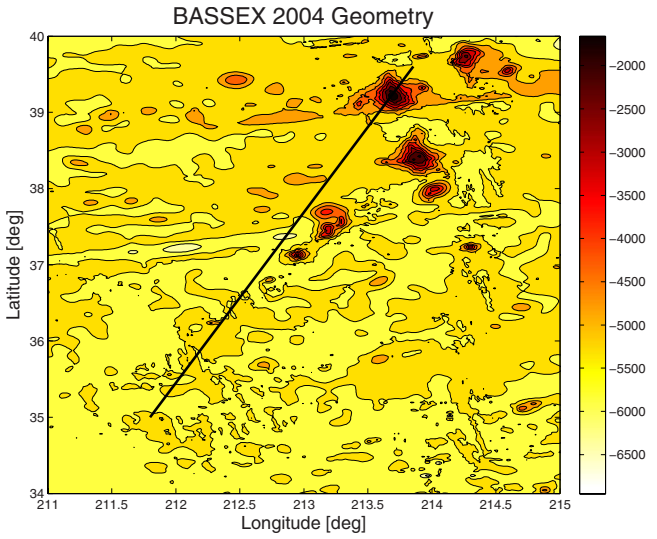


FIG. 7. (Color online) Bathymetry and source-receiver path from BASSEX 04 scattering off the Kermit–Roosevelt seamounts.

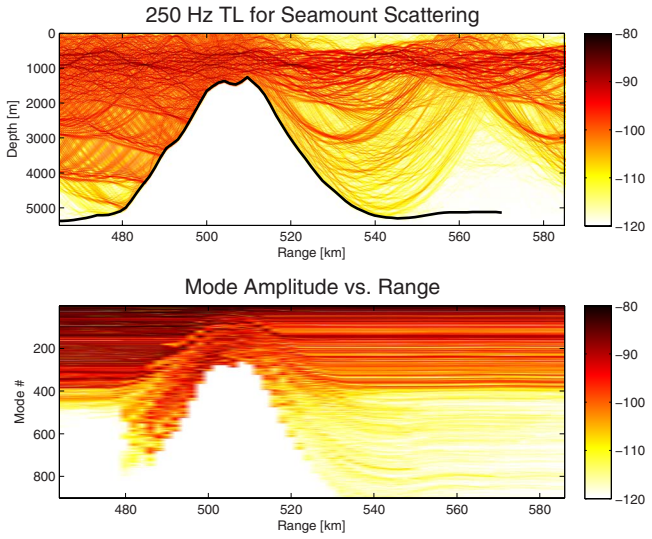


FIG. 8. (Color online) Narrowband 250 Hz PE TL field (upper panel) and mode amplitudes (lower panel) for range-dependent deep-water case. Propagation over the Kermit–Roosevelt seamount from an axial source 510 km away.

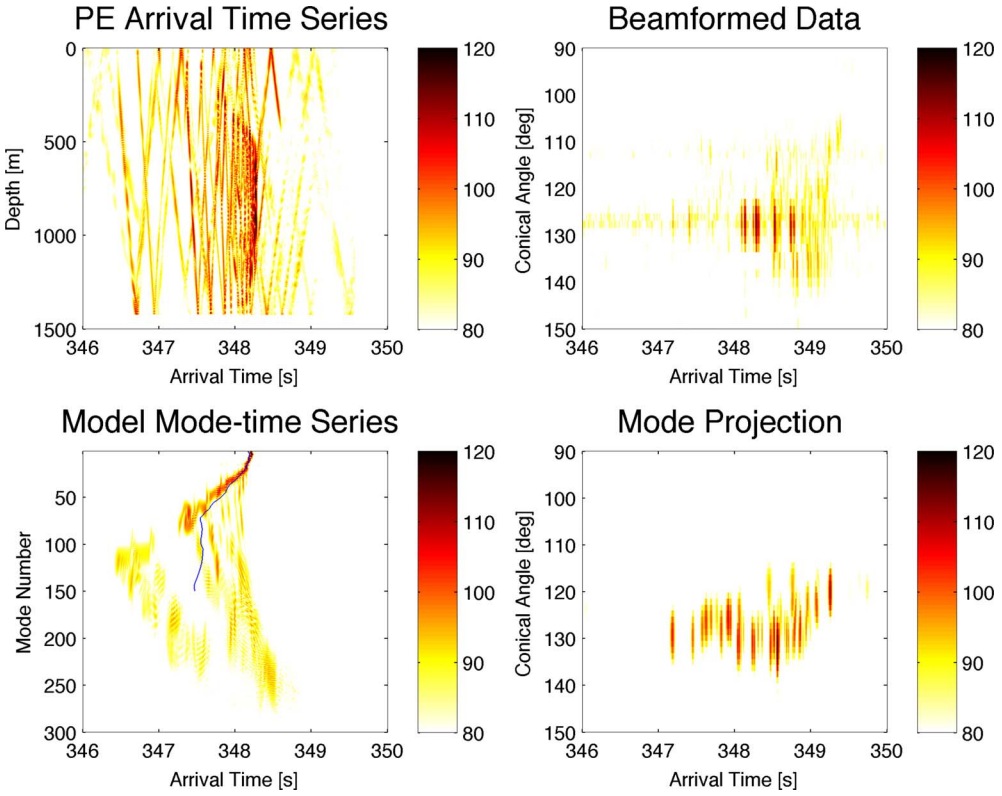


FIG. 9. (Color online) Broadband model and data comparisons for position just above the seamount. (Upper left) broadband time/depth model output showing deep-water dispersion pattern and seamount interacting arrivals. (Lower left) mode arrivals from PE projection and adiabatic mode (line). (Upper right) Data results for beam-time series from 300 m HLA. (Lower right) Simulated beam-time series using the mode projection algorithm.

higher modes. As the seamount rises, higher modes are cut off and energy is either transferred to lower modes or is absorbed into the bottom. The sound that is scattered backwards is not accounted for in the version of the forward PE model used here. Three dimensional propagation is also neglected in this computation. The fact that the cut-off mode number is a linear function of water depth is evident[2] in the shape of the seamount being visible in the mode-amplitude plot. As the sound field propagates downslope, there is a conversion to higher modes, analogous to the first kilometers in the downslope propagation example shown in Fig. 4 (lower left). Once beyond the seamount (ranges greater than 540 km) propagation is again nominally adiabatic.

In order to compare the measured data with simulations, the beam-time series must be computed for the HLA positioned at 514 km in range, just above the seamount. A data section is used when the array was oriented with the source at bearing of 135° relative to the forward endfire beam. To generate the broadband arrival time structure, the field is computed for each frequency from 200 to 300 Hz (a FFT window size of 4 s will be used to handle the deep water dispersion). At the receiver range, the frequency dependent mode amplitudes are computed [Eq. (4)]. Then the field is propagated to each element of the 200 m array, positioned with the source at endfire. Hanning shaded conventional plane-wave beamforming [Eq. (11)] is then applied. These steps are done efficiently by performing the single matrix computation of Eq. (13).

The experimental and simulation results are shown in Fig. 9. There were difficulties with the time-synchronization during the experiment and the data time scale has been aligned by eye to match the simulations. The upper left panel shows the simulated depth-time series computed directly from the PE. The deep-water accordion pattern is seen here, particularly at the lower axial modes (arriving at a time of 348.2 s). Note the presence of wavefronts near 200 m depth that are interpreted as scattered from the sea floor, in particular, the strong path at 348.5 s. The lower left panel shows the mode arrival time series. The line overlaid on this is the theoretical mode arrival structure (at the receiver). In this figure the authors see the evidence of the scattering to higher mode angles from reflections off the seamount. The right two panels are the beam-time series for the data (upper) and simulation (lower). In the data there are strong arrivals at nearly horizontal (as would be observed without the presence of the seamount) and there are later arrivals at higher conical angles (interpreted as higher mode energy reflected from the seamount). In the simulations the authors see similar behavior. The higher conical angle arrivals in the simulations are more identifiable than in the data. This could be the result of bottom roughness (which is not included in the simulation) or an inaccurate geo-acoustic model. Note that the relative levels of the refracted and scattered energies do not agree exactly. Another difference between the model and data is that in data there is scattered energy arriving at a conical angle of 110°–115°, whereas in the simulations the minimum angle is 120°. It is possibly a result of neglecting rough surface scattering.

V. SUMMARY

In this paper a method for efficiently simulating the beam response of an arbitrary volumetric array from the field computation of a PE is computed. Examples were presented using a HLA, but that specific geometry is required. The method involves decomposing the pressure field output of the PE in the vertical into local complex mode amplitudes, propagating via normal mode theory from the PE center to each array element and then applying conventional plane-wave beamforming. Each procedure can be written as a matrix multiplication and the entire algorithm can be computed as a single matrix product, transforming from complex pressure field in depth to complex beam response in bearing. A physical interpretation of the first two matrix products (mode decomposition and mode propagation) shows that one step is applying the range-independent (stratified medium) Green's function to propagate from the PE field location to each array element location. The other pair of operations (mode-projection and beamforming) is shown to be equivalent to performing the Hankel-transform on the field of an individual mode arrival.

The technique is applied to two range-independent cases. These examples, one in shallow and one in deep water, are presented to provide the reader with a working intuition of mode filtering and mode projection. Two examples are performed in range-dependent environments. For the shallow-water range-dependent case, data and model results are compared for the downslope propagation environment off the east coast of Florida. In this case, mode coupling is observed and the time-spread and angular behavior of the data are qualitatively represented by the model approach. The deep-water range-dependent case is taken from the BASSEX 2004 experiment where sound was scattered off of a seamount in the open ocean. For this experiment, a broadband signal was transmitted from an axial source in deep water and received on a HLA (oriented at rear-endfire) located just above a seamount. Mode coupling at the seamount is significant. In both the measurements and the model, the authors observe purely refracted horizontal energy as well as delayed high angle energy that has scattered off of the seamount. The technique presented here has applications in any environment were a range-dependent model is required and it is desired to simulate the array response (either in phone complex data or in beam-space).

ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research, Code 32, Ocean Acoustics. The author would like to acknowledge that the CALOPS-08 experiment was conducted by Jeff Vuono, Phil Dinolpho, and James Murray. For the BASSEX04 experiment, the array data were collected by the author and Chief Scientist Art Baggeroer (MIT). The moored source was deployed by Matt Dzieciuch and Chief Scientist Peter Worcester (both of Scripps Institution of Oceanography-UCSD).

[1]M. Collins, "A split-step Pade solution for the parabolic equation method," J. Acoust. Soc. Am. **93**, 1736–1742 (1993).
[2]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1993).
[3]K. D. Heaney and J. J. Murray, "Three dimensional propagation measurements in the continental shelf," J. Acoust. Soc. Am. **125**, 1394–1402 (2008).
[4]K. D. Heaney, A. B. Baggeroer, K. M. Becker, E. Scheer, and K. Vonderheydt, presented at the Underwater Acoustic Measurements, FORTH, Crete, Greece (2005).
[5]W. H. F. Smith and D. T. Sandwell, "Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry," J. Geophys. Res. **99**, 21803–21824 (1994).

Angular scattering of sound from solid particles in turbulent suspension

Stephanie A. Moore and Alex E. Hay
Department of Oceanography, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

(Received 9 October 2008; revised 6 June 2009; accepted 25 June 2009)

Sound scattering by solid particles suspended in a turbulent jet is investigated. Measurements of the scattered amplitude were made in a bistatic geometry at frequencies between 1.5 and 4.0 MHz, and at scattering angles from 95° to 165° relative to the forward direction. Two types of particle were used: nearly spherical lead-glass beads and aspherical natural sand grains. For each particle type, experiments were carried out using ~200 and ~500 μm median diameter grain sizes, corresponding to $0.7 \lesssim ka \lesssim 4$. The sphericity of the sand grains, defined as the ratio of projected perimeter size to projected area size, was 1.08. The lead-glass bead results are consistent with an elastic sphere model. A rigid movable sphere model provides the best fit to the sand data, and the best-fit diameter is within 4% of the equivalent volume size. However, the scattering pattern for sand is systematically smoother than predicted: that is, the undulations in the angular scattering pattern predicted by spherical scatterer theory are present, but muted. This observed departure from spherical scatterer theory is attributed to disruption of the interference among creeping waves by the irregular surfaces of natural sand grains.
 [DOI: 10.1121/1.3180696]

PACS number(s): 43.30.Ft, 43.20.Fn [KGF] Pages: 1046–1056

I. INTRODUCTION

Acoustic remote sensing systems are becoming a primary tool for sediment dynamics research and sediment transport monitoring in aqueous environments (see Ref. 1 for a recent review). The sound scattered from particles in suspension can provide information on their velocity, size, and concentration and thus, importantly, on the sediment flux. Inverse methods for extracting the size and concentration estimates, however, are dependent on a suitably accurate representation of the scattering cross section which, for randomly-oriented assemblages of irregularly shaped particles, is not readily amenable to calculation from first principles. Thus, in the interpretation of scattering measurements from suspensions of natural sand, the grains are assumed to be spherical on average.

Previous experimental investigations have shown that a modified spherical scatterer model can be used to represent the total scattering and backscattering cross sections of suspensions of natural sand.[2–4] These results provide an adequate basis for the inverse problem in dilute suspensions, for which multiple scattering is unimportant. At concentrations exceeding approximately 1% by volume, however, multiple scattering effects cannot be ignored.[5] Such concentrations are encountered in close proximity to the bed, and become especially important in high energy (i.e., sheet flow) conditions.[6] Theoretical estimates of the effects of multiple scattering on the detected signal necessarily require knowledge of the dependence of single particle scattered energy on scattering angle. However, no experimental evidence exists to support the use of a spherical scatterer model at scattering angles other than 0° and 180°.

This paper presents measurements of the variation with scattering angle of the differential scattering cross section of solid particles in aqueous suspensions as a function of acoustic frequency and mean grain size. Both natural sand grains and manufactured lead-glass beads are used in the experiments. The observed cross sections are compared to predictions based on spherical scatterer theory. The primary goals of the paper are to determine the following: (1) whether a spherical scatterer model provides a good fit to the angular variation of sound scattered from irregularly shaped particles like sand, (2) whether an elastic or a rigid sphere model provides the better fit, and (3) whether the experiments provide physical grounds for seeking improvements to the spherical scatterer approximation for natural sand grains.

The measurements were made for values of $0.7 \lesssim ka \lesssim 4$ (k being the acoustic wavenumber, a the scatterer radius). This ka range is typical of acoustic remote sensing studies of sand transport dynamics. As the lowest frequency resonance for a quartz-like or glass-like sphere in water occurs at $ka \sim 5.5$,[7] resonance scattering should be relatively unimportant for $ka < 4$. It is expected therefore that variations in scattered energy with scattering angle over the ka range of the present measurements should be mainly due to the effects of diffraction. Thus, the focus here is on $O(1)$ values of ka, and the effects of the irregular shapes of natural sand grains on diffraction-induced features in the differential scattering cross section.

II. THEORY

Since the waves scattered from individual particles embedded in turbulence add incoherently on average, the average scattered intensity is proportional to the particle number density, N. The ensemble mean-square scattered pressure, $\langle p_s^2 \rangle$, can then be written as (Ref. 8, pp. 438–441)

$$\langle p_s^2(\theta)\rangle = \frac{p_i^2}{r^2}\exp\left[-2\alpha_0 r + \int_0^r N(r)\Sigma_s dr\right]\int_V N\sigma_s(\theta)dV, \tag{1}$$

where $\langle\ \rangle$ denotes the ensemble average, p_i the incident pressure amplitude, α_0 the attenuation in water, Σ_s the total scattering cross section, σ_s the differential scattering cross section, and θ the scattering angle. V is the detected volume, assumed here to have characteristic dimensions small compared to r, the radial distance from V to the receive transducer. As we are considering solid scatterers, it has been assumed that the absorption cross section of the particles can be ignored relative to Σ_s. Note also that Eq. (1) applies to the low scatterer concentrations for which multiple scattering is unimportant.

Assuming that the detected volume changes very little with scattering angle, and that $\sigma_s(\theta)$ is constant within V (the validity of these assumptions is examined in Appendix A), then the ratio of the mean-square pressures at a given scattering angle, θ_0, and at a reference scattering angle, θ_r, is

$$\frac{\langle p_s^2(\theta_0)\rangle}{\langle p_s^2(\theta_r)\rangle} = \frac{\sigma_s(\theta_0)}{\sigma_s(\theta_r)}. \tag{2}$$

The dependencies on incident pressure, particle concentration, and attenuation along the scattered path are thus eliminated. Equation (2) provides the basis for the comparisons between theory and experiment which are presented later in this paper: the measurements yield the ratio on the left; theory that on the right.

Ignoring thermal and viscous effects, the scattered pressure, p_s, for a plane-wave incident on a solid elastic sphere can be written as a sum of partial scattered waves[9] (p. 273):

$$p_s = \frac{|p_i|}{kr}\sum_{n=0}^{\infty}(2n+1)i^{n+1}\sin\eta_n e^{-i\eta_n}P_n(\cos\theta)e^{i(kr-\omega t)}, \tag{3}$$

where ω is the angular frequency, t is time, P_n is the Legendre polynomial of order n, and η_n is the phase shift of the nth partial wave.[10] The amplitude of the scattered wave can also be written in terms of the far-field form factor, f_∞, as

$$|p_s| = \frac{|p_i||f_\infty|a}{2r}, \tag{4}$$

where f_∞ is given by

$$f_\infty = \frac{2}{ka}\sum_{n=0}^{\infty}(2n+1)i^{n+1}\sin\eta_n e^{-i\eta_n}P_n(\cos\theta). \tag{5}$$

Thus the differential scattering cross section, which is the ratio of scattered power to incident intensity[8] (p. 426), is

$$\sigma_s = \frac{|f_\infty|^2 a^2}{4}. \tag{6}$$

The elastic properties of the scatterer enter the computations through the phase shift. The expressions for the phase shift used here are based on Eq. (30) in Ref. 10. Other than scatterer size and the incident wave frequency, the physical properties upon which the resulting phase shifts explicitly depend are the densities of, and compression wave speeds in, the fluid medium and the solid scatterer, and the shear wave speed in the scatterer.[10,11] Thus, an error pointed out by Hickling[12] (see also Ref. 13) involving Poisson's ratio does not enter the computations. The relevant parameter values used here are $\rho=2650$ kg/m^3, $c_p=5100$ m/s, and $c_s=3200$ m/s for quartz;[7] $\rho=2870$ kg/m^3, $c_p=4870$ m/s, and $c_s=2930$ for lead-glass;[14] and $\rho=998$ kg/m^3, $c_p=1483$ m/s for water.[15] The quartz values are used for the comparisons of the theory with the measurements for sand. For the rigid movable sphere case, the shear modulus is infinite, and computations were made using the expressions for the phase shifts in this limit.[7,10]

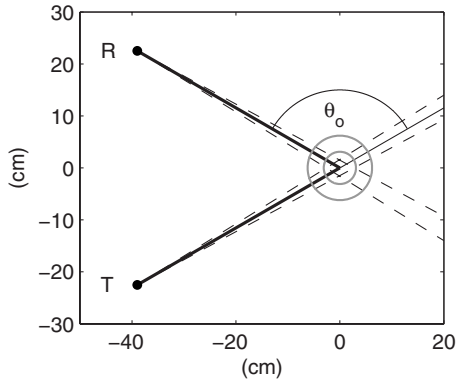


FIG. 1. Scale diagram of the scattering geometry in plan view. The jet axis is into the page, and the jet location is indicated by the concentric grey circles centred at the origin, with radii of σ_J and $2\sigma_J$, respectively, where σ_J is the standard deviation of the sediment concentration profile transverse to the jet axis. T and R represent the transmit and receive transducers, and the heavy black lines indicate the incident and scattered rays for a particle at the jet centreline. The scattering angle for a particle at the centerline, θ_0, is also shown and has a value of 120° in the diagram. The dashed lines represent $\pm 5\beta_0$ (i.e., five times the transducer half-power beamwidth) at 2.9 MHz, to illustrate the narrowness of the beams relative to the width of the jet.

III. METHODS

A. Scattering

The jet tank facility in which the measurements were made is a slightly modified version of that described previously.[16] A vertically-oriented turbulent jet and recirculation system maintain a suspension of particles with a quasi-steady mean concentration, except for a small ($<15\%$) decrease with time over the course of an ~40 min duration experiment (see below). Since the particles are spatially confined within the jet, it is possible to carry out far-field scattering experiments in a crossed-beam geometry without having to correct for attenuation due to particles along the incident and scattered paths (Fig. 1). The modifications to the facility involved: (a) adding a speed controller to the sediment circulation pump and (b) adding an insert to the discharge orifice, thereby reducing the nozzle diameter. As a result of the latter modification, the region of the jet in which the scattering measurements were made was located at a greater distance (measured in nozzle diameters) from the point of discharge.

The suspension is discharged from a 0.95 cm diameter circular nozzle, and recirculated by a variable-speed centrifu-

gal pump. Two broadband transducers were positioned 52.8 cm below the nozzle and 45 cm from the jet centerline. Particle concentrations at this height were determined gravimetrically from samples drawn by suction from the jet axis. Centerline concentrations ranged from ~1 to ~6 kg/m^3. These concentrations are low enough (i.e., <0.3% by volume) that multiple scattering should have been negligible. Prior to the experiments, the water in the tank had been aged for several months to avoid contamination of the signal by scattering from microbubbles. The absence of bubbles was confirmed by checking that no detectable signal was received from the range interval spanning the jet prior to the addition of particles. To prevent any buildup of algae or dust within the tank, the water was continuously filtered through a 1 μm pore-size filter, and 200–300 ml of chlorine bleach were added to the water approximately once a month.

To measure the scatterer concentration, 1 L samples were drawn by suction from the jet centerline at the level of the transducers. The samples were filtered (18.5 cm diameter Q8 Fisher brand filters) and oven-dried at 40 °C for 24–48 h. The dry sand was removed from the filter and weighed to 0.1 mg precision using a Mettler AJ100 balance. This procedure was tested by filtering and drying 12 samples with known initial weights of ~5 g for 24, 47, 71, and 80 h. Three samples were removed from the oven at each drying time, and the final weights measured. The average difference between initial and final weights was −0.17%, and ranged between −0.08% and −0.23% for the four drying times. No discernable trend with drying time was observed.

Three suction samples were drawn at the start and end of each scattering experiment. Suspended sediment concentration, M, was determined from the dry weight of the sand and the volume of water in the sample. The standard deviation of M among the three samples as a percentage of the mean ranged from 0.8% to 12.5%. A second set of triplicate samples was drawn at the end of each experiment, i.e., approximately 40 min after the first set. Scatterer concentrations decreased by 5%–15% over this time interval, because particles inevitably escape from the jet: i.e., the capture cone[16] at the base of the tank is not 100% efficient. The trends with scattering angle associated with this 5%–15% drop in concentration were not removed from the observations.

The broadband piezocomposite transducers (Imasonic) used in this study have a center frequency of 3.0 MHz. Measurements were made between 1.5 and 4.0 MHz at 0.125 MHz intervals: i.e., at 21 separate frequencies. The transmit pulse was produced by a programmable arbitrary waveform generator (National Instruments 5411). The transmit pulse length was 16 μs, corresponding approximately to a 0.0625 MHz energy bandwidth in the pulse, i.e., half the 0.125 MHz frequency interval between adjacent frequencies, assuring negligible frequency overlap in the data at a given scattering angle. This pulse length also corresponds to 1.2 cm resolution in range, which is much less than the 6 cm characteristic width of the jet (see Fig. 1, and below). Thus, the signal at the range corresponding to the jet centerline is representative of the more uniform mean concentration at that location. The ping interval was 10 ms, long enough for

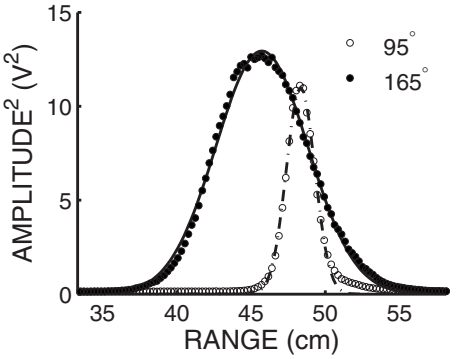


FIG. 2. Squared mean signal amplitude versus range for the large (approximately 500 μm median grain diameter) sand at scattering angles of 95° and 165°. The lines are the best-fit Gaussians, with standard deviations of 0.96 and 3.10 cm, respectively.

echoes to die out at these frequencies. The receive signal, after amplification and rectification, was digitized at 320 kHz with 12-bit resolution. The signals from 10 pings were ensemble-averaged and stored, and 100 ensemble-averaged profiles were acquired at each frequency and angle.

The transducers were mounted equidistant from the jet centerline in a bistatic configuration with their acoustic axes in the same horizontal plane and intersecting at the jet centerline (Fig. 1). Each transducer was suspended from a rigid arm which could be rotated in the horizontal plane about a pivot point at the jet centerline. Measurements were made at scattering angles from 95° to 165° in 5° increments, starting at 165°. At the end of each experiment, a duplicate set of measurements was collected at 165°. The resulting differences in signal amplitude were typically less than 5%, and non-systematic.[17]

The discharge velocity at the nozzle was 4 m/s. Using the nozzle diameter as a length scale, the discharge Reynolds number was $\sim 3 \times 10^4$. The transducer beams intersected the jet at an axial distance from the discharge of 55 nozzle diameters. At this distance, time-averaged transverse profiles of velocity and scalar quantities, including suspended sediment concentration, are expected to be Gaussian for turbulent round jets.[16,18] The square of the time-averaged scattered signal provides a measure of concentration [Eq. (1)]. Profiles of the squared rms voltage for the large sand grains at $\theta = 95°$ and 165° are shown in Fig. 2, together with the best-fit Gaussians. Designating σ_J as the standard deviation of the Gaussian fit, the characteristic width of the jet, $2\sigma_J$, is 6.2 cm at the 165° scattering angle. As the figure indicates, the profile at 95° is much narrower. This reduction in apparent width is a geometric effect, arising primarily from the reduced overlap of the transmit and receive transducer beam patterns as the scattering angle approaches 90° (see Appendix A). Since the received voltage is proportional to the scattered pressure, the peak value in the rms scattered voltage profiles (i.e., the data, not the fit in Fig. 2) was used for the left-hand side of Eq. (2) in the comparisons between theory and experiment presented later.

As a quantitative measure of how well the various models fitted the data, the γ^2 statistic was computed. (γ^2 is one minus the error variance skill score defined in Ref. 19.) For

observations, X, and predictions, Y, both of which are functions of frequency and scattering angle, this statistic is given by

$$\gamma^2 = \frac{\mathrm{Var}(X-Y)}{\mathrm{Var}(X)}, \tag{7}$$

where Var denotes the variance with respect to the mean over all values in the $[\theta,f]$ domain spanned by the observations. Since this statistic is normalized by the variance of the observations, differences in γ^2 values provide unbiased measures of the relative predictive skill of different models. The theoretical prediction with the smallest value of γ^2 represents the best fit. In addition, the correlation coefficient, R^2, between the values of X and Y with means removed, and the rms deviation $\epsilon=\sqrt{\mathrm{Var}(X-Y)}$, also with means removed, were computed for each set of observations and corresponding best-fit theory.

B. Particle size

For both the lead-glass beads and sand two sizes were used: one with a median diameter of ~200 μm and the other ~500 μm, referred to hereinafter as the "small" and "large" particles. The sand used was first sieved into narrow (1/4-phi) fractions. (The phi-scale is given by $-\log_2 d$, with d the particle diameter in millimeters.) The sand retained between the 180 and 212 μm sieves is the small sand, while the large sand was that retained between the 425 and 500 μm sieves. High resolution sand size distributions were determined via electroresistance (Coulter Counter) and image analysis using the sand from the suction samples drawn from the jet. The lead-glass beads were ordered from the manufacturer in relatively narrow size distributions. High resolution size distributions for the lead-glass beads were obtained by Coulter Counter using the suction samples taken from the jet. The image analysis methodology could not be applied to these particles because they are transparent.

For the sieve analyses, samples weighing 10–15 g were shaken in a stack of 20 cm diameter sieves for 15 min following accepted procedures.[20] National Institute of Standards and Technology (NIST) 1017b and 1018b glass beads were used to calibrate the sieves.

For the Coulter Counter measurements, roughly 0.5 g of particles from the jet suction samples were suspended in a 4 L beaker containing a solution of 64% de-ionized distilled water, 35% glycerine, and 1% sodium chloride. The Coulter Counter (model Multisizer II) was used with a 1000 μm diameter aperture. Between 5000 and 10 000 particles were counted in each run. Two separate runs were averaged for each particle type, yielding distributions of the (volume) equivalent spherical diameter.[21] The Coulter Counter size estimates are designated by d_c.

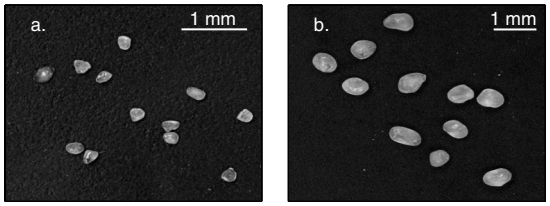


FIG. 3. Photographs of (a) the small and (b) the large sand grains.

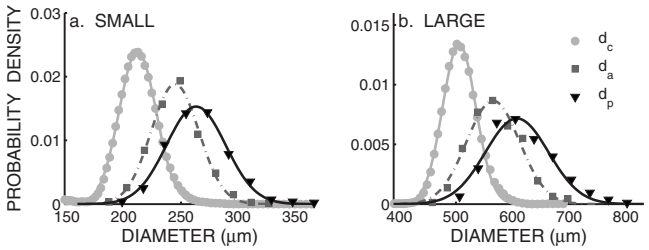


FIG. 4. Size distributions determined by Coulter Counter and image analysis for (a) the small sand (approximately 200 μm median grain diameter) and (b) the large sand (approximately 500 μm median grain diameter). The symbols in the legend indicate the different diameter estimates, as follows: d_c, the Coulter Counter diameter; d_a, diameter based on the projected area; and d_p, diameter based on projected perimeter. The d_a and d_p estimates were determined by image analysis. Points represent data and lines represent the best-fit Gaussian distributions.

Size determination by image analysis was based on 3872×2592 pixel photographs of the particles resting on a flat surface taken with a Nikon D80 digital single-lens reflex camera equipped with a 60 mm focal length macro lens. Example images of the small and the large sand grains are shown in Fig. 3. Particle boundaries were determined using an edge detection method based on a pixel brightness threshold. Equivalent diameters corresponding to the equal projected area sphere, d_a, were determined from the area enclosed by each boundary and the equal projected perimeter sphere, d_p, from the perimeter of the enclosed area. The ratio d_p/d_a for each particle, averaged over a large number of particles, is approximately equivalent to averaging over particle orientation. This average, $\langle d_p/d_a\rangle$, provides a measure of particle sphericity. Photographs of nominally spherical (black) basalt beads of comparable size were analyzed using the same approach to validate the method.

The Coulter Counter and image analysis size distributions are used for the comparisons between theory and experiment because (1) being based on the samples drawn from the jet centerline, they are representative of the actual size distributions of the scatterers in the jet, and (2) they resolve the distributions within each 1/4-phi sieve fraction.

IV. RESULTS

A. Particle size

The size distributions for the small and large sand, as measured with the Coulter Counter, are shown in Fig. 4. Both distributions are nearly Gaussian. Table I lists d_{16}, d_{50},

TABLE I. Coulter Counter size distribution statistics.

Particle	d_{16} (μm)	d_{50} (μm)	d_{84} (μm)
Small sand	196	211	227
Large sand	476	502	532
Small beads	182	212	242
Large beads	384	424	471

TABLE II. Image analysis size distribution statistics. N is the number of analyzed particles while d and σ are the distribution mean and standard deviation. Subscripts a and p denote parameters from projected area and projected perimeter, respectively.

Particle	N	$d_a \pm \sigma_a$ (μm)	$d_p \pm \sigma_p$ (μm)	$d_p/d_a \pm a$
Small sand	999	245 ± 21	265 ± 26	1.08 ± 0.03
Large sand	582	569 ± 44	618 ± 55	1.08 ± 0.03
Small basalt	442	188 ± 10	193 ± 11	1.024 ± 0.009
Large basalt	176	455 ± 24	475 ± 26	1.044 ± 0.006

and d_{84} for both the sand and the lead-glass beads. These diameters represent to the 16th, 50th, and 84th percentiles of the cumulative distribution, and correspond to the mean and one standard deviation below and above the mean for a Gaussian distribution.

The sand size distributions determined by image analysis are also plotted in Fig. 4, and the statistics summarized in Table II. The average value of d_p/d_a for both sand sizes was 1.08 ± 0.03. The same analysis for the nominally spherical basalt beads yielded 1.02 ± 0.009 and 1.04 ± 0.006 for the small and large bead sizes, respectively. As a further comparison, averaging over the projections of all possible orientations of a unit cube in a simulation gave $\langle d_p/d_a \rangle \sim 1.4$. The latter value is larger than the 1.08 measurement for sand, indicating that the sand grains used in the experiments were less angular than cubes, consistent with their somewhat rounded appearance (Fig. 3).

As expected for non-spherical particles, the size distributions for sand obtained with the various methods differ noticeably. The ratios of mean Coulter Counter size to mean projected area size, d_c/d_a, are 0.80 for the small sand and 0.84 for the large sand. Some of the difference between these two measures of size can be attributed to biases intrinsic to the different techniques. For comparison, other investigators[22] have found that, for a unimodal distribution of 300–500 μm sieve diameter "spherical" and "nearly spherical" standard reference glass beads, the average value of d_c/d_a was 0.86, similar to the values obtained here.

B. Scattering results

The observed and best-fit theoretical angular scattering patterns for the small lead-glass beads and small sand are shown in Fig. 5. The data for lead-glass are the average of three experiments, those for sand the average of two. The corresponding results for the large particles are presented in Fig. 6, the data being the average of three experiments for both particle types. All scattered amplitudes have been normalized by the values at 165°: i.e., in Eq. (2), $\theta_r = 165°$.

The normalized theoretical scattered amplitude, Y, is given by

$$Y = \frac{\sqrt{\int_0^\infty |f_\infty(\theta,a)|^2 a^2 n(a) da}}{\sqrt{\int_0^\infty |f_\infty(165\,°,a)|^2 a^2 n(a) da}}, \tag{8}$$

where $n(a)$ is the particle size probability density. Y was computed for a range of mean diameters in steps of 10 μm using Gaussian size distributions with the same breadth-to-mean diameter ratios (i.e., σ/d_{50}) as the Coulter Counter results. For each particle type, the mean diameter and model (rigid or elastic sphere) yielding the minimum value of γ^2 were deemed the best fit.

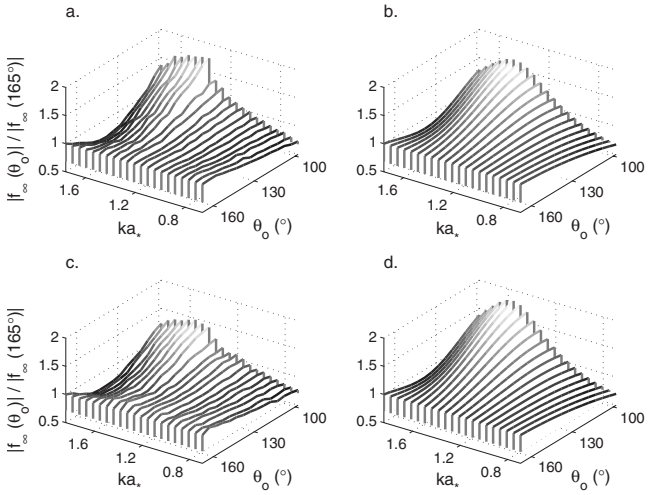


FIG. 5. Observed and predicted angular scattering patterns normalized by the amplitude at $\theta_0 = 165°$ for [(a) and (b)] small lead-glass beads and [(c) and (d)] small sand. Data are on the left, best-fit theory on the right. The theoretical computations assume elastic spheres for the glass beads ($d_* = 2a_* = 210$ μm, $\sigma_* = 30$ μm), and rigid spheres for the sand grains ($d_* = 2a_* = 220$ μm, $\sigma_* = 17$ μm).

The γ^2 values corresponding to the best-fit models for the different particle types are plotted in Fig. 7 versus d/d_*, where d is the theoretical mean diameter, and d_* is the best-fit theoretical mean. The minimum values of γ^2 (i.e., the values at $d = d_*$) for the best-fit models in each case are listed in Table III, together with the corresponding values of ϵ and R^2. In all cases, the best-fit diameter, d_*, is very close to the median Coulter Counter size. The minima in Fig. 7 tend to

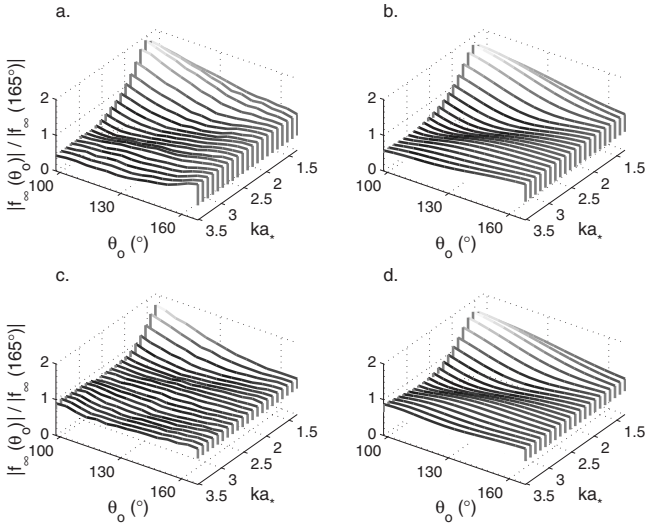


FIG. 6. Observed and predicted angular scattering patterns normalized by the amplitude at $\theta_0 = 165°$ for the large particles: [(a) and (b)] large lead-glass beads; [(c) and (d)] large sand. Data are on the left, and best-fit theory on the right. (Note: the viewpoint is different from Fig. 5.) The theoretical computations assume elastic spheres for the glass beads ($d_* = 2a_* = 420$ μm, $\sigma_* = 44$ μm) and rigid spheres for the sand grains ($d_* = 2a_* = 520$ μm, $\sigma_* = 29$ μm).

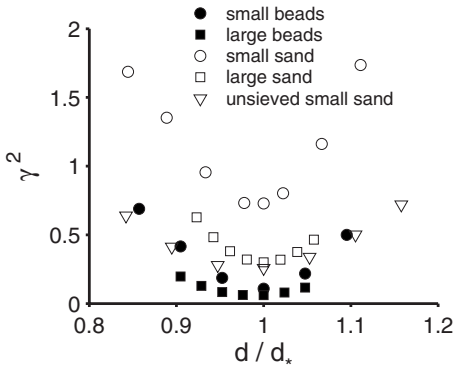


FIG. 7. The values of γ^2 for the best-fit models versus the ratio of theoretical mean diameter, d, to the best-fit theoretical mean diameter, d_*, for all scatterer types.

be more pronounced for the small and large sand, due to their size distributions being relatively narrow compared to those of the lead-glass beads and the unsieved sand (Table III).

Over the measured frequency range, the values of d_* correspond to ka_* ranges of 0.66–1.76 for the small lead-glass beads, and 0.69–1.84 for the small sand. In contrast, the ranges for the large particles are $1.32 \leqslant ka_* \leqslant 3.52$ for lead-glass and $1.63 \leqslant ka_* \leqslant 4.36$ for sand. For quartz-like and/or glass-like particles, the rigid and elastic sphere model predictions differ by less than 7% for $ka < 1$, whereas for $1 < ka < 4$ they differ by as much as 30%. Thus, the effects of the scatterer elastic constants on the predictions are much greater at the higher ka values corresponding to the larger particles. This is reflected by the greater differences in γ^2 between the two models for the large particles, compared to the corresponding differences for the small particles (Table IV). Consequently, the present results demonstrate that (1) the elastic model provides the better fit to the lead-glass bead data and (2) the rigid model the better fit to the sand data.

For the most part, the observed and predicted surfaces in Figs. 5 and 6 exhibit good qualitative and quantitative agreement. For sand, however, while the overall shapes of the experimental and theoretical scattering patterns are similar, the ridges and valleys in the observed patterns are less pronounced than the predictions. This effect, which is clearest in the results for large sand grains [Figs. 6(c) and 6(d)], contributes to the systematically higher values of γ^2 for sand compared to lead-glass.

The large bead and sand data are plotted versus scattering angle for three frequencies in Fig. 8, together with the predictions from the best-fit rigid and elastic sphere models.

TABLE IV. The values of γ^2 computed for the elastic and rigid sphere models using d_* as the theoretical mean diameter.

Scatterer	d_* (μm)	γ^2 (elastic)	γ^2 (rigid)
Small beads	210	0.11	0.12
Large beads	420	0.06	0.19
Small sand	220	1.01	0.73
Large sand	520	0.71	0.30

Also shown are the predictions of the high-pass model. This model was originally put forward by Johnson[23] for backscatter from a fluid sphere, and later modified for total scattering from sand grains.[2] The latter study also suggested a form of the high-pass model for angular scattering, but with no supporting observational evidence, as none was available at the time. Figure 8 indicates that the elastic sphere model best reproduces the measured values for the lead-glass beads, whereas the rigid sphere model best reproduces the large sand data. In addition, the results in Fig. 8 demonstrate that the high-pass model does an unacceptable job of reproducing the observations for either scatterer type.

Additional scattering measurements were carried out with unsieved small sand: that is, the small sand before it was sieved into 1/4-phi size fractions. The values of γ^2 from the best-fit theory to the data are plotted versus d/d_* in Fig. 7. The overall best-fit parameters are summarized in Table III. As for the narrow sand fractions, the best-fit model for these sand data is the rigid sphere. The fit yielded a minimum γ^2 value of 0.26, comparable to the 0.30 value for the large sand.

V. DISCUSSION

A. The effects of irregular particle shape: Diffraction smearing

Overall, the best agreement between measured and predicted scattering was obtained for the large lead-glass beads and the elastic sphere model, yielding a minimum value of γ^2 of 0.06 (Table IV). For comparison, the minimum γ^2 value for the large sand and the rigid sphere model was 0.30. It is concluded that, for the ka range of the present measurements, a rigid sphere is not as good a model for scattering by natural sand grains as is the elastic sphere model for the lead-glass beads. The likely cause of this difference is the irregular shape of the sand grains.

TABLE III. Summary statistics for the best-fit models to the angular scattering data. Particle size statistics (from Coulter Counter analysis except for the "unsieved" sand, which are from sieve analysis) are included for convenience.

Scatterer	d_{50} (μm)	$(d_{84}-d_{16})/2$ (μm)	Model	d_* (μm)	ϵ	R^2	γ^2	d_*/d_{50}
Small beads	212	30	elastic	210	0.09	0.95	0.11	0.99
Large beads	424	44	elastic	420	0.10	0.97	0.06	0.99
Small sand	211	16	rigid	220	0.18	0.91	0.73	1.04
Large sand	502	28	rigid	520	0.08	0.92	0.30	1.04
Unsieved sand	203	62	rigid	190	0.08	0.86	0.26	0.94

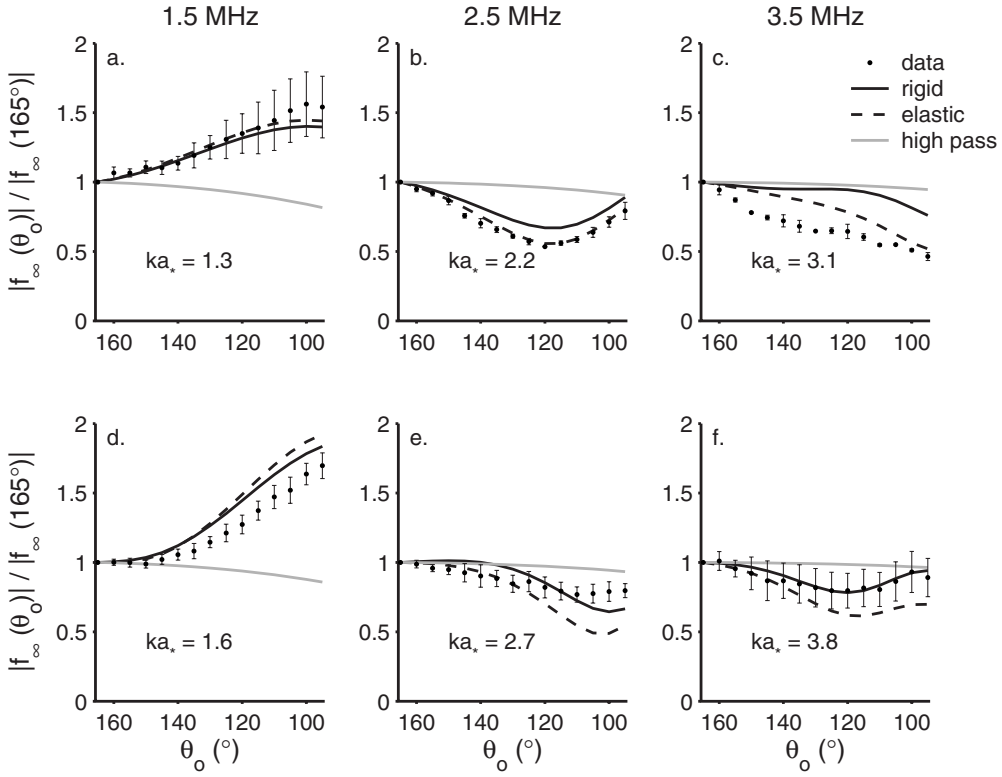


FIG. 8. Measured and predicted scattered amplitudes at three frequencies for suspensions of [(a)–(c)] large lead-glass beads and [(d)–(f)] large sand, normalized to $\theta_0=165°$. The error bars represent ± the standard error (i.e., the standard deviation divided by $\sqrt{2}$, for three repeat experiments). The equivalent plots for the small particles are presented and discussed in Appendix B.

For a plane-wave incident on a sphere, the diffraction pattern is due to interference between the incident wave field and energy leaked by creeping waves traveling along geodesics on the surface of the sphere.[24–26] For an irregularly shaped particle, the creeping wave circuits will not be of equal length, and the constant phase relationships required for interference will be disrupted. Thus, as with the lack of resonant features in backscattering and total scattering cross section measurements for sand at higher values of ka,[3,4,16] the diffraction pattern for a suspension of randomly-oriented sand grains should be a blurred version of that for a smooth sphere. To mimic the smearing effect, we tried smoothing the best-fit predictions along the θ-axis using a Gaussian weighting function with a standard deviation of 5°. For the smearing mechanism to be consistent, smoothing the theory in this way should lead to improved agreement, and indeed does. For large sand, the smoothed theory led to a reduction in γ^2 from 0.30 to 0.21.

B. Effective particle size

Thorne and Buckingham[4] showed that measurements of the backscattering and total scattering cross sections for sand in aqueous suspensions can be collapsed onto smoothed versions of the predictions for a sphere, using non-linear scaling with a single free parameter. However, the value of the scaling parameter differed among the available data sets, and varied systematically with particle size (by almost a factor of 2) for some of the total scattering cross section measurements.

In principle, there are two independent particle size scales involved when comparing measurements of scattering from particulate suspensions with narrow size distributions to spherical scatterer theory. One scale is the circumference of an equivalent sphere to scale the wavelength, yielding ka for the scattering computations. The second scale is the diameter of the equal volume sphere for converting particle mass concentration, M, to particle number density, N: that is, $\rho_s N=M/v_p$, where ρ_s is the grain density and v_p is the particle volume. M is measured (gravimetrically here, using the suction samples), but N is required in the theory [see Eq. (1)]. For irregularly shaped scatterers, there is no reason to expect, *a priori*, that the two scales should be the same.

Schaafsma and Hay[3] found that measurements of acoustic attenuation in aqueous suspensions of natural sand could be brought into agreement with rigid sphere theory using a two-parameter approach. Their scaling is linear: that is, the scaled diameter is given by Bd, where d is the measured diameter and B is one of the parameters. The values obtained for the wavelength scaling parameter were relatively constant. In contrast, the volume scaling parameter varied with mean size in a manner consistent with the fact that the particles tended to be less rounded and more angular with decreasing size. It is convenient here to designate the two parameters as B_λ and B_N, the subscripts denoting their respective physical roles.

Because normalizing the scattered amplitudes eliminates the dependence on scatterer concentration [see Eq. (2)], the results presented here should depend on B_λ only. Thus, it is interesting that the ratio of best-fit acoustic size to measured

mean size, i.e., B_λ, is within 4% of unity for the Coulter Counter measured mean sizes (Table III). Since the Coulter Counter method yields a volume equivalent size, and since B_N should be identical to unity for volume-based size measurements, an implication is that the two linear scaling parameters could be reduced to one when the measured size is based on particle volume, at least for the ka range of the present data, and for sand grains of comparable sphericity.

For the unsieved sand, the best-fit acoustic size differed from the measured median diameter by only 6% (Table III), but the measured size in this case was obtained by sieving. Given the greater breadth of the size distribution for the unsieved sand, this result might suggest that, for the broader distributions more likely to be encountered in natural environments, sieve size could be used instead of Coulter Counter size. However, as particle volume is proportional to d^3, an error of 5%–10% in size would lead to an error of 15%–30% in number density N, which might not be acceptable.

VI. SUMMARY AND CONCLUSIONS

The angular dependence of sound scattering from particles suspended in a turbulent jet has been investigated. The measurements were made from 1.5 to 4.0 MHz with 125 kHz resolution, at scattering angles ranging from 95° to 165°, for suspensions of both nominally spherical lead-glass beads and natural sand. Two narrow size fractions for each of the beads and sand were used: one with a nominal mean diameter of ~200 μm and the other, ~500 μm. Particle size was determined by Coulter Counter for both particle types. Optical image analysis was used to quantify the irregularity of the sand grain shapes. While the transparency of lead-glass precluded the use of the image analysis method for the beads, their shapes had been examined previously by scanning electron microscopy.[3] The size distributions were nominally Gaussian for all particles, and especially so for the small and large sand size fractions. The sphericity of the sand grains, defined as the ratio of projected perimeter size to projected area size, was 1.08 ± 0.03. The scattering measurements were made at non-dimensional wavenumbers of $0.7 \leqslant ka_* \leqslant 1.8$ for the small lead-glass beads and $1.3 \leqslant ka_* \leqslant 3.5$ for the large beads, and at $0.7 \leqslant ka_* \leqslant 1.8$ and at $1.6 \leqslant ka_* \leqslant 4.4$ for the small and large sand size fractions respectively.

Theoretical angular scattering patterns were computed for both rigid movable and elastic movable spheres assuming Gaussian size distributions with the same breadth-to-mean diameter ratios as the measured distributions. Effective acoustic mean diameters were determined by least-squares fitting the experimental results to the predicted scattering patterns, resulting in an overall best-fit mean diameter (d_*) and overall best-fit model (rigid or elastic) for each scatterer type and size fraction. Consistent with previous measurements of the backscattering and total scattering cross sections for glass bead and natural sand particles in suspension, the elastic sphere model provided the best fit to the lead-glass bead data; the rigid sphere model provided the best fit for sand. The modified high-pass model[2] does not fit the data, except at the smallest values of ka.

For each particle type and each size fraction, the best-fit mean diameter was very close to the Coulter Counter value: within 1% for the lead-glass beads and within 4% for the sand. Since the possible error in the Coulter Counter measurements is a few percent at least,[21,27] these values are probably not significantly different from zero. Since particle size measured by the Coulter Counter should be close to the diameter of a sphere of equal volume, the results for the nominally spherical lead-glass beads are expected. The results for sand indicate that the diameter of an equal volume sphere can be used to scale the acoustic wavenumber, at least over the ka range of the present measurements and for sand grains with sphericity comparable to the sand used here. This result has potential implications for inverting acoustic scattering data to suspended sand size and concentration.

The lead-glass bead data are in better agreement with the elastic sphere model than are the sand data with the predictions for a rigid sphere. Since the measurements were made at ka values below the resonances for quartz spheres, but well above the Rayleigh range, the departures from spherical scatterer theory for sand must be related to diffraction and must involve sand grain shape. The authors conclude that the irregular shapes of natural sand grains partially disrupt the creeping wave interferences responsible for the diffraction pattern for a smooth sphere. As a result, the diffraction-induced undulations in the scattering pattern are smoother than predicted by spherical scatterer theory. One implication of this result is that it provides justification for smoothing the scattering cross sections predicted by spherical scatterer theory in the so-called diffraction region of ka-space, even for suspensions of sand grains with very narrow size distributions.

ACKNOWLEDGMENTS

We thank Wesley Paul and Robert Craig for technical support, and Brent Law and Tim Milligan for the Coulter Counter measurements. This project was supported by the Natural Sciences and Engineering Research Council of Canada.

APPENDIX A: BISTATIC SCATTERING FROM ISOTROPIC, FREQUENCY-INDEPENDENT SCATTERERS IN THE JET

The purposes of this Appendix are (1) to verify that the change in detected volume with scattering angle is small, an assumption made in obtaining Eq. (2); and (2) to explain the change in apparent jet width with scattering angle indicated in Fig. 2. To do so, the angular dependence of scattering from the jet is investigated assuming the scattering cross section to be independent of frequency and scattering angle. The predicted variations with scattering angle will thus be entirely due to geometric effects, and the sole source of frequency variation will be the transducer beam patterns. The coordinate system is sketched in Fig. 9. The axis of the jet is

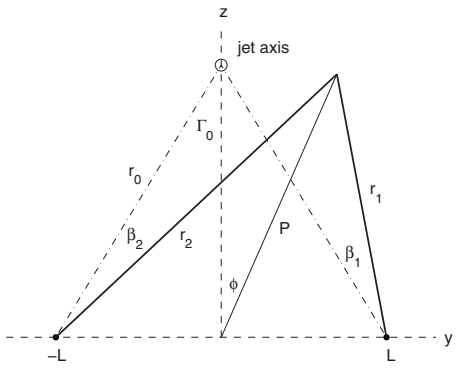


FIG. 9. Sketch of bistatic scattering geometry. The transducers are located at $y=\pm L$. The x-axis (not shown) is out of the page and anti-parallel to the axis of the jet (indicated by the circle).

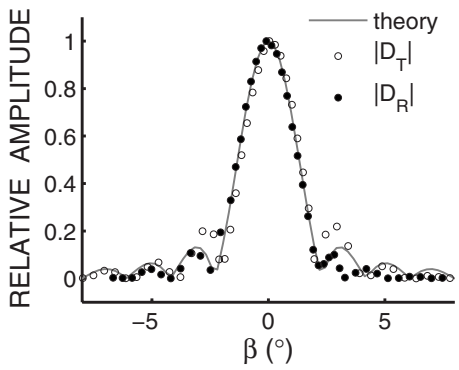


FIG. 10. Measured transducer directivity at 2.9 MHz for the transmit, D_T, and receive, D_R, transducers compared to Eq. (A11) (a_0=0.75 cm). β is the angle of the field point relative to the acoustic axis (see Fig. 9).

anti-parallel to the positive x-axis, and the origin is the midpoint between the two transducers, each located at $y=\pm L$ in the $x=0$ plane.

Scattered signals from different particles arrive at the receiver at the same time (i.e., within a wave period) if

$$r_1 + r_2 = 2r_0 \tag{A1}$$

This relation defines an ellipsoidal surface centered on the origin with its axis of symmetry coincident with the y-axis. Consider any field point with position vector $\boldsymbol{P}=x\hat{i}+y\hat{j}+z\hat{k}$ that satisfies Eq. (A1). In spherical polar coordinates, $x=P\sin\phi\cos\varphi$, $y=P\sin\phi\sin\varphi$, and $z=P\cos\phi$, ϕ being the polar angle and φ the azimuthal angle. Thus, $\boldsymbol{r}_1=\boldsymbol{P}-\boldsymbol{L}$ and $\boldsymbol{r}_2=\boldsymbol{P}+\boldsymbol{L}$, where $\boldsymbol{L}=L\hat{j}$. It follows that

$$r_2^2 - r_1^2 = 4Ly, \tag{A2}$$

and

$$P = \left[\frac{r_1^2 + r_2^2}{2} - L^2\right]^{1/2} \tag{A3}$$

Equations (A1) and (A2) yield

$$r_1 = r_0 - \frac{L}{r_0}y \tag{A4}$$

and

$$r_2 = r_0 + \frac{L}{r_0}y. \tag{A5}$$

Using the relation $\boldsymbol{r}_0\cdot\boldsymbol{r}=r_0 r\cos\beta$, the angle of the field point relative to the acoustic axis of the transducer at $y=L$ is given by

$$\cos\beta_1 = \frac{r_0 z\cos\Gamma_0 - L(y-L)}{[x^2+(y-L)^2+z^2]^{1/2}[L^2+r_0^2\cos^2\Gamma_0]^{1/2}} \tag{A6}$$

and that relative to the axis of the transducer at $y=-L$ by

$$\cos\beta_2 = \frac{r_0 z\cos\Gamma_0 + L(y+L)}{[x^2+(y+L)^2+z^2]^{1/2}[L^2+r_0^2\cos^2\Gamma_0]^{1/2}}, \tag{A7}$$

where $2\Gamma_0$ is the angle subtended at the jet centerline by the transducer baseline, and is related to the scattering angle θ_0 by $2\Gamma_0=\pi-\theta_0$.

The radial distance to the field point from the jet centerline is

$$\rho = [y^2 + (z - r_0\cos\Gamma_0)^2]^{1/2}. \tag{A8}$$

Let σ_{J0} be the jet standard deviation at $x=0$ and x_* the distance from the nozzle to the plane of the transducers. Then $\sigma_J=\sigma_{J0}(x_*-x)/x_*$ accounts for the linear spreading of the jet with downstream distance, and

$$N(x,y,z) = \frac{x_* N_0}{x_* - x}\exp[\rho^2/2\sigma_J^2] \tag{A9}$$

is the number density of scatterers. N decays hyperbolically with distance from the nozzle, as required for round jets.

For omnidirectional scatterers with scattering cross sections independent of frequency, the mean-squared scattered pressure is then proportional to

$$\int\int \frac{D_1^2(\beta_1)}{r_1^2}\frac{D_2^2(\beta_2)}{r_2^2} NP^2\sin\phi d\phi d\varphi, \tag{A10}$$

where $D(\beta)$ is the transducer directivity. The theoretical directivity for a circular piston transducer is given by[28]

$$D(\beta) = 2J_1(ka_0\sin\beta)/ka_0\sin\beta, \tag{A11}$$

J_1 being a cylindrical Bessel function. Figure 10 shows the measured beam patterns for both transducers compared to Eq. (A11). The measurements were made using a standard target located at the jet axis, by rotating one transducer about a vertical axis while the other remained fixed. The standard target was a 1 m long, 0.236 mm diameter stainless steel rod, suspended vertically below the nozzle. Scattering of spherical waves by long cylinders and their use as standard targets has been discussed elsewhere.[29,30] In the figure, the comparison between theory and experiment is quite good, although the value of 0.75 cm used for a_0 in the calculation is 15% larger than the manufacturer's specified radius for the active element in the transducer. Based on Eq. (A11) with a_0 =0.75 cm, the full width at half maximum (FWHM) of D^2, i.e., the beamwidth of one transducer, is 2°, and of $D_1^2D_2^2$ is 1.4°. This angular FWHM corresponds to an arc length of 1.1 cm at 45 cm range, which is much less than the 6 cm ($2\sigma_{J0}$) characteristic width of the jet (see Fig. 1).

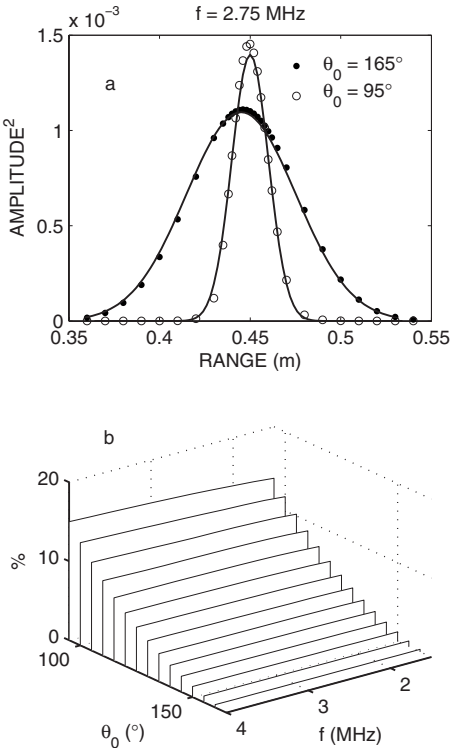


FIG. 11. (a) Theoretical profiles of mean-square scattered signal level at scattering angles of 95° and 165° assuming isotropic, frequency-independent scatterers suspended in the jet. Points are the predictions from Eq. (A10) with $\sigma_{J0}=3$ cm in Eq. (A9). Solid curves are the Gaussian fits to these points, after convolution with a constant-amplitude transmit pulse of length $c\tau/2=1.2$ cm. (b) Theoretical scattering pattern for isotropic, frequency-independent scatterers in the jet, normalized by the values at $\theta_0=165°$.

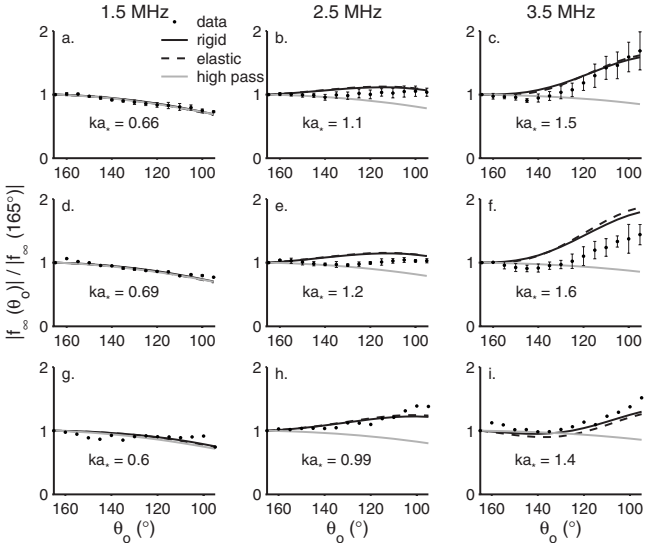


FIG. 12. Measured and predicted scattered amplitudes at three frequencies for suspensions of [(a)–(c)] small lead-glass beads, [(d)–(f)] small sieved sand, and [(g)–(i)] small unsieved sand, normalized to $\theta_0=165°$. Error bars indicate ± the standard error for the small lead-glass beads and small sieved sand. Only one experiment was run with the unsieved sand, so no error estimates are available for this case.

The resulting theoretical profiles of the mean square scattered signal levels from Eq. (A10) are plotted in Fig. 11(a) for scattering angles of 165° and 95°. D was computed using Eq. (A11) with $a_0=0.75$ cm. The integration was carried out for values of x and y ranging between $\pm 3\sigma_{J0}$. The solid lines are the Gaussian fits to the predicted points, after convolution with a boxcar function of length equal to the 1.2 cm pulse length (i.e., $c\tau/2$, τ being the 16 μs pulse duration). The pulse length is short enough that the convolution has a noticeable effect only for the narrow profile at 95°. Comparing Fig. 11(a) with Fig. 2, the predicted reduction in apparent jet width with decreasing scattering angle (from 3.0 to 0.97 cm) is nearly identical to that observed (from 3.1 cm to 0.96 cm). As indicated earlier, this reduction in apparent width is due physically to the reduced overlap area between the transmit and receive beams as the scattering angle approaches 90°.

Figure 11(b) shows the scattering pattern obtained using Eq. (A10), normalized by the values at $\theta_0=165°$. Rather than a scattering amplitude independent of scattering angle, as would be expected for isotropic scatterers, the theory predicts a geometry-induced bias in the observed scattering patterns. The value of γ^2 represented by the bias is 0.0049, small compared to those in Table IV. Because the bias is relatively small ($\lesssim 15\%$), no correction has been applied to the data.

Finally, the scattering angle clearly changes with position in the jet. This variation, $\Delta\theta_0$, was computed using the above equations for $r_1+r_2=90$ cm: i.e., for the ellipsoidal surface tangent to the jet centerline. For points with $10\log(D_1^2D_2^2)>-6$ dB, the computed values of $\Delta\theta_0$ were 0.1° or less for all measured frequencies and scattering angles. Thus, departures from θ_0 due to the finite size of the detected volume should have been negligible.

APPENDIX B: SMALL SCATTERERS REVISITED

The plots equivalent to those in Fig. 8, but for the ~200 μm diameter particles, are shown in Fig. 12. There are three main points to note about Fig. 12 in comparison to Fig. 8.

The first is that there is very little difference between the predictions of the elastic and rigid models. This is expected because, for $ka_*\lesssim 1.6$, the scattered amplitude is dominated by the $n=0$ and $n=1$ partial waves, and the amplitudes of these waves are relatively insensitive to the shear wave speed, at least for particles with the density of quartz or lead-glass suspended in water.

The second point to note is that the modified high-pass model, while agreeing with the data at low frequencies, provides progressively worse agreement as frequency increases. The good agreement at 1.5 MHz is expected, since $ka_*<1$, and the modified high-pass model is exact in the Rayleigh region. The relatively poor agreement with the high-pass model at 3.5 MHz is consistent with the results for the large particles in Fig. 8.

The third point of note is the relatively good agreement between the data and spherical scatterer theory in eight of the nine panels. The exception is panel f, the small sand at 3.5 MHz: the elastic and/or rigid models exhibit trends similar to the data, but the data are systematically lower. This

disagreement is consistent with the best-fit value of γ^2 being larger for the small sand (Fig. 7). In contrast, the unsieved small sand agrees comparatively well with the model predictions, suggesting that spherical scatter theory ought to provide a good fit to the small sieved sand data. In addition, much better agreement between theory and experiment is exhibited by the large sand data at the same value of ka_* [Fig. 8(d)]. Thus, the only explanation the authors have for the relatively poor quantitative agreement between theory and the sieved small sand data is that these data are in error, but we have been unable as yet to identify the cause.

[1]P. D. Thorne and D. M. Hanes, "A review of acoustic measurement of small-scale sediment processes," Cont. Shelf Res. **22**, 603–632 (2002).
[2]J. Sheng and A. E. Hay, "An examination of the spherical scatterer approximation in aqueous suspensions of sand," J. Acoust. Soc. Am. **83**, 598–610 (1988).
[3]A. S. Schaafsma and A. E. Hay, "Attenuation in suspensions of irregularly shaped sediment particles: A two-parameter equivalent spherical scatterer model," J. Acoust. Soc. Am. **102**, 1485–1502 (1997).
[4]P. D. Thorne and M. J. Buckingham, "Measurements of scattering by suspensions of irregularly shaped sand particles and comparison with a single parameter modified sphere model," J. Acoust. Soc. Am. **116**, 2876–2889 (2004).
[5]V. K. Varadan, Y. Ma, and V. V. Varadan, "Scattering and attenuation of elastic waves in random media," Pure Appl. Geophys. **131**, 577–603 (1989).
[6]J. Ribberink, "Bed-load transport for steady flows and unsteady oscillatory flows," Coastal Eng. **34**, 59–82 (1998).
[7]A. E. Hay and D. G. Mercer, "On the theory of sound scattering and viscous absorption in aqueous suspensions at medium and short wavelengths," J. Acoust. Soc. Am. **78**, 1761–1771 (1985).
[8]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill Book, New York, 1968), p. 927.
[9]Lord Rayleigh, *Theory of Sound*, 2nd ed. (Dover, New York, 1945), Vol. **2**, p. 504.
[10]J. J. Faran, "Sound scattering by solid cylinders and spheres," J. Acoust. Soc. Am. **23**, 405–418 (1951).
[11]A. E. Hay and R. W. Burling, "On sound scattering and attenuation in suspensions, with marine applications," J. Acoust. Soc. Am. **72**, 950–959 (1982).
[12]R. Hickling, "Analysis of echoes from a solid elastic sphere in water," J. Acoust. Soc. Am. **34**, 1582–1592 (1962).
[13]K. G. Foote, "Optimizing copper spheres for precision calibration of hydroacoustic equipment," J. Acoust. Soc. Am. **71**, 742–747 (1982).
[14]A. E. Hay and A. S. Schaafsma, "Resonance scattering in suspensions," J. Acoust. Soc. Am. **85**, 1124–1138 (1989).
[15]G. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge University Press, New York, 1967), p. 615.
[16]A. E. Hay, "Sound scattering from a particle-laden, turbulent jet," J. Acoust. Soc. Am. **90**, 2055–2074 (1991).
[17]S. A. Moore, "The angular dependence of acoustic scattering from suspended sediment," MS thesis, Dalhousie University (2008), Nova Scotia, Canada.
[18]H. B. Fischer, E. J. List, R. C. Koh, J. Imberger, and N. H. Brooks, *Mixing in Coastal and Inland Waters* (Academic, New York, 1979), p. 472.
[19]V. V. Kharin and F. W. Zwiers, "Climate predictions with multimodel ensembles," J. Clim. **15**, 793–799 (2002).
[20]R. L. Ingram, "Sieve analysis," in *Procedures in Sedimentary Petrology*, edited by R. E. Carver (Wiley-Interscience, Toronto, 1971).
[21]T. Allen, *Powder Sampling and Particle Size Determination* (Elsevier, New York, 2003), p. 627.
[22]J. P. M. Syvitski, K. W. G. LeBlanc, and K. W. Asprey, "Interlaboratory, interinstrument calibration experiment," in *Principles, Methods, and Application of Particle Size Analysis*, edited by J. P. M. Syvitski (Cambridge University Press, Cambridge, 1991), pp. 174–193.
[23]R. K. Johnson, "Sound scattering from a fluid sphere revisited," J. Acoust. Soc. Am. **61**, 375–377 (1977).
[24]W. Franz and K. Deppermann, "Theorie der beugung am zylinder unter berücksichtigung der kriechwelle (Theory of diffraction by a cylinder accounting for creeping waves)," Ann. Phys. **445**, 361–373 (1952).
[25]W. G. Neubauer, "Observation of acoustic radiation from plane and curved surfaces," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston, (Academic, New York, 1973), Vol. **10**, pp. 174–193.
[26]H. Überall, "Surface waves in acoustics," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1973), Vol. **10**, pp. 1–60.
[27]I. N. McCave and J. P. M. Syvitski, "Principles and methods of geological particle size analysis," in *Principles, Methods, and Application of Particle Size Analysis*, edited by J. P. M. Syvitski (Cambridge University Press, Cambridge, 1991), pp. 174–193.
[28]C. S. Clay and H. Medwin, *Acoustical Oceanography: Principles and Applications* (Wiley-Interscience, Toronto, 1977).
[29]D. T. DiPerna and T. K. Stanton, "Fresnel zone effects in the scattering of sound by cylinders of various lengths," J. Acoust. Soc. Am. **90**, 3348–3355 (1991).
[30]J. Sheng and A. E. Hay, "Spherical wave backscatter from straight cylinders: Thin-wire standard targets," J. Acoust. Soc. Am. **94**, 2756–2765 (1993).

High resolution population density imaging of random scatterers with the matched filtered scattered field variance

Mark Andrews, Zheng Gong, and Purnima Ratilal
Department of Electrical and Computer Engineering, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 11 February 2009; revised 16 June 2009; accepted 17 June 2009)

The matched filter enables imaging with high spatial resolution and high signal-to-noise ratio by coherent correlation with the expected field from what is assumed to be a discrete scatterer. In many physical imaging systems, however, returns from a large number of randomized scatterers, ranging from thousands to millions of individuals, are received together and the coherent or expected field vanishes. Despite this, it is shown that cross-spectral coherence in the matched filtered variance retains a pulse compression property that enables high-resolution imaging of scatterer population density. Analytic expressions for the statistical moments of the broadband matched filtered scattered field are derived in terms of the medium's Green's function, object scatter function, and spatial distribution using a single-scatter approximation. The formulation can account for potential dispersion in the medium and target over the signal bandwidth, and can be used to compare the relative levels of the coherent and incoherent scattered intensities. The analytic model is applied to investigate population density imaging of fish distributions in the Gulf of Maine with an ultrasonic echosounder. The results are verified with numerical Monte-Carlo simulations that include multiple scattering, illustrating that the single-scatter approximation is valid even for relatively dense Atlantic herring (*Clupea harengus*) schools.
[DOI: 10.1121/1.3177271]

PACS number(s): 43.30.Pc [RCG] Pages: 1057–1068

I. INTRODUCTION

The matched filter is applied in many remote sensing systems to provide high-resolution imaging of groups of objects or organisms. In underwater sonar applications, scattered returns are matched filtered to determine the population density of fish and other marine organism,[1,2] localize underwater vehicles,[3] and image seafloor and sub-bottom geomorphology.[4] The matched filter has also been used to image migrating bird populations with radar for ecological surveys.[5] Matched filter theory for maximizing signal-to-noise ratio and improving spatial resolution in the detection and localization of a discrete target is well established.[6–8] It is a coherent process implemented by correlating the target generated or scattered field with a replica waveform, which for an active system is usually the transmitted source signal. This pulse compression leads to a range resolution $\Delta r = c/2B$, where c is the wave speed and B is the signal bandwidth. In contrast, without pulse compression, range resolution $\Delta r = c\tau/2$ is determined by signal duration τ and requires short duration pulses to achieve the same resolution. The matched filter then allows long duration broadband waveforms to be transmitted with sufficient energy to illuminate a target above background noise and still maintain high spatial resolution.[6–8]

When many random scatterers are present, such as an independently moving group of objects, it is typically assumed that the number of scatterers in a given resolution cell is proportional to the intensity of that cell.[1,2,5] This implies a summation of the fields from each scatterer, where upon averaging multiple measurements of the scattered returns from the randomized group, the expected field or coherent intensity is negligible and the expected intensity is dominated by the variance.[9–12] It is also typically assumed that the resolution cell in range for incoherent returns after matched filtering is defined by the source signal's autocorrelation function.[1,2,5,7,8,13] Analytic solutions for the statistically coherent and incoherent intensities scattered from randomized groups have been formulated in the literature[9–12] but have generally not included the matched filter.

Here the authors show that the incoherently scattered intensity maintains spatial resolution through *cross-spectral coherence* in the matched filtered variance. A full field theory is presented for the statistical moments of the broadband matched filtered field simultaneously scattered from a random distribution of scatterers when the single scattering approximation is valid. The moments depend on the characteristic function of the scatterer's spatial distribution, which may extend over multiple range resolution cells of the imaging system. The moments are the output of a filtering process involving the characteristic function, source spectrum, scatter function, and Green's function. For the mean field, the characteristic function is filtered over a bandwidth centered at the *carrier frequency*. This leads to oscillatory contributions from cross-spectral products in the coherent intensity that are insignificant, on average, over the large spatial extent of the scatterers. On the other hand, the variance of the field depends on the characteristic function filtered at the probing signal's *baseband*, leading to a stable output upon averaging. The variance is shown to resample the scatterer distribution at a spatial rate inversely proportional to the signal bandwidth. Since the scatter function and Green's function are

employed, the model can account for dispersion caused by the target or medium.

The analytic model is applied to the physical example of imaging a school of Atlantic herring in the water column with an ultrasonic echosounder. High-resolution population density imaging of scatterer distributions can be achieved using the matched filtered variance since it is directly proportional to the mean number of scatterers within each resolution cell. The results are verified using numerical Monte-Carlo simulations of the matched filtered scattered field statistics that include multiple scattering. The single-scattering approximation is shown to be valid for imaging Atlantic herring schools exhibiting relatively high volumetric densities near 1 fish/m^3 observed in the Gulf of Maine with an echosounder.

An advantage of the analytic and numerical models developed here is that scattering from the entire distribution extending over multiple range resolution cells of the imaging system, within the sonar beam, can be simultaneously analyzed. By applying the matched filter theory from first principles, scatterers over the entire distribution are automatically localized in range, thereby avoiding the need to artificially break-up the scatterers in the distribution to within each range resolution cell. We provide conditions for when the incoherent scattering assumption is valid for the scatterer distribution and when coherent effects can be neglected. We show that it depends on the distribution size, shape, and scatterer density.

II. ANALYTIC MODEL FOR STATISTICAL MOMENTS OF MATCHED FILTERED SCATTERED FIELD FROM A GROUP OF RANDOM SCATTERERS

Here we formulate the theory for an active bistatic imaging system, comprised of a source located at $\mathbf{r}_0$, and a receiving array located at the origin of the coordinate system, imaging a group of scatterers centered at location $\mathbf{r}$ in the far-field of both the source and receiver. The formulation assumes that the receiver is not in the forward scatter direction, so that the scattered field at the receiver can be separated from the incident field. The position of any scatterer in the group is $\mathbf{r}_p=\mathbf{r}+\mathbf{u}_p$, where $\mathbf{u}_p$ is its displacement from the group center. The source transmits a broadband waveform $q(t)$ with Fourier transform $Q(f)$ and bandwidth B. For imaging systems used when only single scattering from each object is significant, the time-harmonic scattered field received from the group can be expressed by summing the contribution from each scatterer,[9,10,12,14,15]

$$\Phi_s(\mathbf{r},f)=Q(f)\sum_p W(\mathbf{r_p})W_0(\mathbf{r_p})G(\mathbf{r_p}|\mathbf{r_0},f)G(\mathbf{0}|\mathbf{r_p},f)\times\frac{S_p(\Omega_i,\Omega,k)}{k}, \tag{1}$$

where $W(\mathbf{r_p})$ and $W_0(\mathbf{r_p})$ are the beampatterns of the imaging system source and receiver, respectively, weighting the contribution from the pth scatterer, G is the medium's Green's function, and $S_p(\Omega_i,\Omega,k)$ is the scatter function, which depends on the wavenumber $k=2\pi f/c$, incident angle from the source Ω_i, and scattered angle in the direction of the receiver Ω. The individual scatterers in the group can be of arbitrary size compared to the wavelength. Here, the group center is assumed to coincide with the main response axis of the receiving array such that the beampattern may be approximated as $W(\mathbf{r_p})W_0(\mathbf{r_p})\approx 1$ for scatterers within the main lobe of the array, and $W(\mathbf{r_p})W_0(\mathbf{r_p})\approx 0$ for scatterers outside of this lobe. Therefore the sum in Eq. (1) is restricted to N, the total number of scatterers imaged within the main lobe of the imaging system, over the full range extent of the group's distribution. Let $\mathbf{k_i}$ and $\mathbf{k_s}$ be the incident and scattered wave vectors for scatterers within the mainlobe in Eq. (1). Approximating the spherical spreading loss, Green's function to each scatterer can be simplied as

$$G(\mathbf{r_p}|\mathbf{r_0},f)=\frac{e^{ik|\mathbf{r_p}-\mathbf{r_0}|}}{4\pi|\mathbf{r_p}-\mathbf{r_0}|}\approx G(\mathbf{r}|\mathbf{r_0};f)e^{i\mathbf{k_i}\cdot\mathbf{u_p}}, \tag{2}$$

where Green's function from the source to the group center is factored out since the scatterer distribution is in the far-field of both the source and receiver. Using similar approximations for the Green's functions from scatterers to receiver, the scattered field from Eq. (1) now simplifies to

$$\Phi_s(\mathbf{r},f)=Q(f)G(\mathbf{r}|\mathbf{r_0},f)G(\mathbf{0}|\mathbf{r},f)\sum_{p=1}^{N} e^{j(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}}\frac{S_p(\Omega_i,\Omega,k)}{k}. \tag{3}$$

The matched filter is now applied,[6–8,16] which is typically a normalized replica of the original transmitted waveform expressed as

$$H(f|t_M)=\frac{1}{\sqrt{E_0}}Q^*(f)e^{i2\pi f t_M}, \tag{4}$$

where t_M is the time delay of the matched filter and $E_0=\int|Q(f)|^2df$ is the source energy. Using Fourier synthesis, the time-dependent matched filtered scattered signal from Eq. (3) is

$$\Psi_s(t_M)=\int\frac{1}{k}\sum_{p=1}^{N} e^{i(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}}S_p(\Omega_i,\Omega,k)\Xi(f)df, \tag{5}$$

where

$$\Xi(f)=\frac{1}{\sqrt{E_0}}|Q(f)|^2G(\mathbf{r}|\mathbf{r_0},f)G(\mathbf{0}|\mathbf{r},f)e^{-i2\pi f(t-t_M)}. \tag{6}$$

The matched filtered signal in Eq. (5) depends on the phase contribution from each scatterer, and is randomized by N, the total number of scatterers imaged within the sonar beam, $\mathbf{u_p}$ and S_p, the location and scatter function for each scatterer, respectively. In general, the position $\mathbf{u_p}$ can be a function of time.

The expected intensity of the matched filtered scattered returns is a sum of the mean field squared $|\langle\Psi_s(t_M)\rangle|^2$, or coherently scattered intensity, and the variance $\mathrm{var}(\Psi_s(t_M))$, or incoherently scattered intensity. If we assume the random variables N, $\mathbf{u_p}$, and S_p are mutually uncorrelated and the scatterers are identically distributed, then the mean squared matched filtered scattered field becomes

$$|\langle\Psi_s(t_M)\rangle|^2 = \left| \int \frac{1}{k} \left\langle \sum_{p=1}^{N} e^{j(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}} S_p(\Omega_i,\Omega,k) \right\rangle \Xi(f) df \right|^2$$
$$= \langle N\rangle^2 \left| \int \frac{1}{k} U(\mathbf{k_i}-\mathbf{k_s})\ S(\Omega_i,\Omega,k)\rangle \Xi(f) df \right|^2. \quad (7)$$

Here, $U(\boldsymbol{\kappa})$ is the characteristic function,[14,17]

$$U(\boldsymbol{\kappa}) = \langle e^{i\boldsymbol{\kappa}\cdot\mathbf{u}}\rangle = \int e^{i\boldsymbol{\kappa}\cdot\mathbf{u}} p_{\mathbf{u}}(\mathbf{u}) d\mathbf{u}, \quad (8)$$

which is a three-dimensional spatial Fourier transform of probability density function (PDF) $p_{\mathbf{u}}(\mathbf{u})$ of scatterer position. The authors assume that the expectation is taken over a measurement time where the distribution is statistically stationary. The coherent intensity in Eq. (7) depends on the expected scatter function, $\langle S(\Omega_i,\Omega,k)\rangle$, which in general depends on the incident and scattered angles, and the statistics of the shape, size, material properties, and orientation of the scatterers in the group.

The second moment of the matched filtered scattered field from Eq. (5),

$$\langle|\Psi_s(t_M)|^2\rangle = \int\int \frac{1}{kk'} \left\langle \sum_{p=1}^{N}\sum_{q=1}^{N} e^{j(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}} e^{-j(\mathbf{k_i'}-\mathbf{k_s'})\cdot\mathbf{u_q}} \times S_p(\Omega_i,\Omega,k) S_q^*(\Omega_i,\Omega,k') \right\rangle \Xi(f)\Xi^*(f') df df', \quad (9)$$

depends on the correlation between two different particle's phase contributions and scatter functions. To evaluate Eq. (9), the joint probability distribution function for the scatterers' positions and scatter functions must in general be specified. Here, it is assumed that the scatter functions are statistically independent from each other and from their relative position among the group. Also, the scatterers' relative positions and phase contributions are assumed to be uncorrelated with the majority of the other scatterers within the imaging system resolution footprint. For many groups in nature, such as fish schools, bird flocks, and insect swarms, each scatterer's position depends only on adjacent and nearby scatterers within some small radius, r_c, beyond which their positions can be considered independent,[18–20] such that

$$\lim_{|\mathbf{u_p}-\mathbf{u_q}|\gg r_c} p(\mathbf{u_p},\mathbf{u_q}) = p(\mathbf{u_p})p(\mathbf{u_q}). \quad (10)$$

The correlation between two particle's positions is typically formulated in terms of the pair distribution function, g,[9] defined by

$$p(\mathbf{u_p},\mathbf{u_q}) = \frac{g(\mathbf{u_p}-\mathbf{u_q})}{V^2}, \quad (11)$$

where V is the group volume. For scatterers separated by distances greater than r_c, their positions can be considered independent, as g approaches unity. When the imaging system beamwidth is much larger than r_c, the contribution to the total scattered intensity from correlated scatterers is negligible. Under these conditions, the expectation over the scatter functions and phase contributions simplify

$$\left\langle \sum_{p=1}^{N}\sum_{q=1}^{N} e^{j[(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}-(\mathbf{k_i'}-\mathbf{k_s'})\cdot\mathbf{u_q}]} S_p(\Omega_i,\Omega,k) S_q^*(\Omega_i,\Omega,k') \right\rangle$$
$$= \int \sum_{p=1}^{N}\sum_{q=1}^{N} ([\langle e^{j[(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}-(\mathbf{k_i'}-\mathbf{k_s'})\cdot\mathbf{u_q}]}\rangle\langle S_p(\Omega_i,\Omega,k) S_q^*(\Omega_i,\psi,k')\rangle - \langle e^{j(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}}\rangle\langle e^{-j(\mathbf{k_i'}-\mathbf{k_s'})\cdot\mathbf{u_p}}\rangle\langle S_p(\Omega_i,\Omega,k)\rangle\langle S_q^*(\Omega_i,\Omega,k')\rangle]\delta_{pq}$$
$$+ \langle e^{j(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}}\rangle\langle e^{j(\mathbf{k_i'}-\mathbf{k_s'})\cdot\mathbf{u_q}}\rangle\langle S_p(\Omega_i,\Omega,k)\rangle\langle S_q^*(\Omega_i,\Omega,k')\rangle) p(N) dN$$
$$= \langle N\rangle[U((\mathbf{k_i}-\mathbf{k_s})-(\mathbf{k_i'}-\mathbf{k_s'}))\langle S(\Omega_i,\Omega,k) S^*(\Omega_i,\Omega,k')\rangle - U(\mathbf{k_i}-\mathbf{k_s})U^*(\mathbf{k_i'}-\mathbf{k_s'})\langle S(\Omega_i,\Omega,k)\rangle\langle S^*(\Omega_i,\Omega,k')\rangle]$$
$$+ \langle N^2\rangle U(\mathbf{k_i}-\mathbf{k_s})U^*(\mathbf{k_i'}-\mathbf{k_s'})\langle S(\Omega_i,\Omega,k)\rangle\langle S^*(\Omega_i,\Omega,k')\rangle, \quad (12)$$

where δ_{pq} is the Kronecker delta and $p(N)$ is the probability of finding N scatterers in the group. Inserting Eq. (12) into Eq. (9) results in the following expression for the second moment,

$$\langle|\Psi_s(t_M)|^2\rangle = \int\int \frac{1}{kk'} (\langle N\rangle[U((\mathbf{k_i}-\mathbf{k_s})-(\mathbf{k_i'}-\mathbf{k_s'}))\langle S(\Omega_i,\Omega,k) S^*(\Omega_i,\Omega,k')\rangle - U(\mathbf{k_i}-\mathbf{k_s})U^*(\mathbf{k_i'}-\mathbf{k_s'})\langle S(\Omega_i,\Omega,k)\rangle$$
$$\times\langle S^*(\Omega_i,\Omega,k')\rangle] + \langle N^2\rangle U(\mathbf{k_i}-\mathbf{k_s})U^*(\mathbf{k_i'}-\mathbf{k_s'})\langle S(\Omega_i,\Omega,k)\rangle\langle S^*(\Omega_i,\Omega,k')\rangle)\Xi(f)\Xi^*(f') df df'. \quad (13)$$

The second moment can be expressed as the sum of the coherent intensity in Eq. (7) and the incoherent intensity given by the variance of the field,

$$\mathrm{var}(\Psi_s(t_M)) = \langle|\Psi_s(t_M)|^2\rangle - |\langle\Psi_s(t_M)\rangle|^2 = \langle N\rangle \int\int \frac{1}{kk'}[U((\mathbf{k_i}-\mathbf{k_s})-(\mathbf{k_i'}-\mathbf{k_s'}))\langle S(\Omega_i,\Omega,k) S^*(\Omega_i,\Omega,k')\rangle$$
$$- U(\mathbf{k_i}-\mathbf{k_s})U^*(\mathbf{k_i'}-\mathbf{k_s'})\langle S(\Omega_i,\Omega,k)\rangle\langle S^*(\Omega_i,\Omega,k')\rangle]\Xi(f)\Xi^*(f') df df' + \frac{\mathrm{var}(N)}{\langle N\rangle^2}|\langle\Psi_s(t_M)\rangle|^2. \quad (14)$$

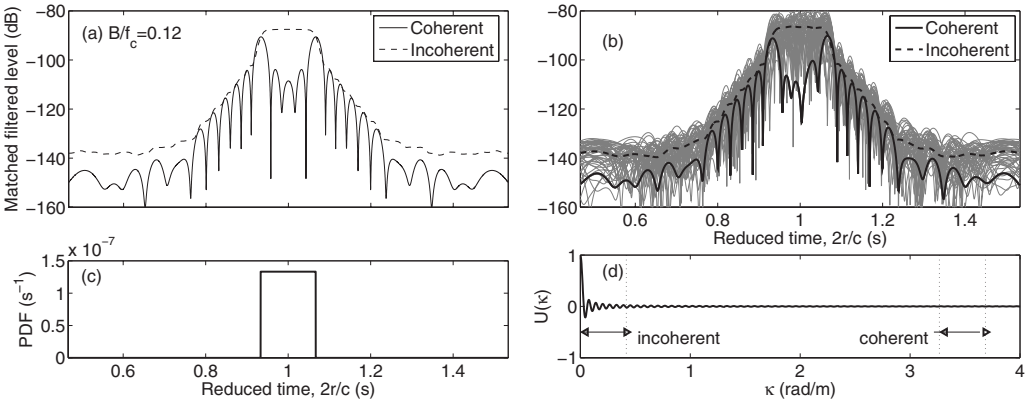


FIG. 1. Broadband matched filtered coherent and incoherent intensity levels scattered from a group of discrete scatterers following a statistically uniform spatial distribution calculated using (a) the analytic model and verified by (b) Monte-Carlo simulations. The spatial PDF for the scatterer distribution is shown in (c) and its wavenumber spectrum in (d). The wavenumber region evaluated by the analytic model coherent intensity corresponding to the source frequency band and the incoherent intensity using the baseband are indicated.

The coherent intensity in Eq. (7) is proportional to $\langle N\rangle^2$. If the standard deviation in the number of scatterers is small compared to its mean, then the incoherent intensity in Eq. (14) is proportional to $\langle N\rangle$. This is important for population density imaging, since population estimation from the matched filtered intensity depends on whether scattering from a group is coherent or incoherent.

An expression is provided for the incoherent matched filtered intensity in the time domain in Refs. 8 and 13 as the output of a convolution between the magnitude square autocorrelation function of the source signal with the mean backscatter cross-section spatial distribution of the scatterer. This causes complications when there is dispersion in the medium or target since the pulse shape may be altered leading to changes in the range resolution. Furthermore, the incoherent intensity expression in Ref. 8 is only valid when the total number of scatterers in each resolution cell is a constant over time or when the coherent intensity is significantly smaller than the variance, as can be seen by comparison with Eq. (14) in this paper. The matched filtered intensity scattered from a continuum is derived in Eq. (13) of Ref. 21 for the ocean seafloor reverberation. However, the final expression for the expected matched filtered intensity, Eq. (14) of Ref. 21, does not retain the time-frequency exponential dependence. This term is essential, along with the wavenumber dependent term for describing how the variance retains high spatial resolution, as can be seen by comparison with Eq. (14) in this paper.

In practical imaging systems, the matched filtered scattered signal and its intensity as a function of time are charted to range by multiplying with the propagation speed of the signal and accounting for the distance travelled from source to scatterer and from scatterer to receiver.

III. ILLUSTRATIVE EXAMPLES

In this section, the analytic model is demonstrated with several canonical spatial distributions of scatterer groups. First, the analytic model results for the matched filtered scattered field statistics are verified with numerical Monte-Carlo simulation in Sec. III A. The effect of scatterer distribution size and shape on the relative levels of coherent and incoherent scattered intensities is examined in Sec. III B. Next, in Sec. III C, the effect of imaging system bandwidth on the spatial resolution of the matched filtered variance is investigated, and it is shown that population density of the scatterers can be inferred from the variance in Sec. III D. Finally, in Sec. IV, the analytic model is applied to the physical example of imaging over depth a school of Atlantic herring, with a relatively large volumetric density of 1 fish/m^3 using an ultrasonic echosounder. The analytic model results are verified with numerical Monte-Carlo simulations that include multiple scattering.

For simplicity, the authors consider only monostatic imaging systems such that $(\mathbf{k_i}-\mathbf{k_s})\cdot\mathbf{u_p}\approx 2kx_p$, so that imaging is achieved with mainly backscattered fields. This simplifies the characteristic function in Eq. (8) to an integration only over range, in this case the x-direction. In all examples, the scatterer groups are located within the main lobe of the receiver, to investigate only the effects of the matched filtered returns in time.

A. Verifying analytic model results with numerical Monte-Carlo simulations for matched filtered field statistics

The analytically determined coherent and incoherent matched filtered scattered intensities are shown in Fig. 1(a) for a group of discrete scatterers centered at $2r/c=1$ s whose spatial distribution follow the uniform PDF in Fig. 1(c). The intensities are calculated directly from Eqs. (7) and (14). The incoherent intensity dominates the scattered field over the entire range extent of the group except for the edges of the distribution where the coherent intensity is non-negligible. For this example, the temporal extent is $2L/c=2/15$ s, where L is the spatial extent of the distribution, and the mean spatial density per unit range for the scatterers is $\langle N\rangle/L=100/\mathrm{m}$. We set $E_0=1$, and $\langle|S/k|^2\rangle=1$ which is proportional to the scattering cross-section of an individual scatterer. For this example, the bandwidth is set to $B/f_c=0.12$, which is a typical bandwidth-to-carrier frequency ratio for many ocean acoustic imaging systems.[1,2,4] The corresponding matched filter temporal resolution $2\Delta r/c=1/B=20$ ms.

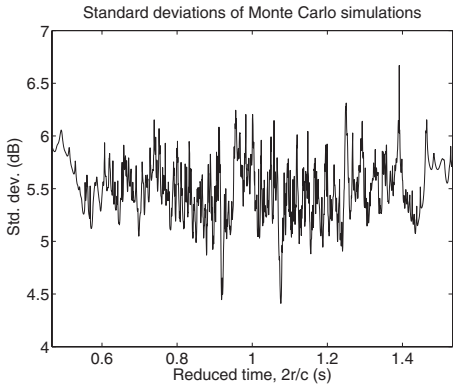


FIG. 2. The standard deviations for the Monte-Carlo simulations for the log transformed matched filtered intensity levels shown in Fig. 1(b).

The results in Fig. 1(a) are verified using Monte-Carlo simulation with 500 realizations in Fig. 1(b), where the locations of the scatterers for each realization are randomly drawn from the uniform spatial PDF of Fig. 1(c). The matched filtered scattered field for each realization is calculated using Eq. (5), and the coherent and incoherent intensities are computed directly as the mean square and variance of the numerical simulations. The analytic model results in Fig. 1(a) are in good agreement with the Monte-Carlo simulation results in Fig. 1(b).

For this distribution, the matched filtered signal behaves as a circular complex Gaussian random process, which implies that the broadband returns are statistically saturated. This can be predicted from Eq. (5) by randomizing the phase from each pth scatterer according to a uniform PDF. For the uniform PDF used here, the extent of the distribution is significantly larger than the wavelength so that the phase of the scattered returns are essentially also uniformly distributed.

The log transformed intensity realizations illustrated in gray in Fig. 1(b) have a standard deviation near 5.6 dB, as shown in Fig. 2. This indicates that the scattered broadband returns are statistically saturated, such that the intensity measurements are independent realizations of a circular complex Gaussian random field.[17,22,23] The fluctuations in the scattered intensity are entirely a result of the randomness of the scatterers' positions, which randomizes the phase in the scattered returns. This can also be predicted by applying the central limit theorem to Eq. (5), and noting that the phase from each pth scatterer follows a uniform PDF when the group's distribution is much larger than a wavelength. The example illustrated here is for a deterministic environment, where the incident intensity does not fluctuate. For an imaging scenario where the environment introduces its own fluctuations, the standard deviations of the scattered returns may differ. In practical imaging systems, the expected matched filtered intensity is measured by averaging over several matched filtered returns to reduce the variance in the estimation of the scatterer population.[17,22,23]

B. Effect of scatterer distribution size and shape on relative levels of coherent and incoherent matched filtered intensities

Here the analytic intensity results obtained for the scatterer group that follows the uniform spatial distribution in Fig. 1 are compared to another two groups that follow Gaussian spatial distributions for the same imaging system. Figure 3 illustrates the coherent and incoherent scattered intensities for a group of scatterers following a Gaussian spatial PDF with extent (a) $\sigma_L=8\lambda$ and (b) $\sigma_L=\lambda/8$, respectively, where σ_L is the standard deviation of the group's distribution in range. The scatterer groups in Figs. 1 and 3(a) have the same number of scatterers as well as the same spatial standard deviation for their distributions in range. For the Gaussian distribution in Fig. 3(a), the incoherent intensity dominates over the entire spatial extent of the group unlike the uniform distribution case where the edge effects are nonnegligible. Comparing Figs. 3(a) and 3(b), the incoherent intensity dominates until the spatial extent of the distribution is reduced to below a wavelength where the coherent intensity then becomes dominant. The point where the coherent intensity overtakes the incoherent intensity depends largely on the number of scatterers, N, because the incoherent intensity, given by Eq. (14), is proportional to N, while the coherent intensity, given by Eq. (7), is proportional to N^2. For a given spatial distribution, larger numbers of scatterers increase the significance of the coherent intensity.

The characteristic functions $U(\kappa)$ for the uniform and the two Gaussian spatial distributions are shown in Figs. 1(d) and 3(c), respectively. For the uniform case, $U(\kappa)=\mathrm{sinc}(\kappa L/2)$, while the characteristic function for the Gaussian spatial distribution is also Gaussian. The coherent intensity in Eq. (7) filters the characteristic function at the carrier frequency where the larger scatterer groups have significantly smaller amplitudes. Furthermore the cross-spectral products $U(2k)U^*(2k')$ in the coherent intensity are highly oscillatory leading to cancellation except when the group's spatial extent is small compared to the wavelength or where

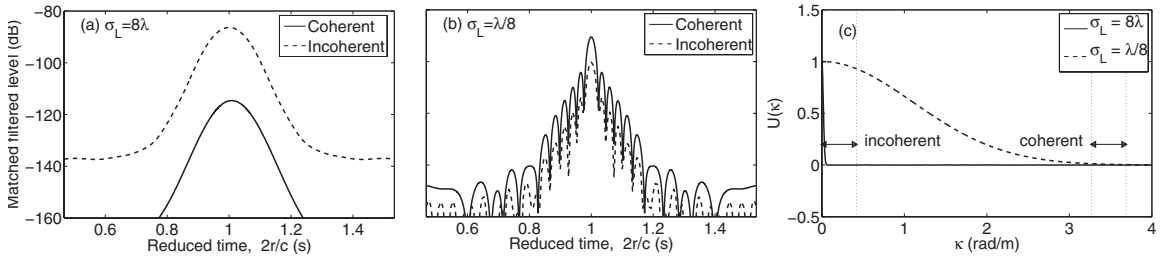


FIG. 3. Broadband matched filtered coherent and incoherent intensity levels scattered from a group of discrete scatterers following a Gaussian spatial PDF with group range extent defined by (a) $\sigma_L=8\lambda$ and (b) $\sigma_L=\lambda/8$. Their respective characteristic functions are illustrated in (c) where the wavenumber region evaluated by the analytic model coherent intensity corresponding to the source frequency band and the incoherent intensity using the baseband are indicated.

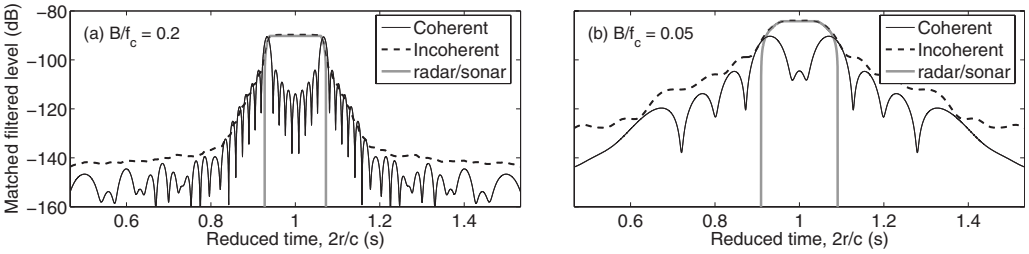


FIG. 4. Effect of varying signal bandwidth to (a) $B/f_c=0.2$ and (b) $B/f_c=0.05$, in comparison with Fig. 1(a). The radar/sonar approximation given by Eq. (15) is plotted over the scattered intensities for the uniform distribution of scatterers with PDF shown in Fig. 1(c).

it has sharply defined edges. In contrast, the incoherent intensity in Eq. (14) filters the characteristic function U $(2(k-k'))$ at the signal's baseband $(k-k')=2\pi(f-f')/c$, where the amplitude is significant for the larger groups. As a result of demodulation to baseband, the variance exhibits cross-spectral coherence that leads to a stable output upon averaging.

C. Effect of imaging system bandwidth on spatial resolution of matched filtered variance

The effect of the imaging system bandwidth on the matched filtered field statistics is investigated in Fig. 4 for the scatterer group that follows the uniform spatial PDF shown in Fig. 1(c). Comparing Fig. 1(a) with Fig. 4(a), increasing the relative signal bandwidth from $B/f_c=0.12$ to $B/f_c=0.2$ provides the matched filtered variance with a larger window in wavenumber to filter the characteristic function. From sampling theory, this allows the scatterer distribution to be sampled with higher spatial resolution $\Delta r=c/2B$ inversly proportional to the signal bandwidth. The variance reconstructs the spatial distribution, maintaining the higher wavenumber fluctuations. In this example, the reduced time resolution is improved from $2\Delta r/c=20$ ms to $2\Delta r/c=12$ ms. Decreasing the system bandwidth, as illustrated in Fig. 4(b), has the effect of sampling the distribution with a low pass filter leading to blurring of the distribution edges. Here, a bandwidth of $B/f_c=0.05$ leads to a reduced time resolution of $2\Delta r/c=48$ ms. While this example can be applied for various imaging systems and frequencies, here $f_c=415$ Hz and $c=1500$ m/s are used for spatial resolutions of $\Delta r=9$ m and $\Delta r=36$ m for (a) and (b) respectively.

These examples illustrate that the variance retains the fine spatial resolution expected of the matched filter, while this analysis here is for a monostatic system, similar analysis can be carried out for bistatic systems in other non-forward scatter directions, applying Eqs. (7) and (14) with the appropriate characteristic function. However, for bistatic systems, the spatial resolution in range depends not only on the temporal resolution of the matched filtered returns but also on the geometry of the problem.

D. Population density estimation using the incoherent matched filtered intensity

When incoherent scattering dominates, the log-transformed matched filtered variance can be approximated for each resolution cell of the imaging system as

$$L_p(\mathbf{r}_M) \approx L_s + \mathrm{TL}_b(\mathbf{r}_M|\mathbf{r_0}) + \mathrm{TL}_b(\mathbf{0}|\mathbf{r}_M) + \mathrm{TS}_b + 10\log_{10}(\langle M(\mathbf{r}_M|\Delta r)\rangle), \quad (15)$$

where $L_p(\mathbf{r}_M)=10\log_{10}\mathrm{var}(\Psi_s(t_M))$, $L_s=10\log_{10}E_0$, $\mathrm{TL}_b(\mathbf{r}_M|\mathbf{r_0})=10\log_{10}(1/E_0)\int|Q(f)|^2|G(\mathbf{r}_M|\mathbf{r_0},f)|^2df)$ is the broadband source spectrum weighted transmission loss, $\mathrm{TS}_b=10\log_{10}(1/E_0)\int|Q(f)|^2(\langle|S(f)|^2\rangle/k^2)df)$ is the broadband source spectrum weighted mean target strength, and $\langle M(\mathbf{r}_M|\Delta r)\rangle$ is the mean number of scatterers within the resolution footprint at $\mathbf{r}_M$. Equation (15) resembles the sonar/radar equation except that it uses the broadband spectrum weighted values. It is in a form that can readily be used to invert for population density imaging, solving for M across the image. Assuming a flat source spectrum across the bandwidth leads to the familiar sonar equation.

The log-transformed level, L_p from Eq. (15) is plotted in Fig. 4 using the population density directly. Here, we see the analytic model matches the results well in the region containing the scatterers, as shown in Fig. 4. This implies that the mean intensity obtained by incoherently averaging matched filtered data can be used to infer scatterer population density when the group's spatial distribution is statistically stationary over the averaging time period by correcting for source power, and spectrum weighted transmission losses and target strength.

IV. IMAGING DENSE SCHOOLS OF FISH WITH AN ULTRASONIC ECHOSOUNDER

The results of Sec. II and the illustrative examples in Sec. III depend on the single-scatter assumption, which is the basis for population density imaging by incoherent summation. The single scattering assumption also implies that attenuation through the group is insignificant. Here, the authors develop a numerical Monte-Carlo simulation model that includes multiple scattering for the statistics of the broadband matched filtered scattered field from a random group of scatterers. The model can be used to determine the conditions for when the single scattering approximation is valid. We compare the results of our analytic model with the single scatter assumption to those of the numerical Monte-Carlo simulation model that includes multiple scattering.

A. Numerical Monte-Carlo simulation model including multiple scattering for the statistics of the broadband matched filtered scattered field

Given a distribution of scatterers, we first generate the time harmonic singly and multiply scattered fields following the approach of Refs. 9 and 24 but generalized for arbitrarily large scatterers with angular dependent scatter functions. Given a source at $\mathbf{r_0}$, receiver at $\mathbf{r}$ and a group of N scatterers, where the position of each scatterer is known, the total scattered field at frequency f,

$$\Phi_s(\mathbf{r},f) = \sum_{p=1}^{N} \Phi_s(\mathbf{r};p,f), \tag{16}$$

is a sum of the scattered fields from each of the pth scatterers,

$$\Phi_s(\mathbf{r};p,f) = Q(f)W(\mathbf{r_p})G(\mathbf{r_p}|\mathbf{r_0},f)G(\mathbf{r}|\mathbf{r_p},f)\frac{S_p(\Omega_i,\Omega,k)}{k} + \sum_{q=1,q\neq p}^{N} \Phi_s(\mathbf{r_p};q,f)G(\mathbf{r}|\mathbf{r_p},f)\frac{S_p(\Omega_{qp},\Omega,k)}{k}, \tag{17}$$

expressed as a sum of the singly and multiply scattered fields. For cases where multiple scattering is negligible, Eqs. (16) and (17) can be combined to give Eq. (1). Here, the second term in Eq. (17) is the multiply scattered field at $\mathbf{r}$ from the pth scatterer, first scattered off all other $N-1$ scatterers, where $\Phi_s(\mathbf{r_p};q,f)$ is the scattered field from the qth scatterer incident on the pth scatterer.

In order to solve Eq. (17) for all N scatterers, the scattered field incident on each scatterer from all other $N-1$ scatterers must be found. This is expressed as

$$\Phi_s(\mathbf{r_n};p,f) = Q(f)W(\mathbf{r_p})G(\mathbf{r_p}|\mathbf{r_0},f)G(\mathbf{r_n}|\mathbf{r_p},f)\frac{S_p(\Omega_i,\Omega_{pn},k)}{k} + \sum_{q=1,q\neq p}^{N} \Phi_s(\mathbf{r_p};q,f)G(\mathbf{r_n}|\mathbf{r_p},f)\frac{S_p(\Omega_{qp},\Omega_{pn},k)}{k}, \tag{18}$$

where n and p are dummy variables indicating that the field is incident on n scattered from p. For a group of N scatterers, there are a total of $N(N-1)$ of these terms representing the field scattered from one particle onto another. Equation (18) has a similar formulation as Eq. (17) in terms of both singly and multiply scattered fields. The scatter functions are in general dependent on both the incident and scattered angles, where Ω_{qp} is the incident angle onto p scattered from q and Ω_{pn} is the scattered angle from p scattered onto n. All of the scattering terms represented by Eq. (18) can be grouped into an $N(N-1)$ column vector $\underline{\phi}$, and solved using an $N(N-1)$ by $N(N-1)$ matrix equation

$$\underline{\phi} = \underline{\phi_1} + \mathbf{A}\underline{\phi}, \tag{19}$$

where $\underline{\phi_1}$ is the $N(N-1)$ column vector with each element representing the singly scattered field from one scatterer onto another given by the first term in Eq. (18). The matrix $\mathbf{A}$ is defined such that the product $\mathbf{A}\underline{\phi}$ is an $N(N-1)$ column vector with each element representing the multiply scattered field from one scatterer onto another given by the second term in Eq. (18). The matrix, Eq. (19), can be rearranged and solved by inverting the $N(N-1)$ by $N(N-1)$ matrix,

$$\underline{\phi} = (\mathbf{I}-\mathbf{A})^{-1}\underline{\phi_1}. \tag{20}$$

However, for a large number N of scatterers, this matrix inversion may not be computationally feasible. Alternatively, it may be solved by first including only singly scattered fields, and then adding subsequent orders of scattering until converging on a stable solution using the recursive relation,

$$\underline{\phi_\mathbf{n}} = \underline{\phi_{\mathbf{n}-1}} + \mathbf{A}\underline{\phi_{\mathbf{n}-1}}, \tag{21}$$

where $\underline{\phi_\mathbf{n}}$ includes all orders of scattering up to the nth order. Together, Eqs. (16)–(19) describe the multiply scattered field off the scatterer group at frequency f at receiver $\mathbf{r}$ taking into account the angular dependence in the scatter function.

The time-dependent matched filtered multiply scattered signal from the scatterer group can now be simulated by calculating the time-harmonic scattered field at discrete frequency intervals over the signal bandwidth and applying Fourier synthesis,

$$\Psi_s(t_M) = \int \Phi_s(\mathbf{r},f)H(f|t_M)e^{-i2\pi ft}df, \tag{22}$$

where the matched filter $H(f|t_M)$ is defined in Eq. (4), and the time-harmonic scattered field is found using Eqs. (16)–(18). The statistics of the multiply scattered matched filtered returns from a group of scatterers following a stationary distribution are estimated using a sample of independent Monte-Carlo simulations. For each simulation, the positions and scatter functions of each scatterer are randomly drawn following their specified statistical distributions. The matched filtered signal scattered from the group is then calculated, and the coherent and incoherent intensities are estimated by the sample mean square and sample variance, respectively, over all of the realizations.

An advantage of applying the matched filter to model the multiply scattered fields from a distribution of scatterers using a broadband pulsed signal is that, under certain scenarios, the various orders of scattering can be distinguished through time delays in subsequent orders of scattering. A simple example of multiple scattering is demonstrated in Fig. 5, where the matched filtered returns from two discrete nonrandom scatterers are simulated using Eq. (22), by integrating over frequency the time harmonic scattered field at the receiver, given by Eqs. (16)–(18). The two scatterers illustrated in Fig. 5(a) are placed equidistant from a monostatic imaging system separated by d from each other at $(0,d/2)$ and $(0,-d/2)$. For illustrative purposes, the scatterers are given very high target strengths so that the multiple scattering effects stand out for only two scatterers. The imaging system transmits a broadband pulse and the scattered returns at the receiver are matched filtered and charted over range using the two-way travel time, $r=ct/2$. Figure 5(b) illustrates the matched filtered scattered returns for three cases: (1) single scattering only, (2) double and single scattering applying the iterative approach, and (3) all orders of scattering using the exact matrix solution. The single scattering pro-

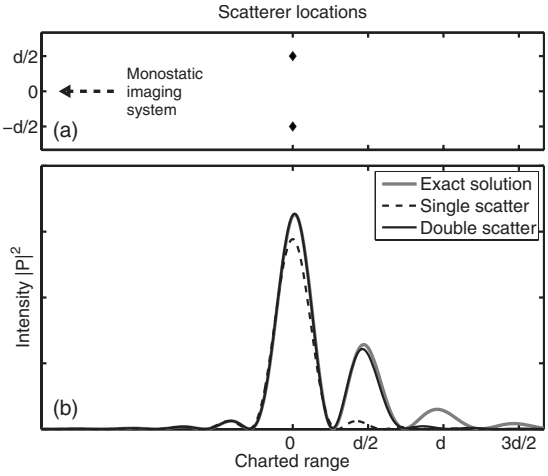


FIG. 5. Broadband matched filtered returns from two non-random discrete targets located at $(0,-d/2)$ and $(0,d/2)$ as shown in (a). The effects of multiple scattering can be seen in (b) by comparing the singly scattered to the multiply scattered matched filtered returns. This example was implemented with $d=60$ m, TS=30 dB, and an imaging system with resolution $\Delta r=15$ m.

duces a single peak, since the returns from each of the two scatterers are received together. Small sidelobes from the matched filtering process can be seen both before and after the primary peak but are less than 10% of the primary peak. The doubly scattered return exhibits a secondary delayed return due to the longer path length traversed by the second order multiply scattered waves. The main peak is also higher including multiple scattering because the sidelobe of the secondary peak coincides with the primary peak. Similarly, including higher orders of scattering produces noticeably delayed returns. The target strengths need not be as high as illustrated in this example to observe significant multiple scattering effects if the population density is increased. In general, as illustrated in this example, multiple scattering effects are characterized by a higher overall scattering level as well as additional delayed returns.

B. Application to ultrasonic echosounder imaging of a school of Atlantic herring in the Gulf of Maine

Here we examine population density imaging of a school of Atlantic herring as a function of water depth in the Gulf of Maine with a conventional fish-finding ultrasonic echosounder. The imaging system uses a broadband linear frequency modulated (LFM) waveform processed with a matched filter.

Many fish-finding echosounders use cw pulses for echolocation, where the pulse duration, T, determines the resolution in depth $\Delta r=cT/2$. The resulting signal bandwidth is inversely related to the time duration, $B=1/T$. Shorter duration pulses offer better depth resolution, however, suffer from reduced signal to noise ratio.[25] Alternatively, some echosounders transmit longer duration broadband LFM waveforms with sufficient energy to provide high signal-to-noise ratio, while still achieving high resolution by applying a matched filter to the scattered returns. In some cases, this method has been found to sufficiently resolve individual fish. For broadband systems, several issues must be addressed in order to accurately calibrate as well as image population denstiy. Both source level and beamwidth may vary over the signal bandwidth, changing the amplitude and ensonified volume as a function of frequency. For this reason, short duration cw systems have been more widely used in the fisheries community and have been extensively calibrated. Both types of echosounders systems offer distinct advantages.

Here, a group of herring are imaged with an EK60 Simrad echosounder, operating with a 38 kHz cw pulse, with a duration of 1 ms, a repetition rate of 1 s^{-1}, and bandwidth $B=1$ kHz, leading to a spatial resolution in depth of $\Delta r=0.75$ m for sound speed $c=1500$ m/s. The imaged fish group is used as a basis for our numerical model in order to simulate observed fish density profiles with a broadband system with a LFM chirp signal with bandwidth $B=3.75$ kHz. The actual fish density distributions may have fluctuations in depth that were not resolved by the cw system that could be observed by a broadband imaging system. However, the volumetric densities observed here are too high to resolve individual fish, even for most broadband systems. The analysis in this section simulates the echosounder imaging a relatively dense group of herring to show that (1) the incoherent intensity is dominant upon averaging, and (2) multiple scattering is negligible.

Shoaling Atlantic herring populations in the Gulf of Maine typically have volumetric densities near 0.05 fish/m^3.[26,27] On rare occasions, however, they have been known to cluster in relatively dense schools with volumetric densities of 1 fish/m^3 as was the case for the school observed here. Figure 6(a) shows an echosounder imagery of the herring school centered at 125 m depth in waters 200 m deep acquired at Georges Basin in the Gulf of Maine on September 22, 2006 using a Simrad EK60 echosounder operating at 38 kHz. The school is assumed to be Atlantic herring based on trawl surveys[26] conducted 4 days later, in the same area of the Gulf of Maine, just north of Georges Bank. The trawl surveys identified several fish schools consisting of more than 99% herring. Historical data[28] have also found Atlantic herring to be by far the most abundant schooling fish in the region.

The EK60 echosounder is hull-mounted and directed downward to provide depth profiles of the fish distributions. The echosounder transducer has a conical beamwidth of 7°, horizontal circular areal resolution of 183 m with diameter of 15.3 m at 125 m water depth. The data shown in Fig. 6(a) are in terms of volumetric backscatter strength (S_v) over depth, with profiles collected over time as the vessel traveled over the fish school at a nominal velocity of 3 m/s.

Volumetric backscatter strength, S_v, can be converted to population density, illustrated in Fig. 6(b) using an estimated mean target strength of −39.6 dB for individual herring.[29] The volumetric densities shown in Fig. 6(b) are for the region between the vertical solid lines in Fig. 6(a), where the density distributions are approximately stationary in time and space. In practice, the depth-dependent volumetric density profiles are incoherently averaged over range to reduce variance in the estimation of scatterer population.[17,22] The total estimated number of fish, N, integrated over the depth of the beam has a mean $\langle N\rangle=1950$ and standard deviation σ_N

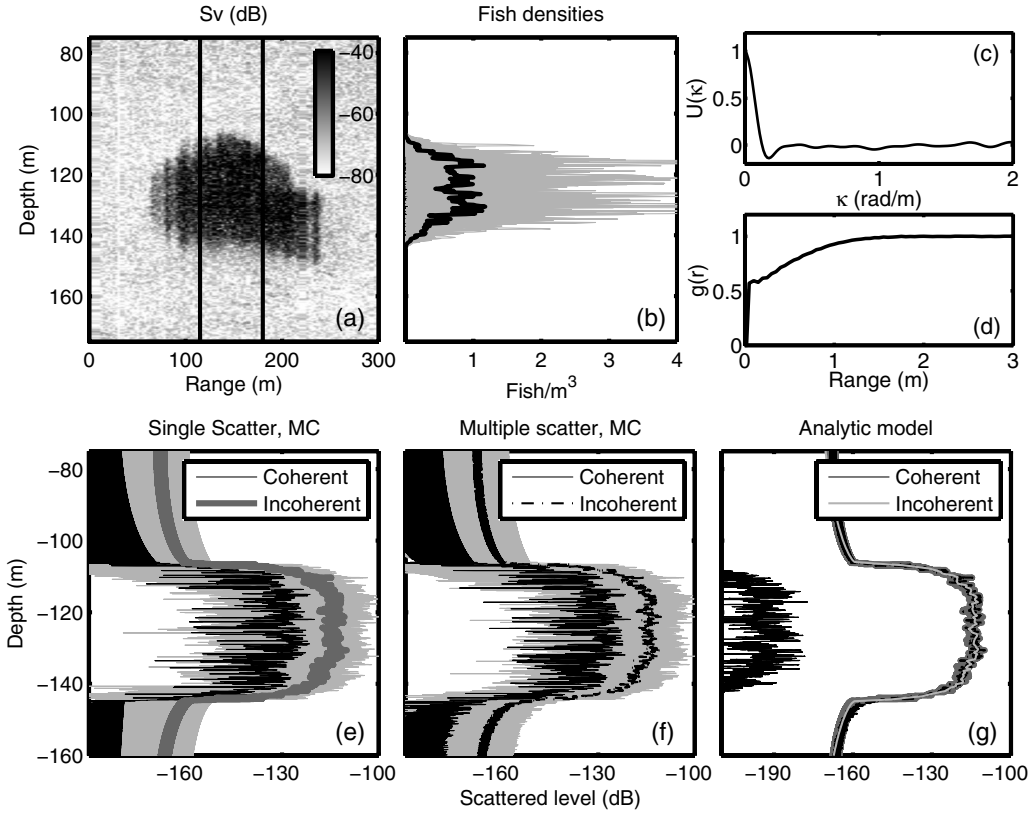


FIG. 6. Broadband matched filtered coherent and incoherent intensity levels scattered from a group of fish. (a) Echosounder data in terms of volumetric scattering strength (S_v) are plotted as a function of depth and range. (b) This is converted to volumetric densities in gray using an estimated mean target strength of −39.6 dB for Atlantic herring assuming single-scattering; the mean over the 25 measurements is shown in black. Numerical Monte-Carlo models simulate the matched filtered returns from this group characterized by the pair distribution function, $g(r)$ illustrated in (d). The Monte-Carlo simulation results using (e) the single-scatter approximation and (f) including multiple scattering show the individual matched filtered signals in light gray, along with the sample mean square, or coherent intensity, and sample variance, or incoherent intensity. Panel (g) compares the incoherent intensities from the numerical Monte-Carlo simulations in (e) and (f) to the coherent and incoherent intensities found using the analytic model. The characteristic function, $U(k)$, for this distribution used in the analytic model is illustrated in (c).

=580 taken over the 25 depth profile samples measured in the region of interest. The standard deviation, σ_N, is a result of both a variation in the number of fish over range and variations in the herring backscatter target strength and other sources of noise and variation. Regardless of the source of the variation, averaging the volumetric profiles over a statistically stationary region reduces the error in both the mean number of fish, $\langle N \rangle$, and the mean volumetric density profile shown in Fig. 6(b). This inversion to volumetric fish density assumes that (1) single scattering is dominant, and (2) the incoherent intensity is dominant over the coherent. Here, the authors show using numerical Monte-Carlo simulations that these assumptions hold even for these relatively dense schools. The analytic model described in Sec. II can then be applied to rapidly compare the matched filtered coherent and incoherent intensities for a variety of fish distributions.

For the numerical Monte-Carlo simulation model, the three-dimensional fish spatial distributions are first generated within the echosounder beam using the fish density profile in Fig. 6(b) to determine their depth dependence and the pair distribution function shown in Fig. 6(d) to determine the inter-fish spacings. Noise in the data and variations in the fish target strength may have resulted in errors in the estimated density profiles. In the Monte-Carlo model, these values are assumed to be correct. By averaging the results of the Monte-Carlo simulations, they reduce this error, as long as the estimated mean target strength is accurate. The pair distribution function of the herring depends on their individual behaviors and group interactions, and has not been thoroughly quantified although it is an area of ongoing research.[18] For dense groups where several fish occupy a resolution cell, it cannot be determined from the data. Here, the fish are assumed to be distributed according to a uniform PDF within an individual resolution cell, except that they cannot be within the near-field of each other. It can be seen in Fig. 6(d) that the pair distribution function converges to unity near 1 m, which is the approximate mean inter-fish spacing for a density of 1 fish/m^3.

The primary scattering mechanism for herring imaged at 38 kHz are their air-filled swimbladder.[29–31] It has been found that the Rayleigh–Born scattering from the body of the fish is important at much higher frequencies; however, it has secondary importance at 38 kHz.[30] The herring swimbladder is a complicated three-dimensional air-filled structure.[32] It is usually modeled as a prolate spheroid.[33] Here, for simplicity, the swimbladder is modeled as a pressure-release sphere in the interest of determining whether or not multiple scattering is significant at all. The swimbladder radius is adjusted to

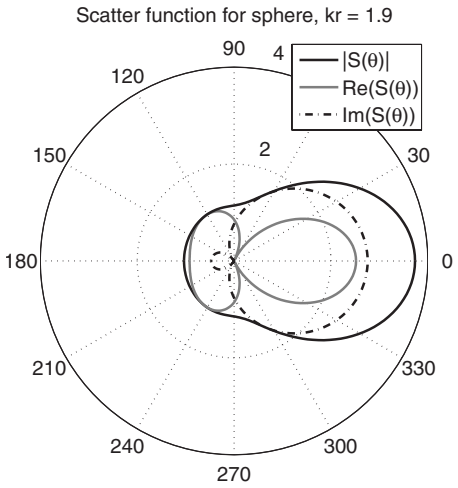


FIG. 7. The angular dependent scatter function for the pressure release sphere used in the numerical and analytic models is illustrated. The magnitude, and real and imaginary parts are plotted in polar coordinates, where 0° indicates the forward scatter direction and 180° indicates backscatter.

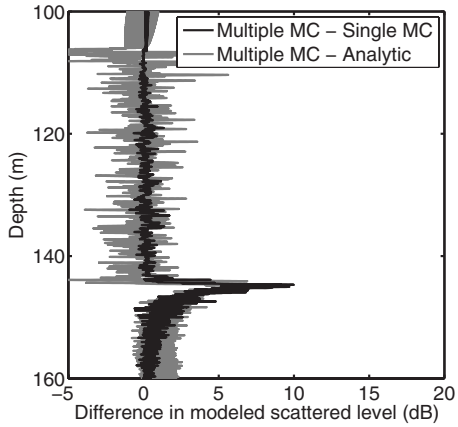


FIG. 8. The differences in the modeled scattered levels from Fig. 6(g) illustrate the effect of multiple scattering over the imaged depth. The difference between the Monte-Carlo multiple scattering model and Monte-Carlo single scattering model is shown in black, while the difference between the Monte-Carlo multiple scattering model and analytic model is shown in gray.

have a mean radius $a=1.2$ cm with standard deviation of 0.2 cm so that the backscattered target strength is consistent with experimentally determined values.[29]

For the herring modeled here, the far-field[15] condition is $r>4a^2/\lambda=1.5$ cm for a 38 kHz echosounder. The scatterers are in the far-field of each other since the mean inter-fish spacing of 1 m is much larger than this distance. Furthermore, the pair distribution function, $g(r)$, shown in Fig. 6(d) is defined so that the fish swimbladders are spaced at least 5 cm apart. This is a realistic requirement because the herring body is on average 25 cm in length and roughly 5 cm in width; mean spacing of the herring is larger than this for mean volumetric density of 1 fish/m^3 as well as in the locally very dense regions where the density can be up to 4 fish/m^3. Since the fish are in each other's far-field, there is negligible attenuation caused by the near-field shadow, which has a length of $(ka/2)^{1/3}a=1.2$ cm. At the operating frequency of 38 kHz, the product $ka\approx 2$ so that the scatterers are non-compact and the angular and range dependent scattered field from the swimbladder can be modeled exactly using Eq. (10.15) of Ref. 34, which includes near-field effects. Since the scatterers here are in the far-field of each other, the scattered field can be approximated by applying the scatter function in Eq. (10.18) of Ref. 34 with negligible error. The half power beam-width of the scatter function employed for the swimbladder at 38 kHz is roughly 90° as shown in Fig. 7.

The broadband matched filtered scattered field simulated using the numerical Monte-Carlo simulation model for 25 realizations of fish distribution within the sonar beam as a function of depth is plotted in gray in Figs. 6(e) and 6(f) for the singly and multiply scattered fields, respectively. In each figure, the sample mean square and variance over the multiple realizations give us the coherent and incoherent intensities. Here the authors see that the incoherent intensity dominates by at least 10 dB. The sample mean square or coherent intensity would be lower had more independent measurements available. Comparing the singly and multiply scattered incoherent matched filtered intensities in Figs. 6(g) and 8 generated using the Monte-Carlo model, they are found to be nearly identical, except at the tail end of the distribution near 145 m depth where the multiply scattered field is slightly higher due to delayed returns. This difference is negligible for population estimation since it is more than 40 dB down from the densely populated region. The single scattering assumption also implies that the incident acoustic intensity is not significantly attenuated through the fish school.

Since the single-scattering assumption is sufficient for fish schools of the given target strength and density, the analytic model can be used to compare the coherent and incoherent intensities using Eqs. (7) and (14), respectively, shown in Fig. 6(g). The mean distribution in Fig. 6(b) is used as the PDF for Eq. (8). The means and variances of the scatter function $S(\Omega_i,\Omega,k)$ and the number of scatterers N used in the analytic model are the same as those used in the Monte-Carlo simulations. Here, we see that the analytic incoherent intensity is in good agreement with the Monte-Carlo result to within 1 dB in the fish region, as shown in Figs. 6(g) and 8. This shows that the correlation between adjacent scatterers has negligible effect on the total intensity and the pair distribution function can be approximated as unity in deriving Eq. (14).

The analytically determined coherent intensity is found to be nearly 80 dB below the incoherent which was not predicted with the numerical models because many more independent realizations would be required to sufficiently converge the sample mean square value to the expected mean square. The convergence of the sample mean square is slow in this case when plotted in log because it is converging to near zero intensity values. Here a drop of more than 10 dB corresponds to a 90% reduction in intensity. Figure 9 illustrates the sample mean square using 4, 12, and all 25 samples to illustrate the slow trend of convergence.

In this section, we could only include up to third order scattering in the Monte-Carlo model because of computational memory constraints. The result obtained here that multiple scattering is negligible for ultrasonic frequency echosounder imaging of dense fish schools is consistent with the

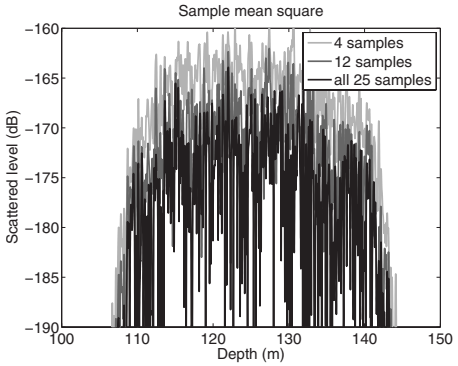


FIG. 9. The coherent intensity for the Monte-Carlo model decreases as more independent samples are added. To obtain the analytic result shown in Fig. 6(g), many more samples would be required.

experimental findings of Ref. 20 and theoretical result of Ref. 35 which considers multiple scattering only up to second order. Both these references, however, do not account for the effects of the matched filter. Further analysis with the three-dimensional Monte-Carlo model developed here indicate that either the fish target strength or volumetric densities need to be 10 dB larger for multiple scattering to matter. These higher densities and target strengths, however, are not realistic for fish distributions that occur in nature.

V. CONCLUSION

Analytic expressions have been derived from first principles, for the statistical moments of the broadband matched filtered scattered field from a random spatial distribution of random targets. The model is applicable when there is dispersion in the medium or target. The theory and analysis presented here explain how high-resolution population density images of randomly distributed objects or organisms can be obtained through cross-spectral coherence in the matched filter variance. The analytic model can be applied to object distributions where (1) the single-scattering approximation is valid, (2) the scatterers are in the far-field of each other, and (3) the correlation between scatterers makes negligible contribution to the total scattered intensity.

The analytic model is verified with a numerical Monte-Carlo model simulating a distribution of Atlantic herring imaged with an ultrasonic echosounder. The Monte-Carlo model illustrates that multiple scattering is negligible, even for relatively high volumetric densities of 1 fish/m^3. Upon averaging, the sample variance of the Monte-Carlo model converges to levels predicted by the analytic model. The authors show that the incoherently scattered intensity can be used to image population densities with resolution inversely related to bandwidth, $\Delta r = c/2B$.

ACKNOWLEDGMENTS

This research was funded by the Office of Naval Research, the National Oceanographic Partnership Program and the Alfred P. Sloan Foundation with administrative support from Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems. This research is a contribution to the Census of Marine Life.

[1]N. C. Makris, P. Ratilal, D. T. Symonds, S. Jagannathan, S. Lee, and R. W. Nero, "Fish population and behavior revealed by instantaneous continental shelf-scale imaging," Science **311**, 660–663 (2006).
[2]D. N. MacLennan and E. J. Simmonds, *Fisheries Acoustics*, 2nd ed. (Chapman and Hall, London, 1992).
[3]W. S. Burdic, *Underwater Acoustic Systems Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
[4]S. Schock, "A method for estimating the physical and acoustic properties of the seabed using chirp sonar data," Inf. Sci. (N.Y.) **29**, 1200–1217 (2004).
[5]E. Eastwood, *Radar Ornithology* (Methuen, London, 1967).
[6]G. L. Turin, "An introduction to matched filters," IRE Trans. Inf. Theory **6**, 311–329 (1960).
[7]C. E. Cook and M. Bernfeld, *Radar Signals: An Introduction to Theory and Appliction* (Artech House, Boston, 1993).
[8]A. W. Rihaczek, *Principles of High-Resolution Radar* (Artech House, Boston, 1996).
[9]L. Tsang, J. A. Kong, and K.-H. Ding, *Scattering of Electromagnetic Waves: Theories and Applications* (Wiley, New York, 2000).
[10]J. W. Strobehn, *Laser Beam Propagation in the Atmosphere* (Springer, Germany, 1978).
[11]H. C. van de Hulst, *Light Scattering by Small Particles* (Dover, New York, 1957).
[12]A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
[13]E. N. Fowle, E. J. Kelly, and J. A. Sheehan, "Radar system performance in a dense-target environment," IRE Int. Conv. Rec. **4**, 136–145 (1961).
[14]P. Ratilal and N. C. Makris, "Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities," J. Acoust. Soc. Am. **118**, 3532–3559 (2005).
[15]M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, Cambridge, 1980).
[16]N. Levanon, *Radar Principles* (Wiley, New York, 1988).
[17]J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).
[18]J. H. Tien, S. A. Levin, and D. I. Rubenstein, "Dynamics of fish shoals: Identifying key decision rules," Evol. Ecol. Res. **6**, 555–565 (2004).
[19]I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks, "Collective memory and spatial sorting in animal groups," J. Theor. Biol. **218**, 1–11 (2002).
[20]K. G. Foote, "Linearity of fisheries acoustics, with addition theorems," J. Acoust. Soc. Am. **73**, 1932–1940 (1983).
[21]B. Newhall, "Continuous reverberation response and comb spectra waveform design," Inf. Sci. (N.Y.) **32**, 524–532 (2007).
[22]N. C. Makris, "The effect of saturated transmission of scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).
[23]B. R. Frieden, *Probability, Statistical Optics and Data Testing*, 3rd ed. (Springer, Germany, 2001).
[24]Z. Ye and A. Alvarez, "Acoustic localization in bubbly liquid media," Phys. Rev. Lett. **80**, 3503–3506 (1998).
[25]J. E. Ehrenberg and T. C. Torkelson, "Fm slide (chirp) signals: A technique for significantly improving the signal-to-noise performance in hydroacoustic assessment systems," Fish. Res. **47**, 193–199 (2000).
[26]Z. Gong, M. Andrews, D. Tran, D. Cocuzzo, S. Dasgupta, S. Jagannathan, I. Bertsatos, D. Symonds, T. Chen, H. Pena, R. Patel, O. R. Godo, R. W. Nero, J. M. Jech, N. Makris, and P. Ratilal, "Atlantic herring (clupea harengus) low frequency target strength and abundance estimation: Ocean acoustic waveguide remote sensing (oawrs) 2006 gulf of maine experiment," J. Acoust. Soc. Am. submitted.
[27]N. C. Makris, P. Ratilal, S. Jagannathan, Z. Gong, M. Andrews, I. Bertsatos, O. R. Godo, R. W. Nero, and J. M. Jech, "Critical population density triggers rapid formation of vast oceanic fish shoals," Science **323**, 1734–1737 (2009).
[28]W. J. Overholtz, J. M. Jech, W. Michaels, L. Jacobson, and P. Sullivan, "Empirical comparisons of survey designs in acoustic surveys of gulf of maine-georges bank atlantic herring," J. Northw. Atl. Fish. Sci. **36**, 127–144 (2006).
[29]E. Ona, "An expanded target strength relationship for herring," ICES J. Mar. Sci. **70**, 107–127 (2003).
[30]K. G. Foote, "Importance of the swimbladder in acoustic scattering by fish: A comparison on gadoid and mackerel target strengths," J. Acoust. Soc. Am. **67**, 2084–2089 (1980).

[31]C. Feuillade, R. W. Nero, and R. H. Love, "A low-frequency acoustic scattering model for small schools of fish," J. Acoust. Soc. Am. **99**, 196–208 (1996).
[32]N. Gorska and E. Ona, "Modelling the effect of swimbladder compression on the acoustic backscattering from herring at normal or near-normal dorsal incidences," ICES J. Mar. Sci. **60**, 1381–1391 (2003).
[33]R. W. Nero, C. H. Thompson, and J. M. Jech, "In situ acoustic estimates of the swimbladder volume of atlantic herring (clupea harengus)," ICES J. Mar. Sci. **61**, 323–337 (2004).
[34]J. Bowman, T. Senior, and P. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (North-Holland, Amsterdam, 1969).
[35]T. Stanton, "Multiple-scattering with applications to fish echo-processing," J. Acoust. Soc. Am. **73**, 1164–1169 (1983).

Temporal and vertical scales of acoustic fluctuations for 75-Hz, broadband transmissions to 87-km range in the eastern North Pacific Ocean

John A. Colosi
Department of Oceanography, Naval Postgraduate School, Monterey, California 93943

Jinshan Xu
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Peter F. Worcester and Matthew A. Dzieciuch
Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0225

Bruce M. Howe
University of Hawaii at Manoa, Honolulu, Hawaii 96822

James A. Mercer
Applied Physics Laboratory, University of Washington, Seattle, Washington 98105

(Received 26 February 2009; accepted 15 June 2009)

Observations of scattering of low-frequency sound in the ocean have focused largely on effects at long ranges, involving multiple scattering events. Fluctuations due to one and two scattering events are analyzed here, using 75-Hz broadband signals transmitted in the eastern North Pacific Ocean. The experimental geometry gives two purely refracted arrivals. The temporal and vertical scales of phase and intensity fluctuations for these two ray paths are compared with predictions based on the weak fluctuation theory of Munk and Zachariasen, which assumes internal-wave-induced sound-speed perturbations [J. Acoust. Soc. Am. **59**, 818–838 (1976)]. The comparisons show that weak fluctuation theory describes the frequency and vertical-wave-number spectra of phase and intensity for the two paths reasonably well. The comparisons also show that a resonance condition exists between the local acoustic ray and the internal-wave field, as predicted by Munk and Zachariasen, such that only internal waves whose crests are parallel to the local ray path contribute to acoustic scattering. This effect leads to filtering of the acoustic spectra relative to the internal-wave spectra, such that steep rays do not acquire scattering contributions due to low-frequency internal waves. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3177259]

PACS number(s): 43.30.Re, 43.30.Ft, 43.30.Dr [ADP] Pages: 1069–1083

I. INTRODUCTION

As a result of the emphasis during the past 2 decades on the application of ocean acoustic tomography to measure gyre-scale ocean variability, observational efforts to study acoustic scattering of low-frequency sound (30–300 Hz) have focused on effects at very long ranges between 1000 and 15 000 km (Duda *et al.*, 1992; Munk *et al.*, 1994; Dushaw *et al.*, 1995; Worcester *et al.*, 1994, 1999, 2000; Worcester and Spindel, 2005). Several important discoveries have been made. First, low-frequency, broadband time fronts recorded on vertical line array receivers have been found to be partitioned into two regimes. The early part of the time front that is composed of steep rays or higher order modes (often termed the ray-like region of the time front) shows well separated, quasi-planar fronts with small fluctuations in intensity and travel time. The later part of the time front that is composed of near-axial rays or low-order modes (often termed the time front finale) shows a complicated multipath interference pattern with large fluctuations similar to Gaussian random noise (Colosi and the ATOC Group 1999; Colosi *et al.*, 1999, 2001). Second, in both regimes there is substantial scattering of acoustic energy into the geometric shadow zone predicted using climatological sound-speed profiles, extending the time front in depth and time (Colosi *et al.*, 1994; Colosi and Flatté, 1996; Worcester *et al.*, 1994, 1999; Dushaw *et al.*, 1999; Van Uffelen *et al.*, 2009). This extension of acoustic energy into the shadow zone shows that the effect of scattering in long-range, low-frequency ocean acoustic propagation is to introduce a significant bias into the time front intensity pattern. The acoustic fluctuations cannot be considered a zero mean effect superimposed upon an otherwise deterministic time front pattern. Finally, in spite of the large fluctuations in the time front finale, the temporal stability of the phase there is surprisingly high and close to that of the ray-like region (Colosi *et al.*, 2005; Wage *et al.*, 2005). Observed coherence times are between 5 and 15 min.

It has been known for some time that the strongest acoustic scattering occurs near ray upper turning points (UTPs) (Flatté *et al.*, 1979). Long-range propagation therefore involves multiple scattering events, which may obscure the fundamental scattering physics. The motivation for this work is to examine the scattering physics of one and two

scattering events using data from the Acoustic Engineering Test (AET) of the Acoustic Thermometry of Ocean Climate (ATOC) project, in order to better understand the results from the long-range experiments. During the AET, broadband acoustic signals were transmitted from a 75-Hz source suspended near the depth of the sound-channel axis to 700-m-long vertical receiving arrays approximately 87 and 3252 km distant in the eastern North Pacific Ocean. Results from 3252-km range have been reported previously (Colosi and the ATOC Group 1999; Colosi *et al.*, 1999; Worcester *et al.*, 1999; Colosi *et al.*, 2001). At 87-km range, the arrival pattern consists of two time-resolved and identifiable time fronts. The first arrival has a ray ID of −3, with a negative (downward) launch angle at the source, one UTP, and two lower turning points (LTPs). (See Munk *et al.*, 1995, for a discussion of ray nomenclature.) The second arrival has a ray ID of +4, with a positive (upward) launch angle at the source, two UTPs, and two LTPs. Previous work on single UTP propagation has been entirely at frequencies of 1000 Hz or more (Worcester, 1979; Worcester *et al.*, 1981; Flatté, 1983; Ewart and Reynolds, 1984), where the acoustic fluctuations are quite strong. Fluctuations are expected to be much weaker at low frequency.

The scattering physics theory to which the data are compared is due to Munk–Zachariasen (MZ) (Munk and Zachariesen, 1976), who modified Rytov's weak fluctuation theory of optical propagation through a turbulent atmosphere (Rytov, 1937) to the considerably more complex problem of ocean acoustic propagation through internal waves. The basic physics of the MZ theory that is to be tested is that there is weak, single forward scattering and that there is a resonance condition between the sloping ray path and the internal waves whose crests are aligned with the sloping ray. The resonance condition leads to an important selectivity in acoustic-internal wave interactions such that steep rays can be too steep to interact with the low-frequency part of the internal-wave field. The MZ theory is used instead of the more sophisticated path-integral theory (Flatté *et al.*, 1979, 1987), of which it is a special case, because of the conceptual simplicity of MZ, and due to the fact that there are not enough observations to make a strongly discriminating comparison.

The measured intensity fluctuations are very weak, as expected, and thus consistent with the application of a perturbative approach like the MZ theory. The scintillation indices (SIs) and variances of log-intensity ($\sigma^2_{\ln I}$) for the two arrivals computed from the 6 days of observations are SI $=0.04$ and $\sigma_{\ln I}=0.8$ dB for ray ID −3 (one UTP), and SI $=0.4$ and $\sigma_{\ln I}=3$ dB for ID +4 (two UTPs). Most important, however, are the time scales of the observed intensity variability. The observed frequency spectra of intensity show that the arrivals from the steeper ray with ID −3 have much less low-frequency variability than the arrivals with ID +4. This result is consistent with MZ and provides direct observational evidence of the resonance condition that prevents steep rays from interacting with low-frequency internal waves. In addition, vertical-wave-number spectra of intensity show order of magnitude agreement with MZ.

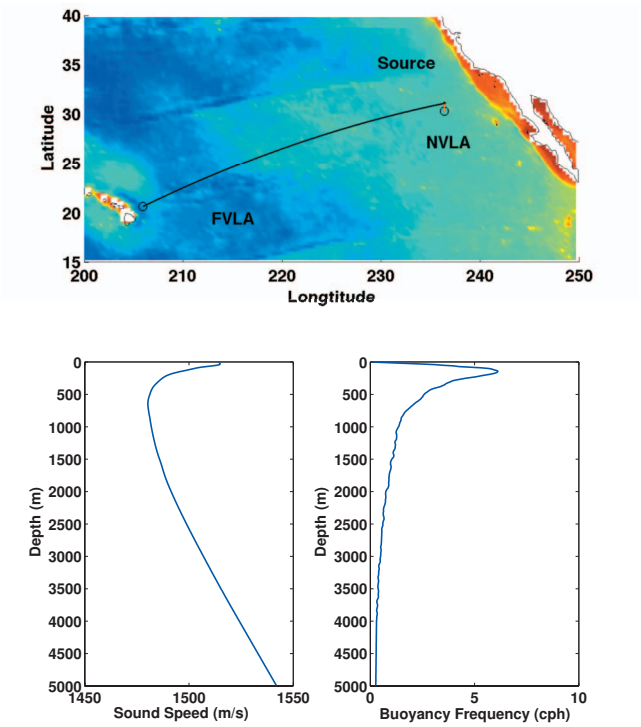


FIG. 1. (Top) Geometry of the ATOC AET in the North Pacific Ocean. The research platform R/P FLIP occupied the source location, and 700-m long vertical line array receivers were located at the near VLA (NVLA) and far VLA (FVLA) positions. (Bottom) Sound-speed and buoyancy-frequency profiles derived from CTD casts made during the deployment and recovery of the NVLA.

The AET data only allow phase fluctuations to be measured over short periods, because the longest transmissions only lasted 40 min. Phase varied too rapidly to be tracked over the gaps between transmissions, which occurred every 2–4 h. Phase variances are very small, of order 0.6 rad rms, and roughly the same for the two ray IDs, over these short time intervals. The frequency spectra of phase are only marginally in the internal-wave frequency band, making meaningful comparisons with MZ difficult. The vertical-wave-number spectra of phase are in good agreement with MZ, however.

The outline of this paper is as follows. Section II describes the observations and the processing needed to obtain the phase and intensity data used in the subsequent analyses. Section III presents various moments of phase and intensity, as well as frequency and wave number spectra. The observations are compared to the MZ theory in Sec. IV. Section V has summary and conclusions.

II. THE AET EXPERIMENT

The ATOC AET was conducted in the eastern North Pacific Ocean from 17–23 November 1994 (Worcester *et al.*, 1999). The acoustic source was suspended at a depth of 652 m, near the sound-channel axis, from the research platform R/P FLIP, which was moored roughly 400 miles south-southwest of San Diego, CA, at 31° 2.050′N, 123° 35.420′W (Fig. 1). The source transmitted periodic, phase-modulated signals with a center frequency of 75 Hz to autonomous vertical line array (NVLA) receivers approxi-

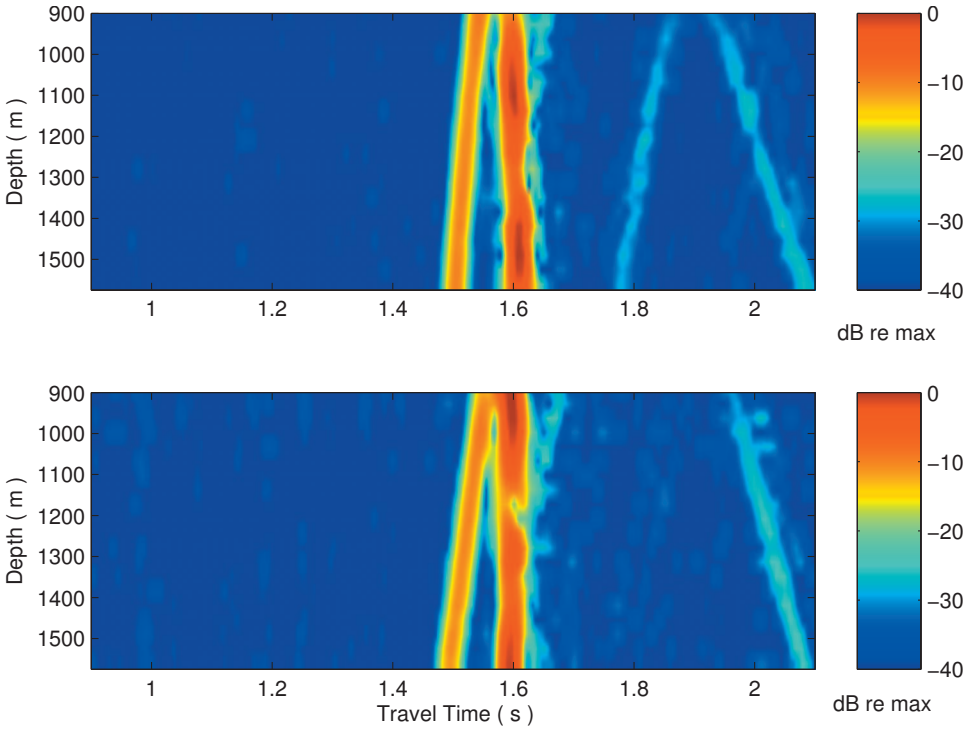


FIG. 2. Two examples of the intensity of the sound fields recorded by the near NVLA. The upper panel is from 08:01:52 UTC on year day 322, while the lower panel is from 00:01:52 UTC on year day 326. The color scale is in dB relative to the maximum value. The travel time is relative to the beginning of the reception.

mately 87.3 and 3252.4 km distant. Each NVLA consisted of 20 hydrophones at 35-m spacing. The near NVLA, for which results are reported here, was located south of R/P FLIP at 30° 14.798′N, 123° 36.491′W (Fig. 1). The near NVLA spanned a depth range of approximately 920–1585 m.

The phase modulation of the transmitted signals was encoded using a periodic linear maximal-length shift-register sequence (*m*-sequence) containing 1023 digits. Each digit contained 2 cycles of the 75-Hz carrier, giving a digit length of 26.667 ms and a sequence period of 27.2800 s. Fifty-four transmissions were made, consisting of a mix of 10-, 20-, and 40-min transmissions, with 2–4 h between transmissions. The 20 (40)-min transmissions consisted of 44 (88) periods of the 27.2800-s *m*-sequence, of which 40 (80) periods were recorded at the NVLA receivers. The recorded signals were processed by replica correlation without any period averaging prior to the correlation, yielding 40 (80) realizations of the time front for each reception. The data analyzed here therefore consist of acoustic pressure $p(z,T,t)$ as a function of depth z and travel time T for each realization of the time front, which are obtained at intervals of 27.2800 s in geophysical time t. Figure 2 shows two time fronts obtained by pulse compression of different 27.2800-s periods. Due to various technical problems, only 24 of the 40-period receptions and 3 of the 80-period receptions were usable. Receptions were deemed unusable if any hydrophone channels were corrupted or missing data over the entire 40- or 80-period receptions.

A. Ray identification

Two high intensity wavefronts swept by the near NVLA, followed by much weaker, bottom-interacting arrivals (Fig. 2). An eigenray calculation using a sound-speed profile computed from conductivity-temperature-depth (CTD) casts during the deployment and recovery of the NVLA identifies the first arrival with ray paths that have a ray ID of −3 (negative (downward) launch angles and three turning points—one UTP and two LTPs) and the second arrival with ray paths that have a ray ID of +4 [positive (upward) launch angles and four turning points—two UTPs and two LTPs] (Fig. 3).

All of the rays associated with the portion of the time front with ray ID −3 recorded on the NVLA have UTPs in the depth range 90–130 m. The mixed layer during the experiment was shallow, of order 20–30 m, so ray ID −3 was not affected by the mixed layer, other than perhaps by diffractive effects. The rays associated with the time front with ray ID +4 have UTPs with a much broader depth distribution, spanning the depth range 225–350 m. The UTPs for ray ID −3 occur close in range to the second UTPs of ray ID +4 (Table I). The differences in UTP depth will be important in the subsequent theoretical analysis, because rays with ID −3 turn at depths where the buoyancy frequency N is roughly 5–6 cph, while rays with ID +4 turn at depths where N is 3–4 cph (Fig. 1).

B. Arrival processing

Estimates of the amplitudes and phases of the two arrivals are made using the maximum-likelihood method of Ehrenberg *et al.* (1981). It is not feasible to simply extract the amplitudes and phases at the times of the peaks in the demodulated signals because the two arrivals overlap in time at the shallower hydrophones in the NVLA and therefore interfere. The pressure signal $r(T)$ at each hydrophone is modeled as the sum of the two arrivals with unknown amplitudes A_j, travel times T_j, and phases θ_j, plus additive noise $n(T)$, i.e.,

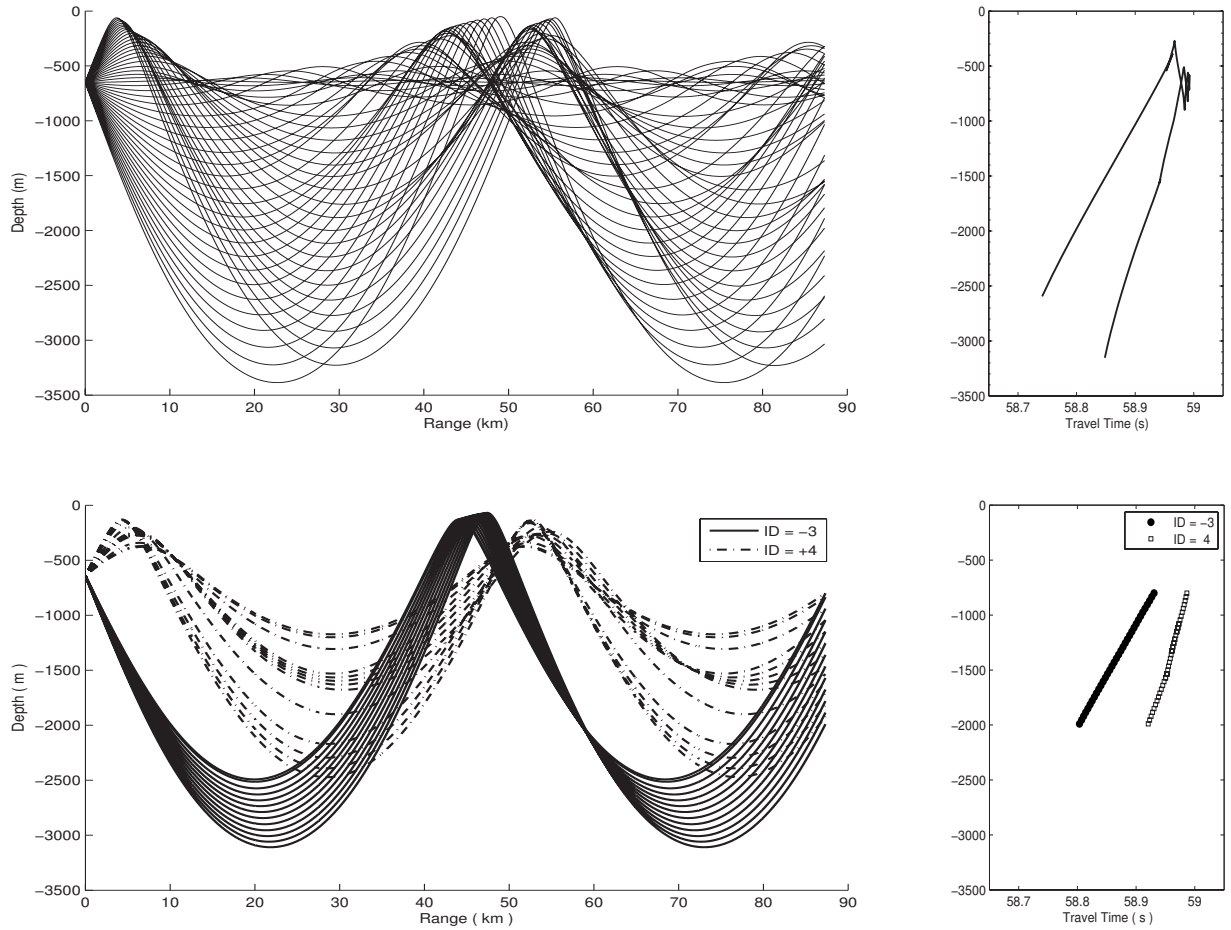


FIG. 3. Ray traces (left) and time fronts (right) for a range of 87.3 km and the range-independent AET sound-speed profile shown in Fig. 1. The upper panels show the rays and time fronts for a uniform fan of rays, while the lower panels show rays that arrive near the NVLA hydrophones for the two observed time fronts. The first time front has ray ID −3, and the second time front has ID +4.

$$r(T) = A_1 E(T - T_1)\cos(\omega(T - T_1) + \theta_1) + A_2 E(T - T_2) \times \cos(\omega(T - T_2) + \theta_2) + n(T). \quad (1)$$

The carrier frequency is $\omega = 2\pi f$, where f is 75 Hz. An estimate of the envelope of the pulse for a single path $E(T)$ is made using fully resolved arrivals from the depth ranges where there is no overlap between the two pulses. The travel times $T_j(z,t)$ established from the peaks of the envelopes are used to align the pulses, which are then averaged to give

$$E_j(T) = \langle I_j(T)\rangle^{1/2} = \left(\frac{1}{N_k N_i}\sum_k \sum_i I_j(T - T_j(z_k, t_i))\right)^{1/2}. \quad (2)$$

The envelopes are estimated separately for the two arrivals, with $j=1$ for ray ID −3 and $j=2$ for ray ID +4. $I_j(T - T_j(z_k,t_i))$ is the square of the absolute value of the complex envelope, and N_k and N_i are the number of hydrophone depths and receptions, respectively, included in the averages. The estimates of the pulse envelopes $E_j(T)$ for the two arrivals are essentially identical, as expected, justifying the use of the same envelope $E(T)$ for both arrivals in Eq. (1) (Fig. 4).

TABLE I. UTP depths and horizontal positions for time fronts with ID −3 and +4.

	ID −3	ID +4	
	UTP	First UTP	Second UTP
Depth Range (m)	90–140	225–350	225–350
Horizontal Position (km)	44–46	5–6	52–55

The maximum-likelihood estimator was applied to all of the receptions, yielding a set of complex demodulates of the form

$$\psi(z,t,\mathrm{ID}) = A(z,t,\mathrm{ID})e^{i\phi(z,t,\mathrm{ID})}, \quad (3)$$

where $A(z,t,\mathrm{ID})$ are the estimated arrival amplitudes. The total phases $\phi(z,t,\mathrm{ID}) = \theta_j - \omega T_j(z,t)$ are derived from the travel time and phase estimates. (In the complex demodulates, the use of the subscript j to identify the two arrivals is replaced by the use of the argument ID, for future clarity.) The amplitudes $A(z,t,\mathrm{ID})$ are re-normalized to remove any variations in hydrophone calibration and non-stationarity in depth by setting the time-mean intensities for each hydrophone to unity, i.e.,

$$\langle I(z,\mathrm{ID})\rangle = \langle A^2(z,\mathrm{ID})\rangle = \frac{1}{N_t}\sum_{j=1}^{N_t} A^2(z,t,\mathrm{ID}) = 1. \quad (4)$$

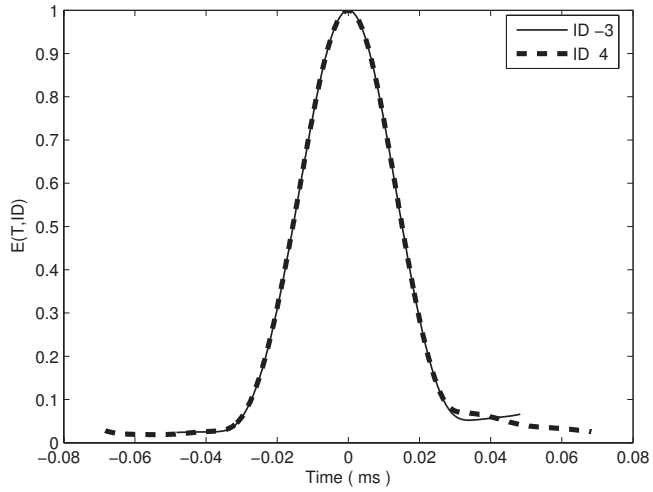


FIG. 4. Mean envelopes $E_j(T)$ for time fronts with ray ID −3 (solid) and ID +4 (dashed). The mean envelope is used in the maximum-likelihood method of Ehrenberg *et al.* (1981).

1. Phase unwrapping

Smooth, unwrapped phases as a function of time and depth can be constructed for the 40 (80) periods recorded for each of the 20 (40) min transmissions. Phase cannot be tracked between the 20 (40) min transmissions, however, because of the 2–4 h between transmissions. The geophysical time coordinate t will therefore be written as $t=\tau+\tau_l$, where the values of τ are only defined over the 40 (80) period reception intervals (i.e., $0\leq\tau\leq39\times27.28$ s, or $0\leq\tau\leq79\times27.28$). The variable τ_l denotes the time of the beginning of reception l. The notation for the complex demodulates then becomes

$$\psi(z,\tau,l,\mathrm{ID})=A(z,\tau,l,\mathrm{ID})e^{i\phi(z,\tau,l,\mathrm{ID})}. \quad (5)$$

A smooth, unwrapped phase function ϕ_u is constructed from ϕ for each reception l such that the mean square difference between gradients calculated from wrapped and unwrapped phases are minimized (Ghilia and Romero, 1994). Applying this smoothness criterion means that the unwrapped phase does not necessarily differ from the wrapped phase by an integer multiple of 2π, but in practice the results are extremely good because the acoustic variability is rather weak. A discussion of the accuracy of the technique is given in Colosi *et al.*, 2005.

Figure 5 shows the wrapped and unwrapped phases (second and third rows) for the two time fronts for one reception. The wrapped phases for time front with ray ID −3 show a rather large vertical gradient of phase due to the tilt of the time front as it encounters the NVLA. The time front with ray ID +4, on the other hand, shows little vertical gradient because it is sweeping past the NVLA closer to normal incidence (Fig. 2).

2. Source and mooring motion

The temporal phase variability seen in Fig. 5 is due in part to motion of the source (during the transmission) and NVLA (during the reception), as well as to time-dependent acoustic fluctuations. The vertical phase structure seen in

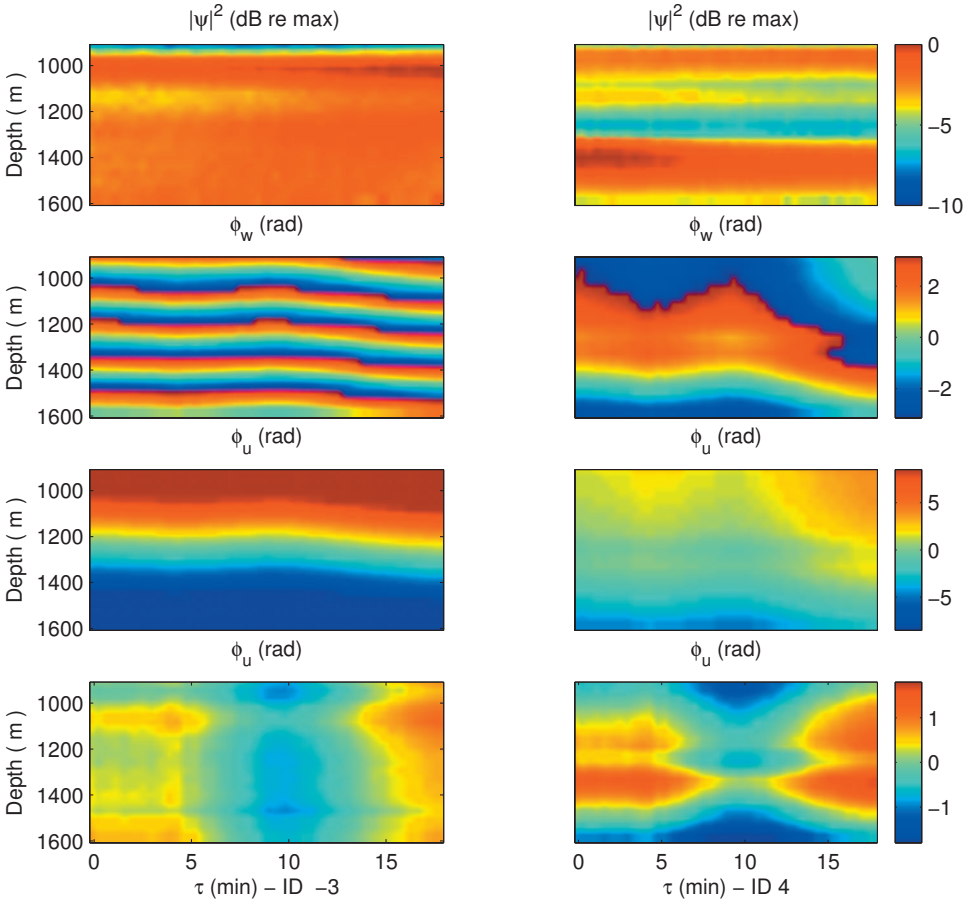


FIG. 5. Observed phase and amplitude fluctuations for time fronts with ray ID −3 (left) and ID +4 (right) over a 40-period reception. The top panel shows the signal intensity in dB relative to the maximum. The lower three panels (starting from the top) show the wrapped phase, unwrapped phase, and the unwrapped phase corrected for source and receiver motion as described in the text.

Fig. 5 is due in part to curvature and tilt of the NVLA, as well as to acoustic scattering. The positions of the source and the center of the NVLA were measured with an accuracy of 1–1.5 m rms by long-baseline acoustic navigation systems, using acoustic transponders on the seafloor (Worcester *et al.*, 1999). The positions of the source and NVLA were only measured once per hour, however, making it difficult to accurately determine the phase corrections required to account for source and receiver motion during the 20- and 40-min transmissions and receptions. In addition, the tilt and curvature of the NVLA were not measured, making it impossible to correct the phases for the relative positions of the hydrophones in the NVLA. (The system designed to determine the array tilt and curvature by recording the navigation signals from the acoustic transponders on six of the same hydrophones used to receive the 75-Hz signals unfortunately failed.) In order to correct at least crudely for source and receiver motion and NVLA tilt, a least-squares fit to the phase function was done to determine linear trends in both depth and time τ (Colosi *et al.*, 2005). For each reception l, the linear trends were subtracted, leading to a corrected phase function $\phi_c(z,\tau,l,\mathrm{ID})=\phi_u(z,\tau,l,\mathrm{ID})-\phi_{lsf}(z,\tau,l,\mathrm{ID})$ (Fig. 5). In the subsequent analysis, the corrected phase ϕ_c will be used to quantify acoustic variability.

III. OBSERVED PHASE AND INTENSITY FLUCTUATIONS

In this section, various moments of the amplitude $A(z,\tau,l,\mathrm{ID})$ and corrected phase $\phi_c(z,\tau,l,\mathrm{ID})$ are presented first, followed by frequency and vertical-wave-number spectra. The moments and spectra will be compared in the Sec. III A with predictions made using MZ (Munk and Zachariasen, 1976).

A. Moments

Calculation of various moments of the phase and amplitude is complicated by the irregular sampling during the AET. Phase moments can be computed only for the 40- and 80-period receptions, because the relative phase between receptions separated by 2–4 h is unknown. Amplitude moments can be computed for the entire 6-day period, however, as well as for 40- and 80-period receptions, because amplitude does not have a similar problem.

1. Phase

The phase variance is computed separately for each reception at each hydrophone for each ID and then averaged over all receptions l and all hydrophone depths z_k to obtain an estimate of the phase variances of the two arrivals

$$\sigma_\phi^2(\mathrm{ID})=\frac{1}{N_z N_l}\sum_{k=1}^{N_z}\sum_{l=1}^{N_l}\left[\frac{1}{N_\tau}\sum_{j=1}^{N_\tau}(\phi_c(z_k,\tau_j,l,\mathrm{ID})-\overline{\phi_c}(z_k,l,\mathrm{ID}))^2\right], \tag{6}$$

TABLE II. rms phase and log-intensity, SI, and variance of log-amplitude for 40-period (20-min), 80-period (40-min), and 6-day observation periods.

		σ_ϕ (rad)	σ_ι (dB)	SI	$\langle\chi^2\rangle$
40-period	ID −3	0.44 ± 0.01	0.26 ± 0.05	0.004 ± 0.002	0.0012 ± 0.0002
	ID +4	0.44 ± 0.01	0.26 ± 0.07	0.005 ± 0.001	0.0014 ± 0.0002
80-period	ID −3	0.61 ± 0.01	0.38 ± 0.10	0.009 ± 0.001	0.0023 ± 0.0003
	ID +4	0.68 ± 0.01	0.40 ± 0.05	0.013 ± 0.004	0.004 ± 0.001
6-day	ID −3		0.80 ± 0.07	0.044 ± 0.013	0.010 ± 0.002
	ID +4		3.1 ± 0.2	0.43 ± 0.05	0.13 ± 0.02

$$\overline{\phi_c}(z_k,l,\mathrm{ID})=\frac{1}{N_\tau}\sum_{j=1}^{N_\tau}\phi_c(z_k,\tau_j,l,\mathrm{ID}), \tag{7}$$

where $N_l=30$ and $N_\tau=40$ for the 40-period receptions, and $N_l=3$ and $N_\tau=80$ for the 80-period receptions. $N_z=20$ in both cases. Table II gives the rms phase variability for the two ray IDs for the two observation times. The error bars are computed using the variation in the phase variance estimates over the different hydrophone depths. As expected, the phase variance increases for the longer observation time, because the phase time series reflects more of the ocean's broadband variability. Even the 80-period (approximately 40-min) observation time is not much longer than the period of the highest frequency internal waves, however, which have a period of roughly 10 min, and the phase variance therefore likely does not include a significant amount of the variance that would be present in a longer time series. There is little difference in the phase variances of the two arrivals. Apparently, the one shallow UTP for the time front with ray ID −3 has the same effect as two deeper UTPs for the time front with ray ID +4, at least over these short observation periods. This result is consistent with that of Flatté and Stoughton (1988), who found that the phase variability is almost independent of time front ID.

2. Amplitude

Two different measures of amplitude variability are the SI and the variance of log intensity ($\iota=\ln A^2$),

$$\mathrm{SI}=\frac{\langle I^2\rangle-\langle I\rangle^2}{\langle I\rangle^2}\quad\text{and}\quad\sigma_\iota^2=\langle\iota^2\rangle-\langle\iota\rangle^2. \tag{8}$$

As before, these moments are computed for the 40- and 80-period observation times. In addition, for amplitude the entire 6-day observation time can be used:

$$\langle I^2\rangle(z_k,\mathrm{ID})=\frac{1}{N_\tau N_l}\sum_{j=1}^{N_\tau}\sum_{l=1}^{N_l}A^4(z_k,\tau_j,l,\mathrm{ID}), \tag{9}$$

$$\mathrm{SI}(\mathrm{ID})=\frac{1}{N_z}\sum_{k=1}^{N_z}\langle I^2\rangle(z_k,\mathrm{ID})-1, \tag{10}$$

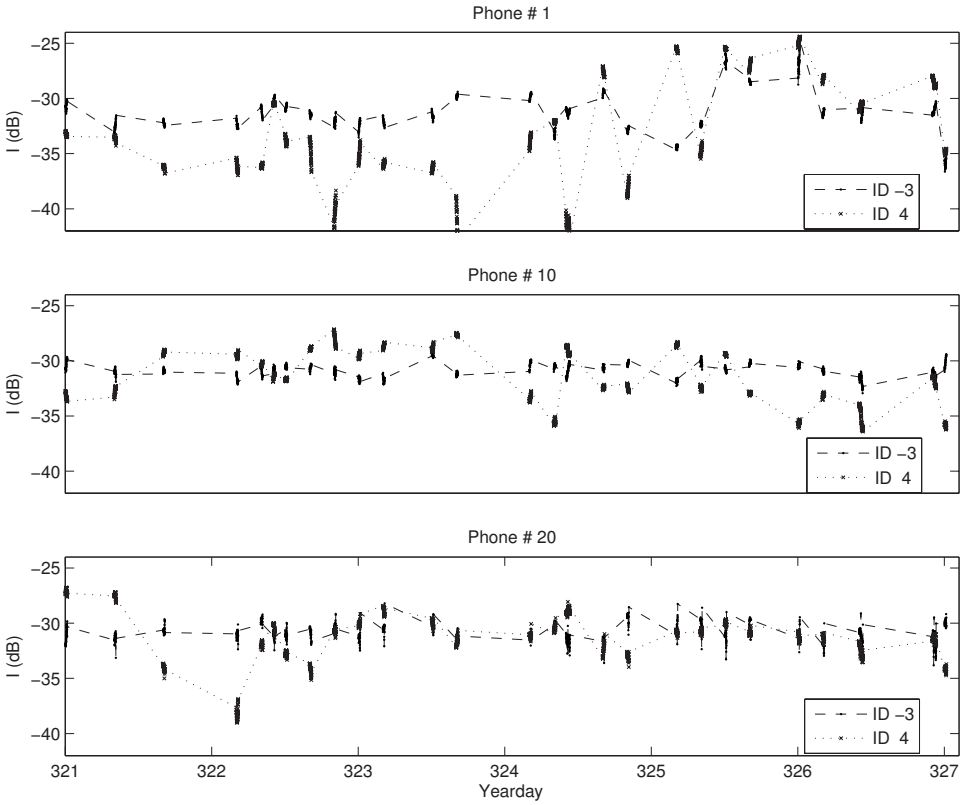


FIG. 6. 6-day time series of acoustic intensity for three different receiver depths. Ray ID −3 is dashed; ray ID +4 is dotted.

$$\sigma_\iota^2(\mathrm{ID}) = \frac{1}{N_z}\sum_{k=1}^{N_z}\left[\frac{1}{N_\tau N_l}\sum_{l=1}^{N_l}\sum_{j=1}^{N_\tau}(\iota(z_k,\tau_j,l,\mathrm{ID}) - \bar{\iota}(z_k,\mathrm{ID}))^2\right], \tag{11}$$

$$\bar{\iota}(z_k,\mathrm{ID}) = \frac{1}{N_\tau N_l}\sum_{l=1}^{N_l}\sum_{j=1}^{N_\tau}\iota(z_k,\tau_j,l,\mathrm{ID}). \tag{12}$$

For the shorter observation times, the calculation proceeds as for phase

$$\langle I^2\rangle(z_k,l,\mathrm{ID}) = \frac{1}{N_\tau}\sum_{j=1}^{N_\tau}A^4(z_k,\tau_j,l,\mathrm{ID}), \tag{13}$$

$$\langle I\rangle(z_k,l,\mathrm{ID}) = \frac{1}{N_\tau}\sum_{j=1}^{N_\tau}A^2(z_k,\tau_j,l,\mathrm{ID}), \tag{14}$$

$$\mathrm{SI}(\mathrm{ID}) = \frac{1}{N_z N_l}\sum_{l=1}^{N_l}\sum_{k=1}^{N_z}\frac{\langle I^2\rangle(z_k,l,\mathrm{ID}) - \langle I\rangle^2(z_k,l,\mathrm{ID})}{\langle I\rangle^2(z_k,l,\mathrm{ID})}, \tag{15}$$

$$\sigma_\iota^2(\mathrm{ID}) = \frac{1}{N_z N_l}\sum_{l=1}^{N_l}\sum_{k=1}^{N_z}\left[\frac{1}{N_\tau}\sum_{j=1}^{N_\tau}(\iota(z_k,\tau_j,l,\mathrm{ID}) - \bar{\iota}(z_k,l,\mathrm{ID}))^2\right], \tag{16}$$

$$\bar{\iota}(z_k,l,\mathrm{ID}) = \frac{1}{N_\tau}\sum_{j=1}^{N_\tau}\iota(z_k,\tau_j,l,\mathrm{ID}). \tag{17}$$

Table II displays the estimates of SI and σ_ι for the two arrivals. As with the phase, the error bars are computed using the variation in the statistical estimates over the different hydrophone depths. As observation time increases, the variation increases, with the largest increase occurring from the 80-period observation time to the 6-day observation time. The intensity fluctuations for the short observation times are similar for the two arrivals, but σ_ι is significantly larger for ID +4 (3.1 dB rms) than for ID −3 (0.8 dB rms) for the 6-day observation time. This result is important because it suggests that the time scales of variability of the two arrivals are different, with ID −3 showing much less low-frequency variability than ID +4. This result will be examined in more detail when the frequency spectra are presented.

Finally, a calculation of the variance of log-amplitude ($\chi = \ln A$) reveals that $\mathrm{SI} \simeq 4\langle\chi^2\rangle$, a result that is valid if the amplitudes A obey a log-normal distribution (Table II). The log-normal distribution is expected to be valid in the weak fluctuation or unsaturated regime (Flatté *et al.*, 1979; Flatté, 1983).

B. Frequency spectra

Frequency spectra of phase and log-amplitude $\chi = \ln A$ are computed for each reception at each hydrophone for each ID, yielding $\hat{S}_{\phi,\chi}(\omega;z,l,\mathrm{ID})$. These spectra are computed by fast Fourier transform of either 40 or 80 samples with a sample interval of 27.28 s, after detrending and windowing

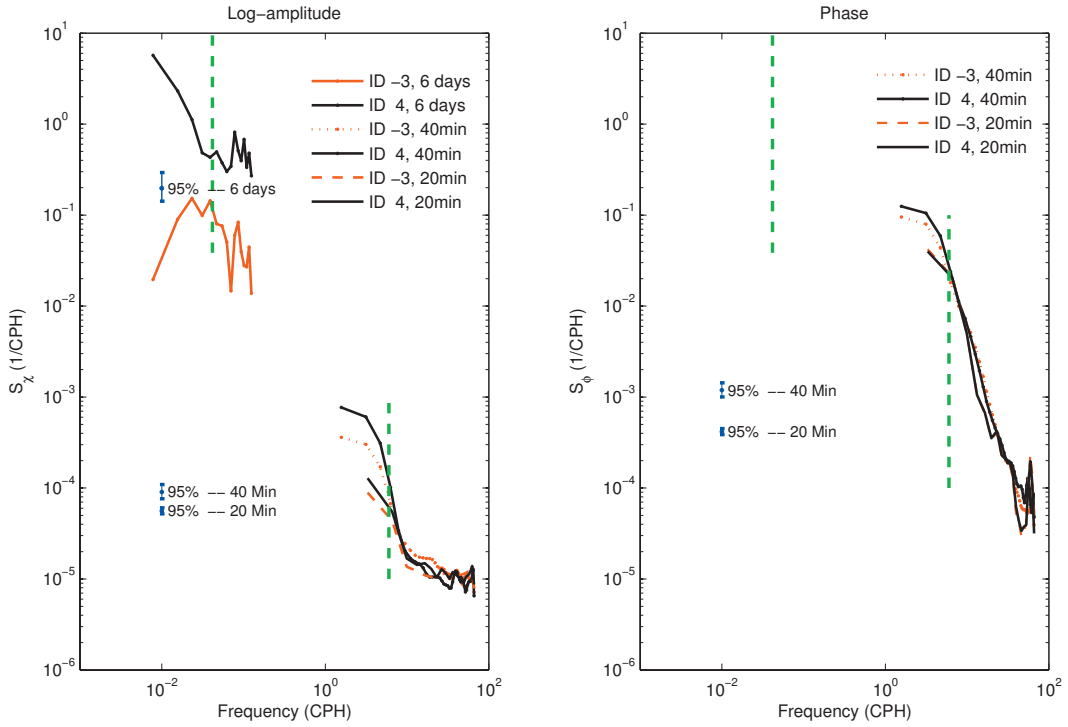


FIG. 7. Observed frequency spectra of log-amplitude (left) and phase (right) for time fronts with ray ID −3 (red) and ID +4 (black). The Coriolis frequency and the maximum local buoyancy frequency (6 cph) are indicated by vertical green dashes.

with a Hanning function. To obtain the final spectral estimate, spectra from all receptions l and hydrophone depths z_k are averaged, giving

$$S_{\phi,\chi}(\omega;\mathrm{ID}) = \frac{1}{N_z N_l}\sum_{l=1}^{N_l}\sum_{k=1}^{N_z}\hat{S}_{\phi,\chi}(\omega;z_k,l,\mathrm{ID}). \tag{18}$$

This procedure is carried out separately for the 40- and 80-period (approximately 20- and 40-min) observation times.

The spectra computed for each reception are limited to relatively high frequencies by the short durations of the receptions. Data are available for log-amplitude over longer time scales (6 days), and thus lower frequencies, than for phase, as discussed above. The data are irregularly sampled, however, causing some difficulty in carrying out spectral analysis (Fig. 6). After experimenting with several methods, the log-amplitude data were interpolated onto a uniform grid every 4 h using cubic splines. Thus, the 30 instances of 40-period receptions over the 6-day experiment were interpolated onto 36 points to carry out the spectral analysis. Fourier transforms were computed for each hydrophone depth, after detrending and windowing with a Hanning function. The final spectral estimates were obtained by averaging over depth.

Figure 7 shows the frequency spectra of phase and log-amplitude for the two arrivals. The maximum local buoyancy frequency (6 cph) and the local Coriolis parameter f are indicated by vertical green lines; this is the internal-wave frequency band. The phase spectra show nearly an ω^{-3} slope over the entire observed frequency range, with a slight flattening of the slope at the buoyancy frequency. The observed spectral slope is consistent with the result that the 80-period observation times had a larger phase variance than the 40-period observation times. In addition, the spectra for ID −3 and ID +4 are similar. Clearly, the data do not adequately resolve the internal-wave frequency band.

The log-amplitude spectra flatten at the highest frequencies, likely due to noise. At about 10 cph, the spectral energy increases rapidly. Again, the increase in spectral energy with decreasing frequency is consistent with the result that the 80-period observation times had more variance than the 40-period observation times. As with phase, the high-frequency end of the spectra are similar for the two IDs, and the time sampling does not adequately resolve the internal-wave frequency band. The low-frequency end of the spectra show very different behaviors. Here, the time front with ray ID −3 has much less low-frequency energy than the time front with ID +4, consistent with the result from the analysis of intensity variance. Some of the additional intensity variance comes from the internal-wave band (i.e., $\omega > f$), but some also comes from the sub-inertial band (i.e., $\omega < f$), especially for ID +4.

C. Vertical-wave-number spectra

Vertical-wave-number spectra of phase and log-amplitude are computed separately for each of the 40 or 80 periods at times τ in each reception l for each ID, yielding an estimate $\hat{S}_{\phi,\chi}(k_z;\tau,l,\mathrm{ID})$. The vertical data are first detrended and then windowed with a Hanning function before computing the Fourier transforms. To obtain the final spectra, an ensemble average is done over all times τ in all receptions l such that

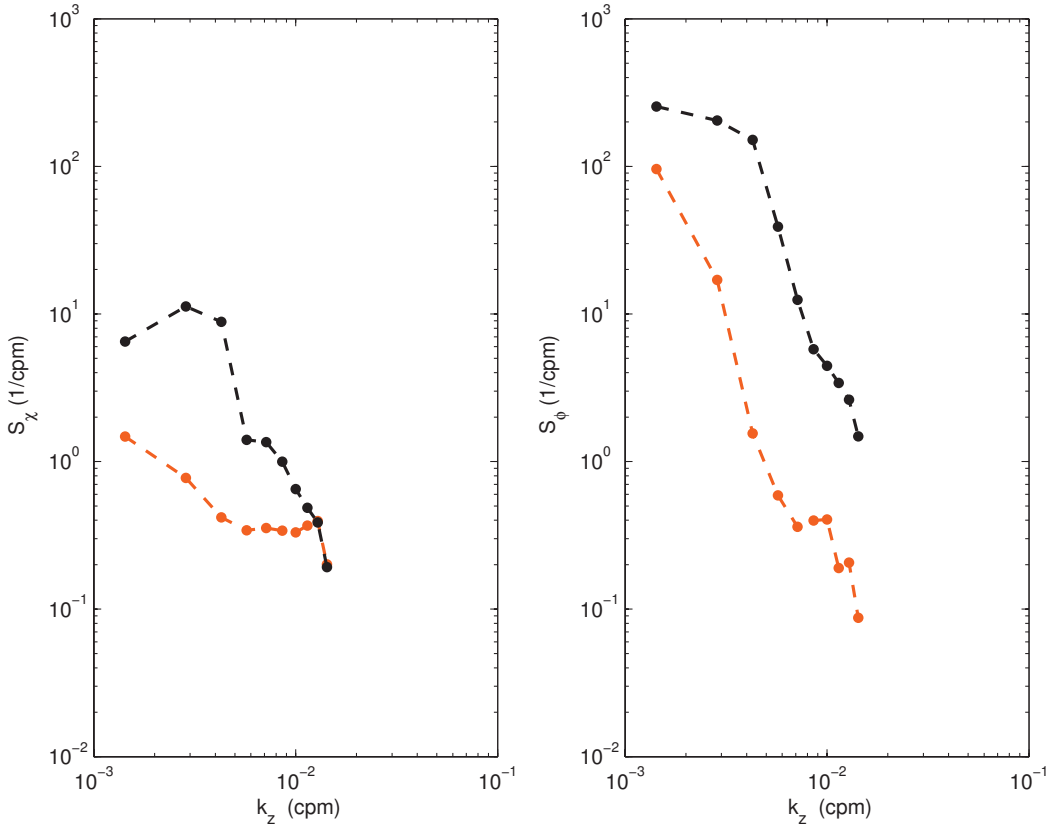


FIG. 8. Observed vertical-wave-number spectra of log-amplitude (left) and phase (right) for time fronts with ray ID −3 (red) and ID +4 (black).

$$S_{\phi,\chi}(k_z;\mathrm{ID}) = \frac{1}{N_\tau N_l}\sum_{l=1}^{N_l}\sum_{j=1}^{N_\tau}\hat{S}_{\phi,\chi}(k_z;\tau_j,l,\mathrm{ID}). \tag{19}$$

Both the 40- and 80-period data are combined in the ensemble average. The resulting vertical-wave-number spectra are shown in Fig. 8. The phase spectra have a roughly k_z^{-3} shape for both IDs, with ID +4 showing somewhat of a roll off at low wave number. The spectra of log-amplitude for the two time fronts, however, are markedly different. The time-front with ray ID −3 has a rather flat wave number spectrum, while ID +4 has a steeper spectrum with a roll off around k_z=0.003 cpm.

IV. THEORY AND COMPARISON TO OBSERVATIONS

Intensity fluctuations were shown in the preceding section to be small (i.e., SI$\ll$1). Because SI$\simeq 4\langle\chi^2\rangle$, the intensity probability density function is likely close to log-normal. This evidence suggests that the observations are in the unsaturated regime (Flatté *et al.*, 1979), so that a perturbation treatment of the acoustic fluctuations is appropriate. Such a model was put forth by Munk and Zachariasen, (1976), and elegant analytical results are available for the Garrett–Munk (GM) internal-wave spectrum. This section discusses the MZ theory and compares the AET observations to predictions based on MZ (Munk and Zachariasen, 1976).

A. M–Z theory

Appendix summarizes the MZ theory. The frequency, vertical-wave-number spectra of phase, and log-amplitude are written as integrals along the ray path $z_r(x)$,

$$S_{\phi,\chi}(R,\omega,k_z) = \pi k_0^2\int_\Gamma ds S_\mu(k_\perp(\omega,k_z);z) \times H[\omega-\omega_L(z_r(x))]H[N(z_r(x))-\omega] \times\left[1\pm\cos\left(\frac{k_z^2R_{fz}^2(x)}{2\pi}\right)\right], \tag{20}$$

where H is the Heavyside step function. The plus sign refers to the spectrum of phase ϕ and the minus sign refers to the spectrum of log-amplitude χ. The integral involves two terms. The first term is the spectrum of relative sound-speed fluctuations μ evaluated for internal-wave wave numbers that are perpendicular to the sloping ray, $S_\mu(k_\perp(\omega,k_z);z)$. The second term (in square brackets) is a diffraction term involving the vertical Fresnel zone R_{fz}.

Equation (20) shows that there is an important resonance condition such that only internal waves whose wave numbers are perpendicular to the ray (i.e., internal waves whose crests are parallel to the ray) contribute to the acoustic fluctuations. Using the GM internal-wave spectrum, the evaluation of $S_\mu(k_\perp)$ under the perpendicular wave number constraint can be written in terms of frequency and vertical wave number, yielding (Appendix)

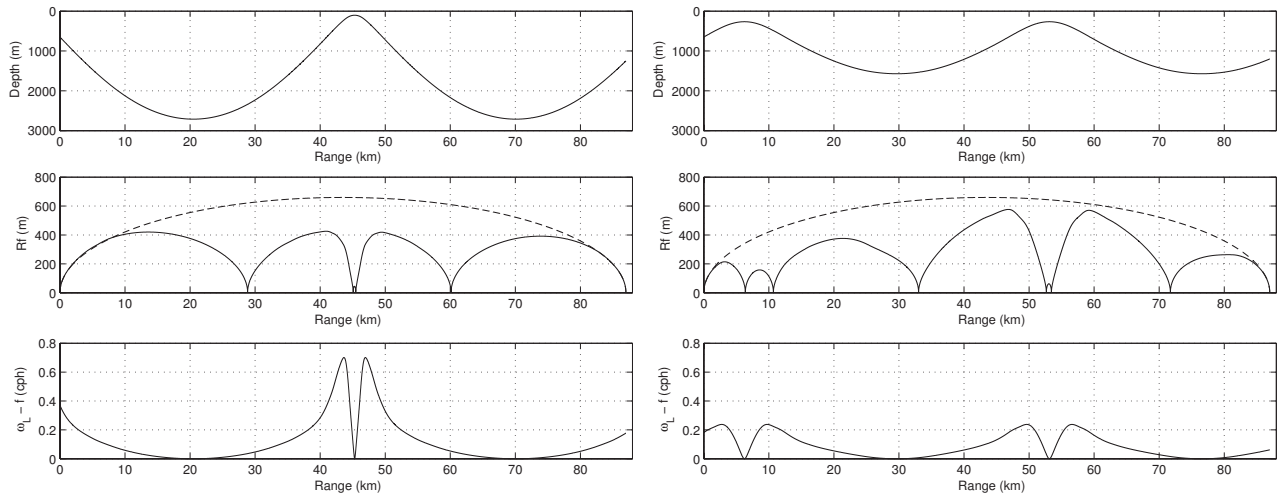


FIG. 9. Ray paths (upper), Fresnel zone $R_{fz}(x)$ (middle), and low-frequency cutoff minus the Coriolus parameter $(\omega_L(x)-f)$ (lower) for eigenrays with ID −3 (left) and ID +4 (right). In the middle panel, the Fresnel zone for constant background sound speed (dashed) is shown for reference.

$$S_\mu(k_\perp(\omega,k_z);z)=\frac{\mu_0^2N^3}{N_0^3}\frac{8}{\pi^3}\frac{k_{z*}}{k_z(k_z^2+k_{z*}^2)}\frac{N(z)f}{\omega^3}\times\left(\frac{\omega^2-f^2}{\omega^2-\omega_L^2}\right)^{1/2},\quad \omega\geq\omega_L, \tag{21}$$

where $k_z(z)=\pi jN(z)/N_0B$, $k_{z*}(z)=\pi j_*N(z)/N_0B$, and $N_0B=\int_0^D N(z)dz=10.3$ rad m/s for the buoyancy-frequency profile used here (Fig. 1). For the GM spectrum, $j_*=3$. The first factor in Eq. (21) is the relative sound-speed variance as a function of depth, which scales as N^3 (Munk and Zachariasen, 1976). Here $\mu_0^2=6.26\times10^{-8}$ is a reference relative sound-speed variance (whose value is not precisely known from AET measurements). The term $\omega_L=(f^2+N^2(z)\tan^2\theta_r)^{1/2}$ is the lowest internal wave frequency that can have a wave number perpendicular to the ray with angle θ_r (see Appendix for details). To make this explicit and to impose the internal-wave cutoff for frequencies greater than $N(z)$, Heavyside functions H are placed in Eq. (20). The spectrum without the perpendicular wave number constraint is

$$S_\mu(\omega,k_z;z)=\frac{\mu_0^2N^3}{N_0^3}\frac{4f}{\pi^2}\frac{k_{z*}}{k_z^2+k_{z*}^2}\frac{(\omega^2-f^2)^{1/2}}{\omega^3}. \tag{22}$$

Equation (22) shows that at high frequencies and mode numbers the spectrum scales as ω^{-2} and k_z^{-2}. The spectrum under the perpendicular wave number constraint [Eq. (21)], however, scales as ω^{-3} and k_z^{-3}, adding an additional ω^{-1} and k_z^{-1} dependence to the spectrum.

The second term in Eq. (20) is often referred to as the Fresnel filter. It can be considered a weighting function controlling the spectral contributions to the variances $\langle\chi^2\rangle$ and $\langle\phi^2\rangle$ at each wave number k_z. The computation of the Fresnel zone $R_{fz}(z_r(x))$ is well known and is discussed in detail in Esswein and Flatté, 1980. The Fresnel zone is the scale at which scattering can cause interference (Flatté, 1983). Furthermore, the product k_zR_{fz}, a ratio of medium sound-speed scales to acoustic scales, measures the relative importance of diffraction (see Flatté *et al.*, 1979, for a discussion of the diffraction parameter Λ); small (large) k_zR_{fz} means small (large) diffraction. The Fresnel filter as a function of k_z has its first maximum at $k_z=0$ for phase and $k_z=\sqrt{2}\pi/R_{fz}(x)$ for log-amplitude. Because the internal-wave spectrum evaluated at the perpendicular wave number is approximately proportional to k_z^{-3}, the largest contributions to the variance of phase come from large scales (i.e., small $k_z\simeq0$), while the largest contributions to the log-amplitude variance come from scales near the Fresnel zone. This interpretation is for the case where there is no waveguide and therefore only applies locally along the ray when a waveguide is present. In the waveguide case, the total effect is the integral over the entire ray path and thus represents the contribution not only from the Fresnel factors but also from the strength of the sound-speed fluctuations as a function of depth and the low frequency cutoff factor ω_L.

Figure 9 shows numerical evaluations of $z_r(x)$, $R_{fz}(x)$, and $\omega_L(x)$ for eigenrays with ray ID −3 and ID +4 for the AET environment (Fig. 1). The launch angles for the eigenrays are −9.8° and 5.3° for ID −3 and ID +4, respectively. The Fresnel zones for both eigenrays roughly follow the envelope for the constant background sound-speed case [i.e., $R_{fz}^2=\lambda x(R-x)/R$], but ray ID +4 has more structure than ID −3. This occurs because caustics are zeros of R_{fz}. The eigenray with ID +4 has gone through more turning points and therefore more caustics than the eigenray with ID −3. The eigenray with ID +4 also has a larger maximum Fresnel zone (600 m) than the eigenray with ID −3 (400 m). Thus, ray ID +4 may have contributions to the log-amplitude variance from slightly larger vertical scales than ID −3. Of critical importance, however, is the behavior of ω_L along the ray path. The eigenray with ID −3 has significantly larger values of ω_L than the eigenray with ID +4, and these large values extend over a significant region around the UTP. The eigenray with ID −3 is therefore expected to have significantly less low-frequency variability than the eigenray with ID +4, which is exactly the result from the AET observations.

TABLE III. Phase and log-intensity moments and SI for 40-period (approximately 20-min), 80-period (approximately 40-min), and 6-day observation times and as predicted by Munk and Zachariasen (1976) theory.

		ID −3			ID +4		
		σ_ϕ (rad)	σ_ι (dB)	SI	σ_ϕ (rad)	σ_ι (dB)	SI
AET	40-period	0.44 ± 0.01	0.26 ± 0.05	0.004 ± 0.002	0.44 ± 0.01	0.26 ± 0.07	0.005 ± 0.001
AET	80-period	0.61 ± 0.01	0.38 ± 0.10	0.009 ± 0.001	0.68 ± 0.01	0.40 ± 0.05	0.013 ± 0.004
AET	6-day		0.80 ± 0.07	0.044 ± 0.013		3.1 ± 0.2	0.43 ± 0.05
MZ		0.55	0.75	0.030	1.03	1.86	0.20

B. Comparison to observations

Modeled rms of log-intensity $\sigma_{\ln I}$ are 2.0 and 0.8 dB for ray ID +4 and ID −3, respectively. These values are in reasonable agreement with the observed values of 3.1 and 0.8 dB (Table III). Most of the discrepancy for ray ID +4 is due to sub-inertial variability in the observed log-amplitude (Fig. 7). The comparisons are similar if the intensity variation is quantified using SI rather than the rms of log-intensity. The modeled values of SI are 0.21 and 0.034 for ray ID +4 and ID −3, respectively, while the observed SIs are 0.44 and 0.044 (Table III).[1] Comparing rms phase variability, the modeled values of σ_ϕ are 1.49 and 0.73 rad for ray ID +4 and ID −3, respectively, while the observed values for the 80-period observation times are 0.68 and 0.61 rad (Table III). Given that the frequency spectra increase with decreasing frequency (i.e., red spectra), the total phase variance should be much larger than obtained from the 80-period observation times. Thus, the model is likely under predicting the value of the rms phase.

Figure 10 shows the observed and modeled frequency spectra of log-amplitude and phase. The two-dimensional spectra $S_{\phi,\chi}(R,\omega,k_z)$ are computed by numerical integration of Eq. (20), using Eq. (21). Integrations over vertical wave number give the modeled frequency spectra

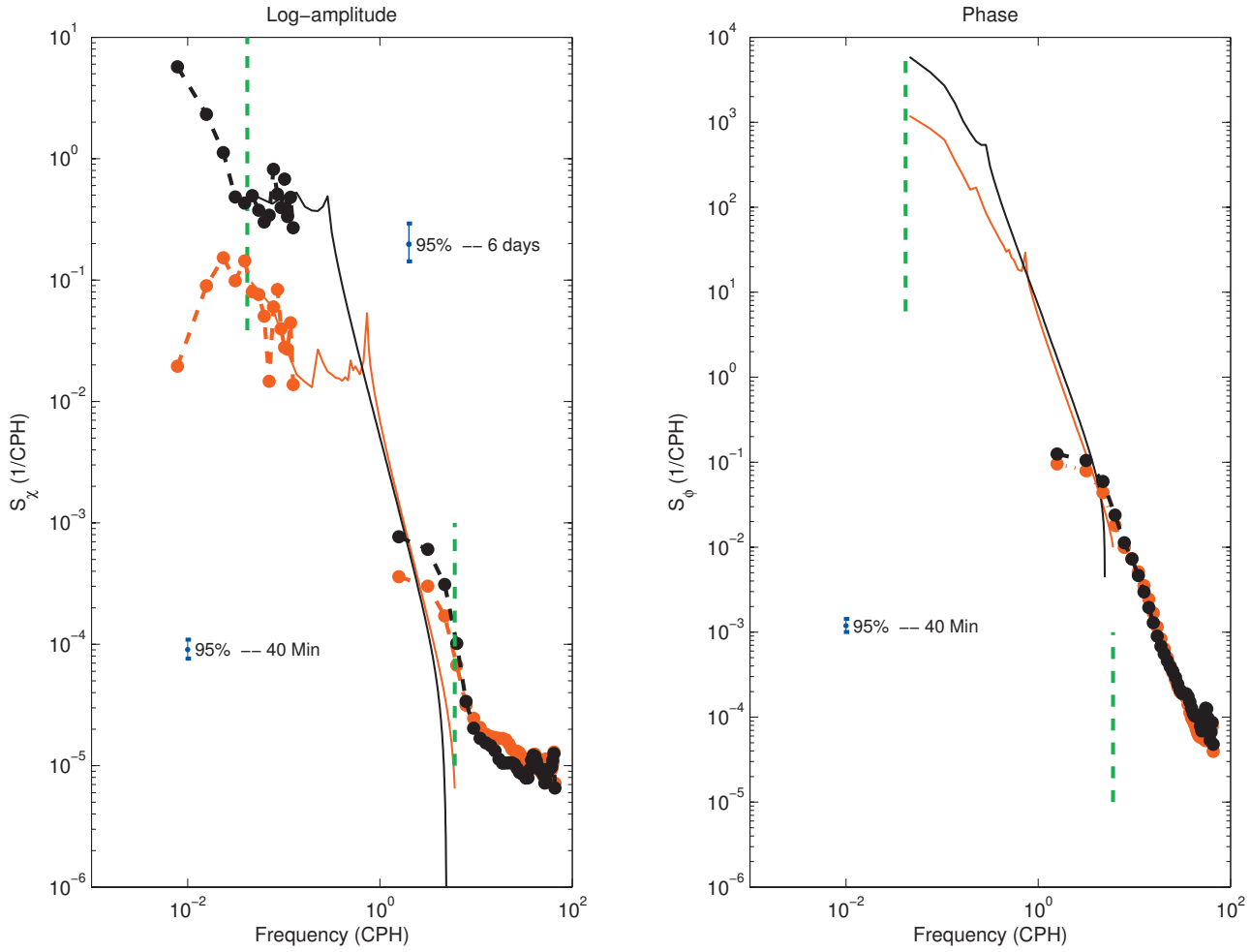


FIG. 10. Observed and predicted frequency spectra of log-amplitude (left) and phase (right) for time fronts with ray ID −3 (red) and ID +4 (black). The Coriolis frequency and the maximum buoyancy frequency (6 cph) are plotted with vertical green dashes.

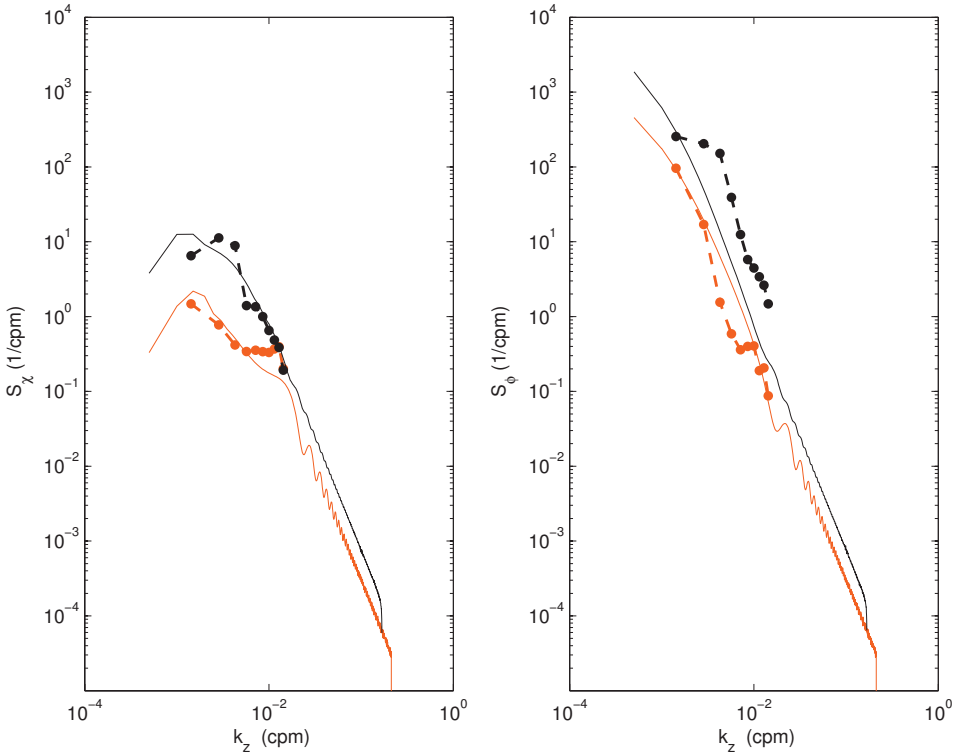


FIG. 11. Observed and predicted vertical-wave-number spectra of log-amplitude (left) and phase (right) for time fronts with ray ID −3 (red) and ID +4 (black).

$$S_{\phi,\chi}(R,\omega) = \int_0^\infty S_{\phi,\chi}(R,\omega,k_z)dk_z. \tag{23}$$

As expected, the modeled spectra cut off at the critical frequencies f and $N_{\max} \simeq 6$ cph, whereas the observations show some variability at frequencies below f (for log-amplitude) and above $N_{\max}$. The model log-amplitude spectrum for ray ID +4 has significantly more low-frequency variability than that for ID −3, in agreement with the observations. The model phase spectra also show low-frequency enhancement for ray ID +4, but no observations are available for comparison. The modeled spectra have distinct high- and low-frequency regions. At high frequencies, the spectra have very nearly a slope of ω^{-3}, and at low frequencies, the spectra are rather flat. Separating these regions is a cusp-like feature which occurs at frequencies of roughly 0.72 and 0.28 cph for ray ID −3 and ID +4, respectively. These transition frequencies corresponds to the peak values of ω_L near the ray UTPs (Fig. 9).

Figure 11 shows the observed and modeled vertical-wave-number spectra of log-amplitude and phase. The modeled spectra are obtained by analytically integrating the two-dimensional spectra over frequency (Appendix). The observed spectra are normalized so that the integral over k_z gives the log-amplitude and phase variances observed over the limited depth aperture, which are necessarily different from the variances observed over time (Colosi *et al.*, 2005). The observed and modeled log-amplitude spectra for ray ID +4 both roll off at $k_z \simeq 0.002$ cpm; this is the effect of the Fresnel filter, which attenuates contributions from low wave numbers (large vertical scales). For ray ID −3, both spectra are relatively flat from 0.01 to 0.001 cpm. The observed and modeled phase spectra are in good agreement, with both having slopes of k_z^{-3}.

V. SUMMARY AND CONCLUSIONS

Observed intensity fluctuations for two time fronts in the AET, 87-km, 75-Hz acoustic transmissions are very weak, and thus are expected to be consistent with the perturbation theory of Munk and Zachariasen (1976). The observed SI and rms of log-intensity $\sigma_{\ln I}$ over the 6 days of observations are SI=0.04 and $\sigma_{\ln I}$=0.8 dB for the time front with ray ID −3 (one UTP) and SI=0.4 and $\sigma_{\ln I}$=3 dB for the time front with ray ID +4 (two UTPs). The rays making up these two time fronts differ significantly in how they sample the upper ocean. The rays with ID −3 are steeper than those with ID +4. The average UTP depths for ID −3 (110 m) are correspondingly shallower than those for ID +4 (285 m). These different sampling properties affect the time scales of the observed intensity variability, because only internal waves whose crests are aligned with the sloping ray contribute to the acoustic variability in MZ (Munk and Zachariasen 1976). Indeed, the observed and modeled frequency spectra show that ray ID −3 has much less low-frequency variability than ID +4. This comparison of the frequency spectra for two different ray geometries provides the first direct observational evidence of the resonance that selects internal waves that are aligned with the sloping ray. In addition, the vertical wave number spectra of intensity and phase also show order of magnitude agreement with the MZ theory.

Henyey and Macaskill (1996), Colosi and the ATOC Group (1999a), and Flatté and Rovner (2000) have shown that because of ray curvature, results for straight rays cannot be strictly applied locally along the ray, as was done by MZ (Munk and Zachariasen, 1976) (Appendix). Ray curvature effects are particularly significant near the UTP depth. Subsequent to MZ (Munk and Zachariasen, 1976), the phase and log-intensity spectra were derived using the path-integral for-

malism. The path-integral results for intensity give expressions expected to be valid in both the unsaturated and partially saturated regimes (Flatté *et al.*, 1987). The path integral results have the MZ results as a limiting case in the unsaturated regime, however. In the case of the phase spectrum, the path integral and MZ predictions are the same. The conclusion is that the present comparisons with predictions made for MZ (Munk and Zachariasen, 1976) show that the resonance between the sloping ray and the internal waves whose crests are aligned with that slope has important physical meaning.

ACKNOWLEDGMENTS

The ATOC Acoustic Engineering Test (AET) was supported by the Strategic Environmental Research and Development Program through Defense Advanced Research Projects Agency (DARPA) Grant No. MDA972-93-1-0003.

APPENDIX: SPECTRA OF PHASE AND LOG-AMPLITUDE

Spectra of phase and log-amplitude are first derived in the absence of a sound channel and then generalized to the case when a sound channel is present.

1. No waveguide

The well-known result for the spectrum of phase ϕ and log-amplitude χ for a point source in the Rytov approximation (Ishimaru, 1977) is

$$S_{\phi,\chi}(R,k_y,k_z) = \pi k_0^2 R \int_0^1 d\hat{x} S_\mu(k_x=0,k_y/\hat{x},k_z/\hat{x}) \times \left[1 \pm \cos\left(\frac{(k_y^2+k_z^2)R_f^2(x)}{2\pi\hat{x}^2}\right)\right], \quad \text{(A1)}$$

where $\hat{x}=x/R$ is the normalized range, $R_f^2(x)=\lambda x(R-x)/R = \lambda R\hat{x}(1-\hat{x})$ is the Fresnel zone, and $k_0=\omega/c_0=2\pi/\lambda$ is the acoustic wave number. In this approximation, only the wave numbers perpendicular to the ray along the x-axis contribute to the acoustic fluctuations. For present purposes, consider the acoustic propagation to occur in a region which is without bound in the y-direction, but with $0<z<D$, where D is the ocean depth, and $0\le x\le R$, where R is the range. The stratification is assumed constant over the ocean depth, yielding a constant buoyancy frequency N_0. The GM three-dimensional (3D) spectrum of relative sound-speed fluctuations is

$$S_\mu(k_x,k_y,k_z) = \mu_0^2 \frac{4}{\pi^3} \frac{k_{z*}}{(k_z^2+k_{z*}^2)} \frac{k_z f}{N_0} \frac{\sqrt{k_x^2+k_y^2}}{(k_x^2+k_y^2+(k_z f/N_0)^2)^2}, \quad \text{(A2)}$$

where f is the Coriolus parameter, $k_{z*}=\pi j_*/D$, and $j_*=3$. The normalization condition gives $\mu_0^2=\int_0^\infty dk_z \int_{-\infty}^\infty dk_x dk_y \times S_\mu(k_x,k_y,k_z)$. The WKB dispersion relation typically used with the GM spectrum is

$$\omega^2 = N_0^2 \frac{k_x^2+k_y^2}{k_z^2} + f^2. \quad \text{(A3)}$$

Changing variables from k_y to ω using $S_{\phi,\chi}(R,\omega,k_z/\hat{x}) = S_{\phi,\chi}(R,k_y/\hat{x},k_z/\hat{x})(1/s)dk_y/d\omega$, with $k_y dk_y/d\omega=\omega k_z^2/N_0^2$ and $k_y=\pm(\omega^2-f^2)^{1/2}k_z/N_0$, gives

$$S_\mu(0,\omega,k_z(j)) = \frac{\mu_0^2}{\hat{x}} \frac{8}{\pi^3} \frac{k_{z*}}{k_z(k_z^2+k_{z*}^2)} \frac{fN_0}{\omega^3}, \quad \omega \ge f. \quad \text{(A4)}$$

Thus, the frequency-wave number spectrum for phase and log-amplitude becomes

$$S_{\phi,\chi}(R,\omega,k_z) = \pi k_0^2 R \int_0^1 d\hat{x} S_\mu(0,\omega,k_z/\hat{x}) \times \left[1 \pm \cos\left(\frac{k_z^2\frac{(\omega^2-f^2)}{N_0^2}R_f^2(\hat{x}) + k_z^2 R_f^2(\hat{x})}{2\pi\hat{x}^2}\right)\right]. \quad \text{(A5)}$$

The extra factor of 2 in Eq. (A4) relative to Eq. (A2) comes from the two contributions at $\pm k_y$. The variances of log-amplitude and phase are obtained by integrating over frequency and vertical wave number,

$$\langle\phi^2\rangle,\langle\chi^2\rangle = \int_0^\infty dk_z \int_f^\infty d\omega S_{\phi,\chi}(R,\omega,k_z). \quad \text{(A6)}$$

2. Waveguide

The waveguide case in which both the mean sound-speed and the buoyancy-frequency profiles are functions of the depth coordinate $[c=c(z)$ and $N=N(z)]$ has been treated by Munk and Zachariasen (1976) by applying the homogeneous condition locally along the unperturbed ray path $z_r(x)$. The stretching effect of $\hat{x}$ is ignored (i.e., the wave acts locally like a plane wave). Thus, the spectra for propagation along a ray path from a point source becomes

$$S_{\phi,\chi}(R,k_y,k_z) = \pi k_0^2 \int_\Gamma ds S_\mu(k_\perp(k_y,k_z;x);z(x)) \times \left[1 \pm \cos\left(\frac{k_y^2(k_\perp)R_{fy}^2(x) + k_z^2(k_\perp)R_{fz}^2(x)}{2\pi}\right)\right], \quad \text{(A7)}$$

where ds is an element of ray arc length. Here $R_{fy}^2(x) = \lambda x(R-x)/R$ is the Fresnel zone in the out of plane direction, $R_{fz}^2(x)$ is the vertical Fresnel zone computed according to Esswein and Flatté (1980), and the wave number that is perpendicular to the ray with angle θ is $k_\perp(z_r(x)) = (-k_z \tan\theta(z_r(x)), k_y, k_z)$. Using the WKB dispersion relation [Eq. (A3) with $N_0=N(z)$] gives $k_y(k_\perp)=\pm k_z(\omega^2-\omega_L^2)^{1/2}/N$, with $\omega_L^2=f^2+N^2\tan^2\theta$; the dispersion relation cannot be satisfied for $\omega<\omega_L$. The 3D spectrum of relative sound-speed fluctuations is

$$S_\mu(k_x,k_y,k_z;z)=\frac{\mu_0^2N^3}{N_0^3}\frac{4}{\pi^3}\frac{k_{z*}}{(k_z^2+k_{z*}^2)}\frac{k_zf}{N}\times\frac{\sqrt{k_x^2+k_y^2}}{(k_x^2+k_y^2+(k_xf/N)^2)^2}, \tag{A8}$$

where $k_{z*}=\pi j_* N(z)/N_0B$, $j_*=3$, and $N_0B=\int_0^D N(z)dz$. As before, changing variables from k_y to ω using $S_{\phi,\chi}(R,\omega,k_z)=S_{\phi,\chi}(R,k_y,k_z)dk_y/d\omega$, with $k_ydk_y/d\omega=\omega k_z^2/N^2$, gives

$$S_\mu(k_\perp(\omega,k_z);z)=\frac{\mu_0^2N^3}{N_0^3}\frac{8}{\pi^3}\frac{k_{z*}}{k_z(k_z^2+k_{z*}^2)}\frac{Nf}{\omega^3}\times\left(\frac{\omega^2-f^2}{\omega^2-\omega_L^2}\right)^{1/2},\quad \omega\geq\omega_L, \tag{A9}$$

and the frequency-wave number spectrum becomes

$$S_{\phi,\chi}(R,\omega,k_z)=\pi k_0^2\int_\Gamma dsS_\mu(\omega,k_z;z)\times\left[1\pm\cos\left(\frac{k_z^2\frac{(\omega^2-\omega_L^2)}{N^2}R_{fy}^2(x)+k_z^2R_{fz}^2(x)}{2\pi}\right)\right]. \tag{A10}$$

For short ranges such as those considered in this paper, R_{fy} is of order 100 m (Fig. 9). Because the horizontal wavelengths of internal waves are generally much larger than 1000 m, one can make the approximation

$$S_{\phi,\chi}(R,\omega,k_z)\simeq\pi k_0^2\int_\Gamma dsS_\mu(\omega,k_z;z)\times\left[1\pm\cos\left(\frac{k_z^2R_{fz}^2(x)}{2\pi}\right)\right]. \tag{A11}$$

Equation (A11) should be compared with Eq. (121) from Munk and Zachariasen (1976). Equation (A11) says that for each frequency ω and each vertical wave number k_z, the spectrum of relative sound-speed fluctuations must be integrated along the ray. However, there are forbidden regions in this integral, specifically where $\omega<\omega_L$ and $\omega>N(z)$. To specifically indicate the forbidden regions, Heavyside step functions are added to the spectrum to give

$$S_{\phi,\chi}(R,\omega,k_z)=\pi k_0^2\int_\Gamma dsS_\mu(\omega,k_z(j);z)\times\left[1\pm\cos\left(\frac{k_z^2R_{fz}^2(x)}{2\pi}\right)\right]\times H[\omega-\omega_L(z_r(x))]H[N(z_r(x))-\omega]. \tag{A12}$$

If the vertical-wave-number spectrum is needed, the frequency integral in Eq. (A12) can be done analytically, yielding

$$S_{\phi,\chi}(R,k_z)=k_0^2\frac{8}{\pi^2}\int_\Gamma ds\frac{\mu_0^2N^3}{N_0^3}\frac{k_{z*}}{k_z(k_z^2+k_{z*}^2)}\frac{Nf}{\omega_L^2}F(N,\omega_L,f)\times\left[1\pm\cos\left(\frac{k_z^2R_{fz}^2(x)}{2\pi}\right)\right], \tag{A13}$$

$$F(N,\omega_L,f)=\int_{\omega_L}^N\frac{\omega_L^2}{\omega^3}\left(\frac{\omega^2-f^2}{\omega^2-\omega_L^2}\right)^{1/2}d\omega=\frac{1}{2}\left[\left(1-\frac{\omega_L^2}{N^2}\right)^{1/2}\left(1-\frac{f^2}{N^2}\right)^{1/2}+\left(\frac{\omega_L}{f}-\frac{f}{\omega_L}\right)\times\ln\left[\frac{\left(1-\frac{f^2}{N^2}\right)^{1/2}+\frac{f}{\omega_L}\left(1-\frac{\omega_L^2}{N^2}\right)^{1/2}}{\left(1-\frac{f^2}{\omega_L^2}\right)^{1/2}}\right]\right]. \tag{A14}$$

[1]The M–Z theory does not actually predict, SI, but with the approximation that log-amplitude has a normal distribution, $SI=e^{4\langle\chi^2\rangle}-1$.

Colosi, J. A., and Flatté, S. M. (**1996**). "Mode coupling by internal waves for multimegameter acoustic propagation in the ocean," J. Acoust. Soc. Am. **100**, 3607–3620.

Colosi, J. A., and the ATOC Group (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. A. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch) (**1999**). "A review of recent results on ocean acoustic wave propagation in random media: Basin scales," IEEE J. Ocean. Eng. **24**, 138–155.

Colosi, J. A., Flatté, S. M., and Bracher, C. (**1994**). "Internal wave effects on 1000-km oceanic acoustic pulse propagation: Simulation and comparison to experiment," J. Acoust. Soc. Am. **96**, 452–468.

Colosi, J. A., Tappert, F. D., and Dzieciuch, M. A. (**2001**). "Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **110**, 163–169.

Colosi, J. A., Baggeroer, A. B., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Dushaw, B. D., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T., and Forbes, A. M. G. (**2005**). "Analysis of multipath acoustic field variability and coherence in the finale of broadband basin-scale transmissions in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1538–1564.

Colosi, J. A., Scheer, E. K., Flatt, S. M., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Worcester, P. F., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T. G., and Baggeroer, A. B. (**1999**). "Comparisons of measured and predicted acoustic fluctuations for a 3250-km propagation experiment in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3202–3218.

Duda, T. F., Flatté, S. M., Colosi, J. A., Cornuelle, B. D., Hildebrand, J. A., Hodgkiss, W. S., Worcester, P. F., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1992**). "Measured wave-front fluctuations in 1000-km pulse propagation in the Pacific Ocean," J. Acoust. Soc. Am. **92**, 939–955.

Dushaw, B. D., Howe, B. M., Mercer, J. A., Spindel, R. C., and and the ATOC Group (A. B. Baggeroer, T. G. Birdsall, C. Clark, J. A. Colosi, B. D. Cornuelle, D. Costa, B. D. Dushaw, M. A. Dzieciuch, A. M. G. Forbes, B. M. Howe, D. Menemenlis, J. A. Mercer, K. Metzger, W. H. Munk, R. C. Spindel, P. F. Worcester, and C. Wunsch) (**1999**). "Multimegameter-range acoustic data obtained by bottom-mounted hydrophone arrays for measurement of ocean temperature," IEEE J. Ocean. Eng. **24**, 202–214.

Dushaw, B. D., Worcester, P. F., Cornuelle, B. D., Howe, B. M., and Luther, D. S. (**1995**). "Baroclinic and barotropic tides in the central North Pacific Ocean determined from long-range reciprocal acoustic transmissions," J. Phys. Oceanogr. **25**, 631–647.

Ehrenberg, J. E., Ewart, T. E., and Morris, R. D. (**1981**). "Signal-processing techniques for resolving individual pulses in a multipath signal," J.

Acoust. Soc. Am. **63**, 1861–1865.
Esswein, R., and Flatté, S. M. (**1980**). "Calculation of strength and diffraction parameters in oceanic sound transmission," J. Acoust. Soc. Am. **67**, 1523–1531.
Ewart, T. E., and Reynolds, S. (**1984**). "The mid-ocean acoustic transmission experiment, MATE," J. Acoust. Soc. Am. **75**, 785–802.
Flatté, S. M. (**1983**). "Wave propagation through random media: Contributions from ocean acoustics," Proc. IEEE **71**, 1267–1294.
Flatté, S. M., and Rovner, G. (**2000**). "Calculation of internal-wave induced fluctuations in ocean acoustic propagation," J. Acoust. Soc. Am. **108**, 526–534.
Flatté, S. M., and Stoughton, R. (**1988**). "Predictions of internal wave effects on ocean acoustic coherence, travel time variance, and intensity moments for very long-range propagation," J. Acoust. Soc. Am. **84**, 1414–1424.
Flatté, S. M., Reynolds, S., and Dashen, R. (**1987**). "Path integral treatment of intensity behavior for rays in the sound channel," J. Acoust. Soc. Am. **82**, 967–972.
Flatté, S. M., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, London).
Ghilia, D. C., and Romero, L. A. (**1994**). "Robust two-dimensional weighted and unweighted phase unwrapping that uses fast transforms and iterative methods," J. Opt. Soc. Am. A **11**, 107–117.
Henyey, F., and Macaskill, C. (**1996**). "Sound through the internal wave field," *Stochastic Modeling in Physical Oceanography* (Birkhauser, Boston).
Ishimaru, A. (**1977**). "Theory and application of wave propagation and scattering in random media," Proc. IEEE **65**, 1030–1061.
Munk, W. H., and Zachariasen, F. (**1976**). "Sound propagation through a fluctuating stratified ocean: Theory and observation," J. Acoust. Soc. Am. **59**, 818–838.
Munk, W., Spindel, R., Baggeroer, A., and Birdsall, T. (**1994**). "The Heard island feasibility test," J. Acoust. Soc. Am. **96**, 2330–2342.
Munk, W. H., Worcester, P. F., and Wunsch, C. (**1995**). *Ocean Acoustic Tomography* (Cambridge University Press, London).
Rytov, S. (**1937**). "Wave and geometrical optics," C. R. Acad. Sci. URSS **18**, 263.
Van Uffelen, L. J., Worcester, P. F., Dzieciuch, M. A., and Rudnick, D. (**2009**). "The vertical structure of shadow-zone arrivals at long range in the ocean," J. Acoust. Soc. Am. **125**(6), 3569–3588
Wage, K. E., Dzieciuch, M. A., Worcester, P. F., Howe, B. M., and Mercer, J. A. (**2005**). "Mode coherence at megameter ranges in the North Pacific Ocean," J. Acoust. Soc. Am. **117**, 1565–1581.
Worcester, P. F. (**1979**). "Reciprocal acoustic transmission in a midocean environment: Fluctuations," J. Acoust. Soc. Am. **66**, 1173–1181.
Worcester, P. F., and Spindel, R. C. (**2005**). "North Pacific Acoustic Laboratory," J. Acoust. Soc. Am. **117**, 1499–1510.
Worcester, P. F., Williams, G. O., and Flatté, S. M. (**1981**). "Fluctuations of resolved acoustic multi-paths at short range in the ocean," J. Acoust. Soc. Am. **70**, 825–840.
Worcester, P. F., Cornuelle, B. D., Hildebrand, J. A., Hodgkiss, W. S., Duda, T. F., Boyd, J., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1994**). "A comparison of measured and predicted broadband acoustic arrival patterns in travel time-depth coordinates at 1000-km range," J. Acoust. Soc. Am. **95**, 3118–3128.
Worcester, P. F., Cornuelle, B. D., Dzieciuch, M. A., Munk, W. H., Colosi, J. A., Howe, B. M., Mercer, J. A., Spindel, R. C., Metzger, K., Birdsall, T., and Baggeroer, A. B. (**1999**). "A test of basin-scale acoustic thermometry using a large-aperture vertical array at 3250-km range in the eastern North Pacific Ocean," J. Acoust. Soc. Am. **105**, 3185–3201.
Worcester, P. F., Howe, B. M., Mercer, J. A., Dzieciuch, M. A., and the Alternate Source Test (AST) Group (T. G. Birdsall, B. M. Howe, J. A. Mercer, K. Metzger, R. C. Spindel, and P. F. Worcester) (**2000**). "A comparison of long-range acoustic propagation at ultra-low (28 Hz) and very-low (84 Hz) frequencies," in *Proceedings of the US-Russia Workshop on Experimental Underwater Acoustics*, edited by V. I. Talanov (Institute of Applied Physics, Russian Academy of Sciences, Nizhny Novgorod), pp. 93–104.

Tracking blue whales in the eastern tropical Pacific with an ocean-bottom seismometer and hydrophone array

Robert A. Dunn and Olga Hernandez[a)]
Department of Geology and Geophysics, University of Hawaii, Manoa 1680 East-West Road, Honolulu, Hawaii 96822

(Received 1 September 2008; revised 21 May 2009; accepted 29 May 2009)

Low frequency northeastern Pacific blue whale calls were recorded near the northern East Pacific Rise (9 °N latitude) on 25 ocean-bottom-mounted hydrophones and three-component seismometers during a 5-day period (November 22–26, 1997). Call types A, B, C, and D were identified; the most common pattern being ~130–135 s repetitions of the AB sequence that, for any individual whale, persisted for hours. Up to eight individual blue whales were recorded near enough to the instruments to determine their locations and were tracked call-by-call using the B components of the calls and a Bayesian inversion procedure. For four of these eight whales, the entire call sequences and swim tracks were determined for 20–26-h periods; the other whales were tracked for much shorter periods. The eight whales moved into the area during a period of airgun activity conducted by the academic seismic ship R/V *Maurice Ewing*. The authors examined the whales' locations and call characteristics with respect to the periods of airgun activity. Although the data do not permit a thorough investigation of behavioral responses, no correlation in vocalization or movement with airgun activity was observed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158929]

PACS number(s): 43.30.Sf, 43.80.Nd [RAS] Pages: 1084–1094

I. INTRODUCTION

The blue whale (Balaenoptera musculus) populates all of the world's oceans, forms vocally distinct groups, and has long-range seasonal migrations. Because of historic whaling pressure, it is considered endangered throughout its range and has been protected internationally since 1965 (Yochem and Leatherwood, 1985). Common to all blue whales is emission of high intensity, low frequency, and long duration acoustic calls in repetitive patterns, possibly used for communication (e.g., Stafford *et al.*, 1999, 2001; Thompson *et al.*, 1996; McDonald *et al.*, 2006). Owing to their high source levels (189 dB re 1 μPa at 1 m) and low frequencies (14–100 Hz), these calls can be detected at up to 200 km distance on bottom-moored hydrophones (Širović *et al.*, 2007). Blue whales are also highly vocal, producing distinct amplitude- and phase-modulated calls in repetitive patterns that allow tracking of individual animals, which is of prime importance to detailed behavioral studies. In an environment where individuals are often dispersed, passive acoustic methods are effective means to study their presence, movements, and calls.

From November 9–28, 1997, an array of ocean-bottom seismometers and hydrophones was deployed along the northern East Pacific Rise to collect seismic data during an active-source seismic study of the magma chambers beneath this volcanic system (Fig. 1). We recently examined the data for whale calls and found that blue whale calls were recorded during 5 of 20 days of recording. We present an analysis of calls from eight blue whales and track the whales using their calls and a localization algorithm based on a probabilistic grid search method. We obtained long, complete call sequences for four of the eight whales, and were able to track their motions for 20–26 h intervals. Reconstructions of continuous swim tracks of individual blue whales lasting more than just a few hours are rare (e.g., Watkins *et al.*, 2004).

The whales entered the area while the academic seismic ship R/V *Maurice Ewing* carried out a seismic experiment using a 20-gun, 139-l airgun source. The proximity of the whales and ship allowed us to examine their calls and swim tracks for any anomalous behavior with regard to airgun use. Sounds produced by airgun arrays have garnered increasing interest as there are concerns regarding the potential impact of airgun noise on marine mammals (Malakoff, 2001, 2002; National Research Council, 2003).

II. INSTRUMENTATION AND DATA

The study site is located in the eastern tropical Pacific Ocean, 750 km southwest of Mexico's coastline, in 2600–3200 m of water, and is centered on a section of the northern East Pacific Rise. An array of ocean-bottom seismometers and hydrophones was deployed over a 200-km-section of the mid-ocean ridge with a minimum station spacing of 12 km (Dunn *et al.*, 2001). Not all instruments recorded blue whale calls; those that did record calls are shown in Fig. 1. The instruments consisted of a mix of ocean-bottom receivers from the Woods Hole Oceanographic Institution: 9 ocean-bottom hydrophones (OBHs), 3 ocean Reftek in a ball (ORB) equipped with a hydrophone, and 13 Office of Naval Research three-component seismometers (OBSs) equipped with a hydrophone. We used both the hydrophone and vertical component seismometer data for this study. Recordings were made with a sampling rate of 200 Hz for the OBH and

[a)]Present address: MEMMS (Marine Ecosystem Modeling and Monitoring by Satellite), CLS, Satellite Oceanography Division, 8-10 rue Hermès, 31520 Ramonville, France.

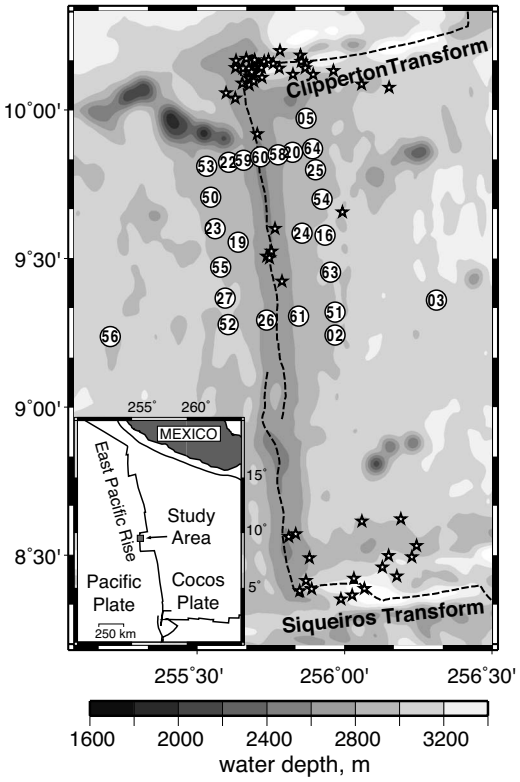


FIG. 1. Seafloor bathymetry map showing location of the East Pacific Rise (darker shading down the center of the figure; the dashed line indicates the axis of the ridge) and locations of the instruments that recorded blue whale calls (numbered circles). The instruments are a mix of ocean-bottom receivers from the Woods Hole Oceanographic Institution: 3 ORB (numbers 2–5) equipped with a hydrophone; 9 OBH (numbers 16–27), and 13 OBS (numbers ≥50) equipped with a hydrophone. Stars indicate epicenters of 58 locatable earthquakes, out of >580 total that occurred in the region during the study period.

ORB and 128 Hz for the OBS. For these instruments the useful band for acoustic detection of blue whales is between 5 and 60 Hz.

All 20 days of the seismic records were examined. Blue whale calls were detected from November 22–26, 1997 on subsets of the ocean-bottom stations, depending on the location of each whale with respect to the instrument array (i.e., only the closest instruments recorded the whale calls with large enough signal-to-noise ratio for analysis). Vocalizing blue whales entered the area during the latter part of the seismic experiment, after 13 days of intermittent airgun activity (Fig. 2). No other marine mammals were identified in the data. Apart from blue whale calls and the *Ewing* activities, the most common signals recorded were regional earthquakes, which tend to occur in high numbers along mid-ocean ridges and impart significant acoustic energy to the water column. Over the 20-day recording period the instruments detected more than 580 distinct earthquake events.

Throughout the seismic study, the position and speed of the R/V Maurice Ewing were digitally logged every minute and information on the status of the airgun array was logged at the time of each airgun pulse. The airgun array consisted of 20 bolt airguns that varied in volume from 145 to 875 in.^3 for a total discharge volume of 8503 in.^3 (~139 l); at the source, the airgun output was 237 dB (re 1 μPa P-P at 1 m). This is the effective output of the airgun array, as if the energy emanated from a point source. Because the array is spread over a large area, the actual output is much lower. Towed 40 m behind the ship and at 10 m depth, the array generated acoustic pulses every 210 s (150 s on November 25) as the ship traveled a pattern within the instrument network. The airgun array is designed to focus energy downward, rather than to the sides, and there is an azimuthal variation of the energy emission, with the highest levels emitted fore and aft of the ship and significantly less energy emitted to the sides.

Under current guidelines, the National Marine Fisheries Services defines the radii around airgun sources with received sound levels of 180 dB as a safety radii for cetaceans (NMFS, 2005); the radii with received levels of 160 dB are considered to be distances within which some cetaceans are likely to be subject to behavioral disturbance (NMFS, 2005). With regard to the Ewing's airgun array, theoretical calculations (Diebold, 2004) and field calibration studies (Tolstoy *et al.*, 2004) show that sound levels produced by the array depend on the depth of observation. Therefore, received levels at the whale will depend not only on the distance from the ship but the depth of the whale. Studies of dive characteristic of blue whales off the central California coast (Lagerquist *et al.* 2000) show that 72% of all dives are between 0 and 16 m and less than 1 min duration; the second most frequent dive interval is 97–152 m, accounting for 15% of all dives and <1.2% of the whale's total time underwater. Blue whales seldom dive to 150 m depth and even more rarely to greater depths. Theoretical calculations for the array used in this experiment (Diebold, 2004) indicate that peak sound levels of ≥180 dB (re 1 μPa rms) occur within 250 m of the array

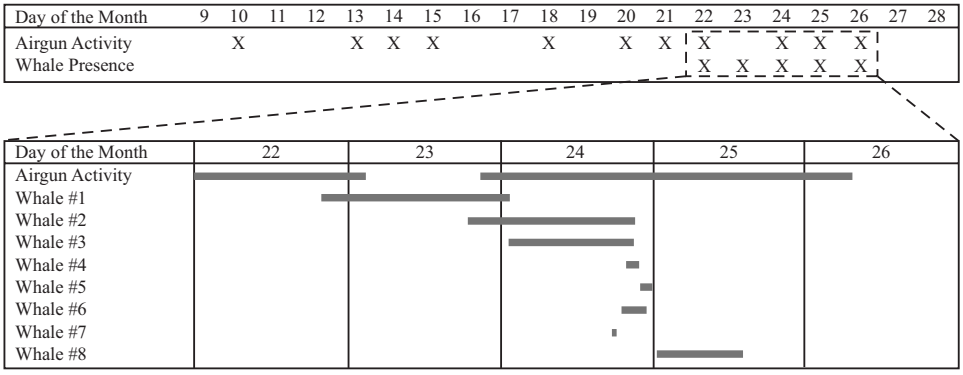


FIG. 2. Time line of events showing periods of airgun activity and periods when whale calls were detected and tracked. The numbering of the whales used here is consistent with a numbering scheme used throughout the text. The whales appeared in the study area after 13 days of on-and-off airgun activity.

at 10 m depth below the sea surface and within 500 m of the array at 125 m depth; received levels of ≥160 dB (re 1 μPa rms) within 600 m of the array at 10 m depth and within 1.75 km at 125 m depth. At 10 km distance, received levels are calculated to be 133 dB at 10 m depth and 136 dB at 125 m depth. A field calibration study (Tolstoy *et al.*, 2004) of the Ewing 20-gun array (with an ~8600 in.^{3} volume array rather than the 8503 in.^{3} array used here) found 180 dB (re 1 μPa rms) levels within ~1 km of the array and 160 dB levels within ~2–3 km of the array (recording depths were 18 and 500 m). At distances greater than 5 km, the received energy was <145 dB, with the majority of the energy in the 5–100 Hz range.

III. TRACKING METHOD

We adopt a Bayesian inversion method for locating the whales (Tarantola and Valette, 1982). The basic idea is to define the solution in terms of an *a posteriori* probability density that incorporates the data (onset times of vocalizations), the model parameters (vocalization position), and the theoretical link between model parameters and calculated synthetic data. From this, the most likely location of the whale is given as the position where the *a posteriori* probability is a maximum. About the maximum likelihood point, a full representation of the 95% probability region for the whale's location is easily extracted from the probability density.

The method involves a brute-force grid search comparing arrival time measurements of whale calls to predicted arrival times for all locations on a spatial grid. The synthetic arrival times are calculated to all grid points in advance and stored for later use, greatly improving the efficiency of the method. The benefits of this method are numerous: it allows for non-linear travel time calculations, it is consistent with respect to a change of variables, it allows for general error distributions in the data, it incorporates theoretical errors that arise from inaccurate parametrizations and theoretical simplifications, it allows for the formal incorporation of any *a priori* information concerning the location parameters (such as a probability distribution for the whale's location derived from an estimate of its position at an earlier time), and it provides a full representation of the probability of a whale's location. In short, the Bayesian inversion method is flexible and provides a mathematically robust location of a whale's location given noisy, sparsely recorded data.

The unknowns in the problem are the spatial coordinates of a whale (x,y), where x is longitude and y is latitude. Given the large station separation (>3 times the water depth), the data are incapable of resolving accurate whale depths. Given that vocalizing blue whales tend to spend the majority of their time near the ocean surface (Oleson *et al.*, 2007), we simply assume that the whale is located at the surface when calculating its lateral position. In practice, we define a grid over the model space $\mathbf{m}=(\mathbf{x},\mathbf{y})$, which is the area of the ocean for which we might find the whale. Because of transmission loss and detection thresholds, any recorded whale will be within ~50 km of the nearest station. Thus, we used a uniform 200×200 km^2 grid centered on the stations. The general solution (without the depth and time dependence) is

$$P(\mathbf{m}) = K\rho(\mathbf{m})\exp\left\{-\sum_{i=1}^{N}\frac{|t^i_{\text{obs}} - t^i(\mathbf{m})|}{\sigma_i}\right\}. \quad (1)$$

The values t^i_{obs} are the N observed arrival times of a single whale call minus a weighted average of all such times [Tarantola and Valette, 1982; Eqs. 10–12]. Measurement of the arrival times can be made by several different methods and is a critical step in the location problem. The best results were obtained by band pass filtering the data to isolate the fundamental frequency of the B call (or the stronger 48 Hz overtone for the case of whale 3), lying the filtered time series over the spectrogram of the call, and then handpicking the onset of the B call from the joint time series and spectrogram plot. Picking onset times is generally less accurate than cross-correlation (e.g., Nosal and Frazer, 2006), but in our case the overlap in both the time and frequency domains of airgun pulses with the whale calls makes cross-correlation impractical. The 1-σ pick uncertainties, ranging from 0.1 to 6 s, are the largest source of error in the location problem but are included in the location method and weight their respective measurements.

The values $t^i(\mathbf{m})$ in Eq. (1) are the theoretical travel times from a grid location in $\mathbf{m}$ to each station for which exists a value t^i_{obs} minus a weighted average of these times [Tarantola and Valette, 1982; Eqs. 10–13]. The travel times were calculated using an algorithm that at distances far from a seismic station (greater than approximately three times the water depth) mimics T-phase propagation: acoustic energy travels along a direct path at a constant acoustic speed. At distances closer to a seismic station, it is important to account for the water depth of the seismic station and our algorithm models the acoustic propagation along a direct path from the whale (at the sea surface) to the station (on the seafloor). Given the experiment geometry (sparse station layout with large separation) and large pick errors, a more accurate acoustic propagation model would not result in appreciably better whale locations.

The values σ_i are a combination of observational and theoretical uncertainties: $\sigma^2=\sigma_t^2+\sigma_T^2$, where σ_t are the uncertainties corresponding to each measured time, t^i_{obs}, derived from the picking procedure, and σ_T are the uncertainties in the travel time calculations from a grid point in $\mathbf{m}$ to a recording instrument. The σ_T values include the uncertainty of the instrument positions, the uncertainty due to the method of travel time calculation, and the uncertainty of the acoustic medium. The instrument coordinates and uncertainties were determined via an inverse procedure using the travel times of the airgun pulses from the ship (whose position is accurately known via global positioning system) to the instruments. The 1-σ errors of the instrument positions are 3–150 m, depending on the amount and distribution of data available for each instrument. The error due to the acoustic path calculations and to unknown deviations in the acoustic velocity is <1 s.

By adding each of the variances of the different error sources, the total expected uncertainty in the theoretical calculation is ~1 s.

$\rho(\mathbf{m})$ consists of any *a priori* information that may exist on the whale's position (other than the travel time measurements) before we calculate an estimate of the position. If no such information exists, then initially the whale has equal probability of being anywhere on the grid and $\rho(\mathbf{m})=1/M$, for all values of $\mathbf{m}$, where M is the number of grid points. Thus a summation of $\rho(\mathbf{m})$ over the total model space yields a value of 1 (100% probability that the whale is somewhere on the grid, with equal probability at all locations). On the other hand, given the speed at which a whale can swim, we could state that the position of the whale at the (i+1)th call must be close to that at the *i*th call. In this case, we could write the *a priori* information function for the whale's position as

$$\rho(\mathbf{m}) \propto \exp\left\{-\frac{1}{2}[\mathbf{m}-\langle\mathbf{m}_i\rangle]^T\left(\frac{1}{R^2}\right)[\mathbf{m}-\langle\mathbf{m}_i\rangle]\right\},$$

where $\langle\mathbf{m}_i\rangle$ is the estimated position at the previous call and R (units of distance) is the product of the average speed of the whale and the amount of time lapsed since the previous position was estimated. In other words, we establish *a priori* a Gaussian probability density such that there is a ~95% probability of the whale being within $2R$ of the previously estimated position. We used this approach because it tends to smooth the track of the whale, which is otherwise noisy due to the large pick uncertainties. We also post-processed the tracks with a three-point averaging filter to further reduce spurious call-to-call position noise.

Equation (1) is normalized by the constant K such that the probability of the whale being somewhere within the grid is 100%. Since the model parameters, $\mathbf{m}$, are discrete, K is defined as

$$K=\left(\sum\sum P'(x,y)\right)^{-1},$$

where P' are the un-normalized values from Eq. (1).

The probability density $\mathrm{P}(\mathbf{m})$ provides a full representation of the probability of a whale's coordinates. The maximum likelihood position of the whale is the position where $\mathrm{P}(\mathbf{m})$ is maximum and is thus the position where the weighted data misfit is a minimum [in the case of constant $\rho(\mathbf{m})$]. Our method uses the L_1 norm to quantify misfit length, because a solution is thereby less biased by outliers in the data. The shape of the distribution, which is not necessarily elliptical about the maximum likelihood position, provides a "map" of the uncertainty of the whale's position. Using a running algorithm over the time series of all identified calls, we calculated whale locations when at least four call observations were available on separate receivers; we rejected any locations when the misfit,

$$\Phi=\frac{1}{N}\sum_{i=1}^{N}\frac{|t^i_{\mathrm{obs}}-t^i(\mathbf{m})|}{\sigma_i},$$

exceeded a value of 1.5 or the location uncertainty exceeded 3 km.

IV. BLUE WHALE CALLS

Comparisons of the calls in our data with those of other studies reveals that our records are from a northeastern Pacific population of blue whales (e.g., Stafford *et al.*, 1999); a typical spectrogram, showing the ACB call components of northeastern Pacific blue whales, is shown in Fig. 3(a). A few D calls were also recorded [Fig. 3(b)] (e.g., Thompson *et al.*, 1996; Aroyan *et al.*, 2000; Mcdonald *et al.*, 2001). On ORB03, D calls are present November 23 around 2100 and 2230 GMT and again on November 25 around 2300 GMT. On ORB02, D calls are present from 1800 GMT November 24 to 0700 GMT November 25.

Of the whales studied here, the A calls have durations that last 20–30 s and the duration of the B calls is approximately 15–20 s (Fig. 4). The time between the onset of the A call and the onset of the B call is variable from whale-to-whale and between the calls emanated by any one whale and tends to be in the 50–60 s range. The duration of C calls is about 12 ± 1 s. We suggest caution when examining call durations in this and other data sets, since multipathing of the acoustic energy tends to elongate the apparent duration of calls in both the time series and spectrograms, and this elongation will be environmentally dependent. Furthermore, the rise time of the A call is very slow and A call duration measurements will be inaccurate for distant, noisy records. However, there are some anomalies in the calls particular to individual whales allowing them to be identified separately from other whales. One whale (whale 1) exhibits a brief discontinuity in its B calls and a pulse in the overtones at the end of its A calls [Fig. 3(a)]. Another whale (whale 3) exhibits a shortened A call rapidly followed by a 7–8 s un-modulated ~16 Hz tone (Fig. 4), which could be considered a separate call, but here we refer to both parts of this call as an "anomalous" A call. Some whales exhibit shorter A call durations and some longer, but this is not sufficient to identify individuals since multiple whales can exhibit one of the two durations, the durations are not necessarily constant across all calls of a single whale, and there is a bias toward measuring shorter times as the whale moves further from the recording instrument and signal-to-noise decreases.

Typically the calls appear in sequences of A and B combinations. The C call was nearly always recorded when a whale was close to an instrument, but generally undetectable at other times due to its low amplitude. Therefore, it is to be understood that, unless specific reference is made to the C call, a C call may have been present but is omitted from the discussion. The most common call pattern in the data is a sequence of AB calls (i.e., ABABAB $\cdots$). Only whale 3 deviated from this pattern, by forming repetitions of one A call followed by more than one B call, such as repetitions of ABB or ABBB (each B call is preceded by a C call in this case). Consecutive A calls were not recorded and each sequence starts with an A call. These patterns were repeated regularly, often for many hours. Using a 95% confidence interval ($\alpha=0.05$), the A-to-B and B-to-B spacings of calls were not statistically distinguishable between night and day. In a few rare instances, an A call was followed by a short silence, rather than a B call.

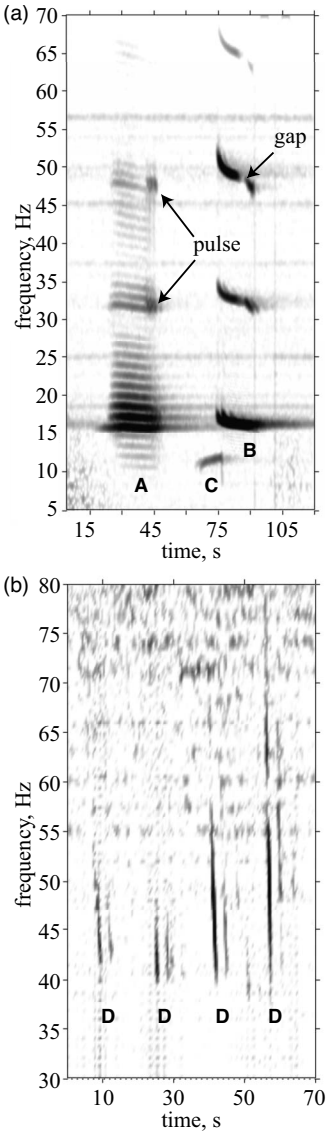


FIG. 3. (a) Spectrogram of an ACB sequence of calls constructed from a stack of 66 individual spectrograms of such calls that occurred after 0800 (GMT) on November 23 (whale 1). The A call begins with ~5–8 s of un-modulated 16 Hz signal that is not always readily apparent in distant noisy recordings and is followed by a train of amplitude modulated short pulses with a fundamental carrier frequency of 16 Hz and at least two harmonics at 32 and 48 Hz. Each pulse includes multiple frequency-offset non-harmonic components. The pulses are not obvious in the spectrogram, but they can be seen in the time series data (Fig. 4). The A call is slightly down-swept in frequency and the 32 and 48 Hz overtones of the A call often terminate with a short 2–3 s pulse, which is likewise smeared in time by the stacking. The low amplitude precursor to the B call, denoted a C call, consists of an upward sweeping call from ~10.5–11.5 Hz. The B call is characterized by a fundamental downward-swept (frequency-modulated) sound from ~17 to 15.5 Hz; a second harmonic that sweeps down from ~34 to 31 Hz; a strong third harmonic that sweeps down from ~52 to 46.5 Hz; and a fourth harmonic that sweeps down from ~68.5 to 62 Hz. The downward sweeping B tones exhibit faint, but persistent, "ghosts" that follow the main pulses by ~2–2.5 s. These ghosts are likely caused by acoustic energy that traveled secondary paths to the instruments (first and second water column multiples). Higher frequency components are expected for northeastern Pacific blue whales (e.g., Thompson *et al.*, 1996; McDonald *et al.*, 2001), but not recorded by our instruments. Narrow vertical lines on the spectrogram are internal instrument noise, the horizontal bands are ship traffic noise. (b) Example spectrogram of D calls (not stacked). These calls have a ~1 s duration, down sweeping from 80 Hz or less to about 40 Hz, and only occurred when more than one whale was present. Each direct arrival of the D call is followed by a fainter arrival at a time expected for the first water column multiple of the acoustic energy.

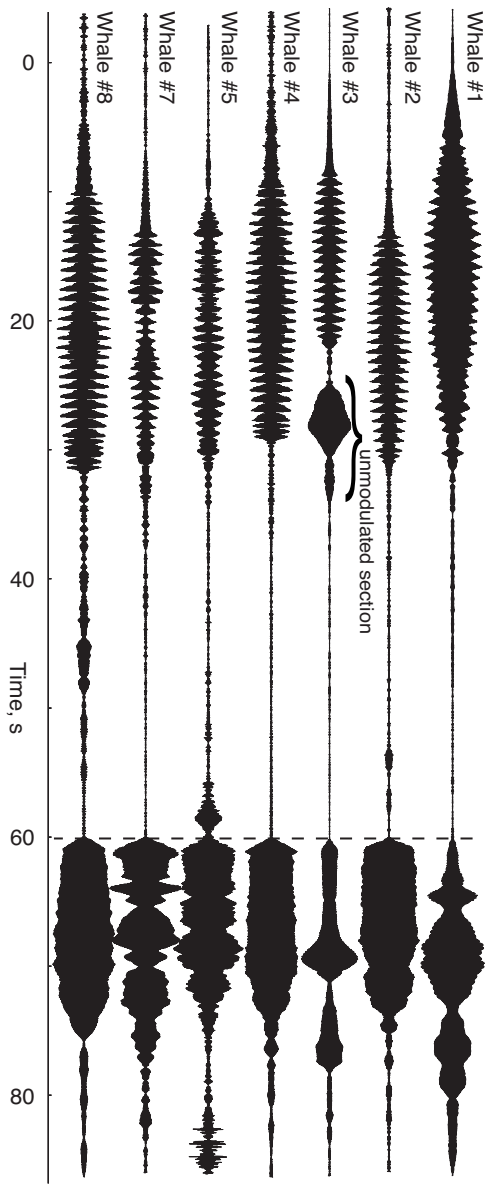


FIG. 4. Time series of the AB calls of blue whales located by this study. These records have been narrow band pass filtered (12–22 Hz) to isolate the fundamental component of the calls. The amplitude modulation is apparent in each of the A calls. The character of the calls changes considerably from call to call due to constructive and destructive interference of the direct path and one or more reflected paths, an effect that changes with the distance of the whale from the receiver and also presumably with the whale's depth (surface and bottom reflected wave interference) and orientation (radiation pattern). However, there are fundamental characteristics of the calls that persist across all calls of a single whale but differ between whales. Specifically, the duration of the A call tends to vary between whales (for example, compare whale 1 to whale 2) and whale 3 exhibits a short modulated A call followed by a separate 10 s un-modulated ~16 Hz call and then the B call. Other distinguishing characteristics may be seen in the overtones (not shown) and in the spectrograms of the calls (Fig. 3).

V. WHALE TRACKING

Over a period of 2.5 days, and after 13 days of on and off airgun activity, up to eight blue whales entered the area of the seismic experiment during continued airgun operations. Figure 2 and Table I summarize basic call detection and tracking information for these whales. While most of the time the whales were located just outside of the seismic array, making detection and tracking difficult, we were able to

TABLE I. Whale tracking summary.

Whale	Start date:time (GMT)	End date:time (GMT)	Number of calls located	Minimum distance traveled (km)	Mean speed over course (km/h)	Min/mean/max distance to airguns (km)
1	Nov 22 1997: 2005	Nov 24 1997: 0053	417	190	6.5	37/62/90
2	Nov 23 1997: 1916	Nov 24 1997: 2136	263	112	4	28/51/97
3	Nov 24 1997: 0200	Nov 24 1997: 1958	287	60	3	15/47/87
4	Nov 24 1997: 2012	Nov 24 1997: 2133	19	4	3	74/78/84
5	Nov 24 1997: 2201	Nov 24 1997: 2342	15	12	7	74/75/78
6	Nov 24 1997: 1910	Nov 24 1997: 2250	6	17	5	61/64/83
7	Nov 24 1997: 1816	···	1	···	···	33/33/33
8	Nov 25 1997: 0045	Nov 25 1997: 1410	68	53	4	72/76/84

identify calls and track some whales over multi-hour intervals. We suggest that individual whales were tracked, rather than multiple whales traveling together because the calls were regularly spaced (i.e., no out-of-sequence calls), not detectably dissimilar, and, perhaps most importantly, did not overlap with other calls emanating from the same location. Having said that, it cannot be ruled out that when one whale stopped vocalizing another whale, located near the same spot, took up where the first left off; or that non-vocalizing whales traveled together with the one vocalizing whale. In one case (whale 3), the A call of the whale is very anomalous as is the B call pattern, providing further support that in that particular case only one individual was tracked. In one or two cases, a whale that was tracked may have been a whale that had been previously identified and tracked over an earlier period of time (based on whale locations and detection times).

A. Blue whale 1

In the final hours of November 22, whale 1 was detected on western stations. By 2005 GMT, as the ship was finishing a seismic line to the north, this whale moved close enough to the seismic stations to be located [Fig. 5(a)]. The whale traveled southeast during the next day, crossing the array. In the final hours of November 23 the airgun activity recommenced southwest of the whale's position. At that point whale 1 continued its easterly heading until 0100 GMT on November 24, when it exited the area to the east and was no longer recorded on sufficient instruments to be located. The distance moved between any two calls is often similar to or smaller than the 1-σ uncertainty of the location, so is difficult to accurately measure the whale's detailed motions and instantaneous velocity. Examining the point-to-point path of the whale, over the 29-h period the whale traveled ~200 km at an average speed of ~6–7 km/h. This is only a rough approximation of the whale's true speed, since the calculated whale track tends to be noisy due to the picking uncertainties; nonetheless it is a typical speed and distance for cruising or migrating whales (Mate *et al.*, 1999). The distance between the whale and the ship during airgun operations was never less than ~37 km (sound levels <145 dB) and there are no detectable changes in the whale's heading nor speed upon the stopping and restarting of the ship's airguns.

We were able to monitor all calls [Fig. 6(a)] from whale 1 over a 24-h period beginning at approximately 0000 GMT on November 23. Throughout this 24-h period, AB calls were repeated semi-regularly; no other call sequence was formed. Repeated sequences of AB calls occur at 135 ± 5 s intervals (all call interval and gap times are measured from the onset of one call to the onset of the next call), with no statistically relevant change from day to night. There are often gaps of both small and large nature that interrupt the repeated AB sequences [Fig. 7(a)]. The majority of gaps are small, between ~160 and 340 s. While it has been suggested that small gaps may represent respiration times (Cummings and Thompson, 1971; Mcdonald *et al.*, 2001), the small gaps in the calls of whale 1 do not repeat at regular intervals and there are 45–60 min intervals when the spacing between calls does not exceed 150 s (or 15 s longer than the main repeat interval), suggesting that a larger call spacing is not required for breathing. Small gaps in the call sequences appear after some of the T-phases of regional earthquakes. Some of these gaps are real, but others may be due to masking of whale calls by the intense broadband earthquake energy. Furthermore, there are other small gaps of this nature in the call pattern that are not preceded by T-phases and several T-phase recordings not followed by gaps. Therefore, there is no obvious correlation between gaps in the call sequences and earthquake T-phases. The *largest* gap in the sequences is ~40 min and the two largest gaps containing no more than one or two AB calls correspond to the time intervals of approximately 1645–1745 and 1900–2015 GMT (0945–1045 and 1200–1315 local time, respectively) on November 23.

Near the end of the 24-h period, the airgun activity recommenced within the seismic array when the whale was located ~90 km from the airgun source. At that distance, sound pressure levels from the airguns are expected to be relatively low (Tolstoy *et al.*, 2004; Diebold, 2004), but seismic instruments near the whale did record the airgun pulses and it is conceivable that the whale detected them as well (airgun pulses occur within the vocalization band of blue whales). There is a small gap just after the first airgun pulses and a larger gap of ~20 min after the airguns had been powered up to full volume. After that, the AB pattern repeats as usual. While the correlation of the call gaps with the airgun activity is of interest, it is not possible to make any causative judgments about these gaps, since similar gaps are

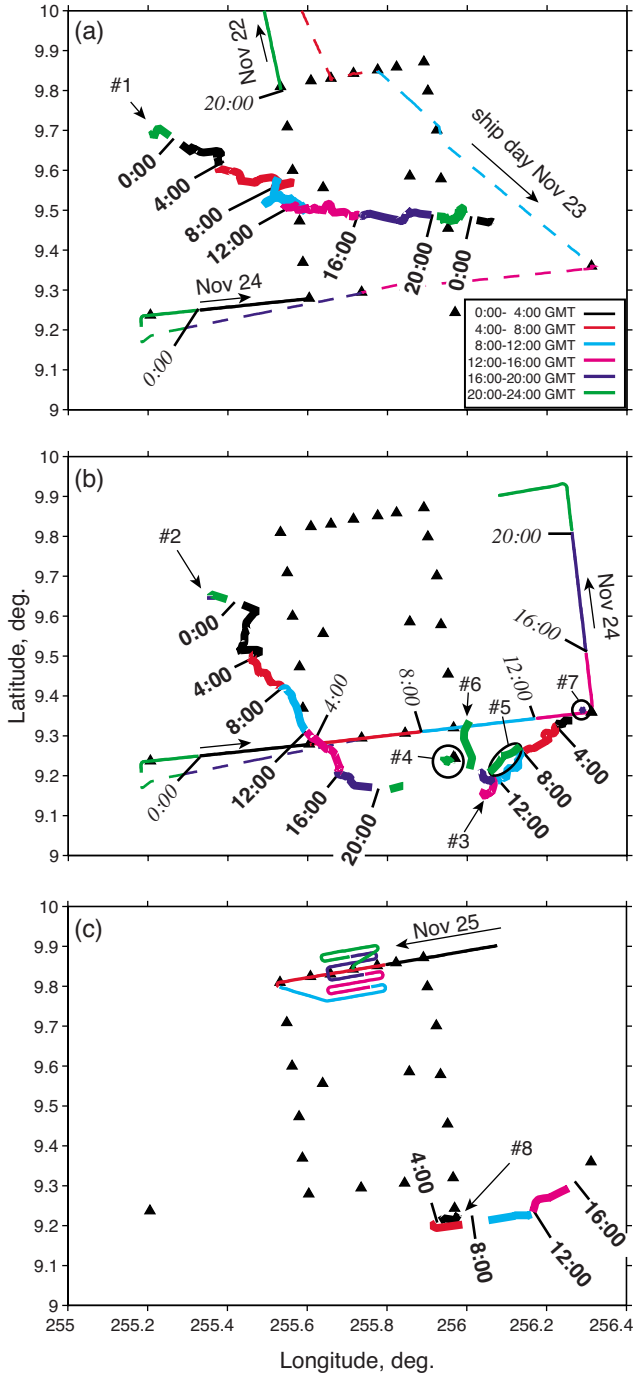


FIG. 5. (Color online) Whale swim tracks for (a) November 23, (b) November 24, and (c) November 25. The number of the whale given on the figure corresponds with the numbers used in the text. Thin dashed (no airgun activity) and solid (airgun activity) lines indicate the ship tracks. Time marks (GMT) are indicated on the tracks in bold (whale times) and italics (ship time). The whale and ship tracks are also color coded, as indicated in panel (a), by 4-h periods.

found throughout the 24-h period. Using high signal-to-noise ratio calls recorded on the station closest to the whale, we compared mean amplitudes of whale calls that occurred 3 h before and 3 h after the airguns restarted and found no statically meaningful difference (i.e., using a 95% confidence interval, we cannot reject the null hypothesis that the amplitudes before and after airgun startup are the same). A similar analysis for when the airguns shutdown earlier in the day has poor resolution because whale 1 was located far from any stations at that point and the calls all have low signal-to-noise ratios; in any case, there was no obvious change in call amplitude. We also compared the mean time intervals of the calls both before and after airgun startup and likewise found no meaningful difference.

Over its entire swim track, whale 1 may have traveled alone, as no overlapping or out-of-sequence calls were recorded that would indicate an accompanying whale or whales. Some of the outlying stations did record more distant whales. The most prominent examples being whale 2 who moved into the area of the study from the west at the end of the 24-h period as whale 1 exited the area to the east, and at least two whales in the vicinity of ORB03 (far eastern region) after 1300 GMT. On ORB03, repeated patterns of AB, ABB, ABBB, ABBBB, and even one clear ABBBBB were recorded, many of them overlapping in time with each other and/or the AB calls of whale 1. We were able to track one of these whales, whale 3, as it moved nearer to the main array on November 24.

B. Blue whale 2

On November 23 at 1900 GMT blue whale 2 approached close enough to the array from the west to be recorded on multiple instruments. At that time its location was determined to be near the location where whale 1 had entered the area 24-h previously [Fig. 5(b)]. Thereafter, whale 2 traveled southeast at ~4 km/h on average until ~2000 GMT at which time it turned east and subsequently stopped vocalizing (last known call occurs at 2136 GMT). The first few locatable calls from this whale occurred just before airgun startup on November 23, thereafter all monitoring of this whale occurred during airgun activities. An analysis of call amplitude changes, upon startup of the airguns, could not be made because the whale was located far from any stations at that point and the calls have low signal-to-noise ratios (scatter in the amplitudes are too large to allow for a meaningful test). Later in the day, the closest distance between the whale and ship was 28 km. There is no indication that the whale tended to avoid the ship, but rather it assumed a heading that crossed the ship's path (aft of the ship at 65 km distance).

We were able to monitor all calls from whale 2 over a 20-h period beginning at approximately 0000 GMT on November 24. Throughout the 20-h period, AB calls [Fig. 6(b)] were repeated semi-regularly. Repeated sequences of AB calls are predominantly spaced at 129 ± 5 s time intervals [Fig. 7(b)], with irregularly occurring gaps that are twice (260 s) and three times (390s) the fundamental spacing. The character of the gaps is thus much different from that of whale 1 (and other whales in this study). During this 20-h period, whale 2 may have traveled alone, as no other calls (overlapping or out of sequence) were recorded that would indicate an accompanying whale or whales. Near the end of the day of November 24, when vocalizations ceased, whale 2 was last detected approaching the positions of at least three other whales in the southeast corner of the study area.

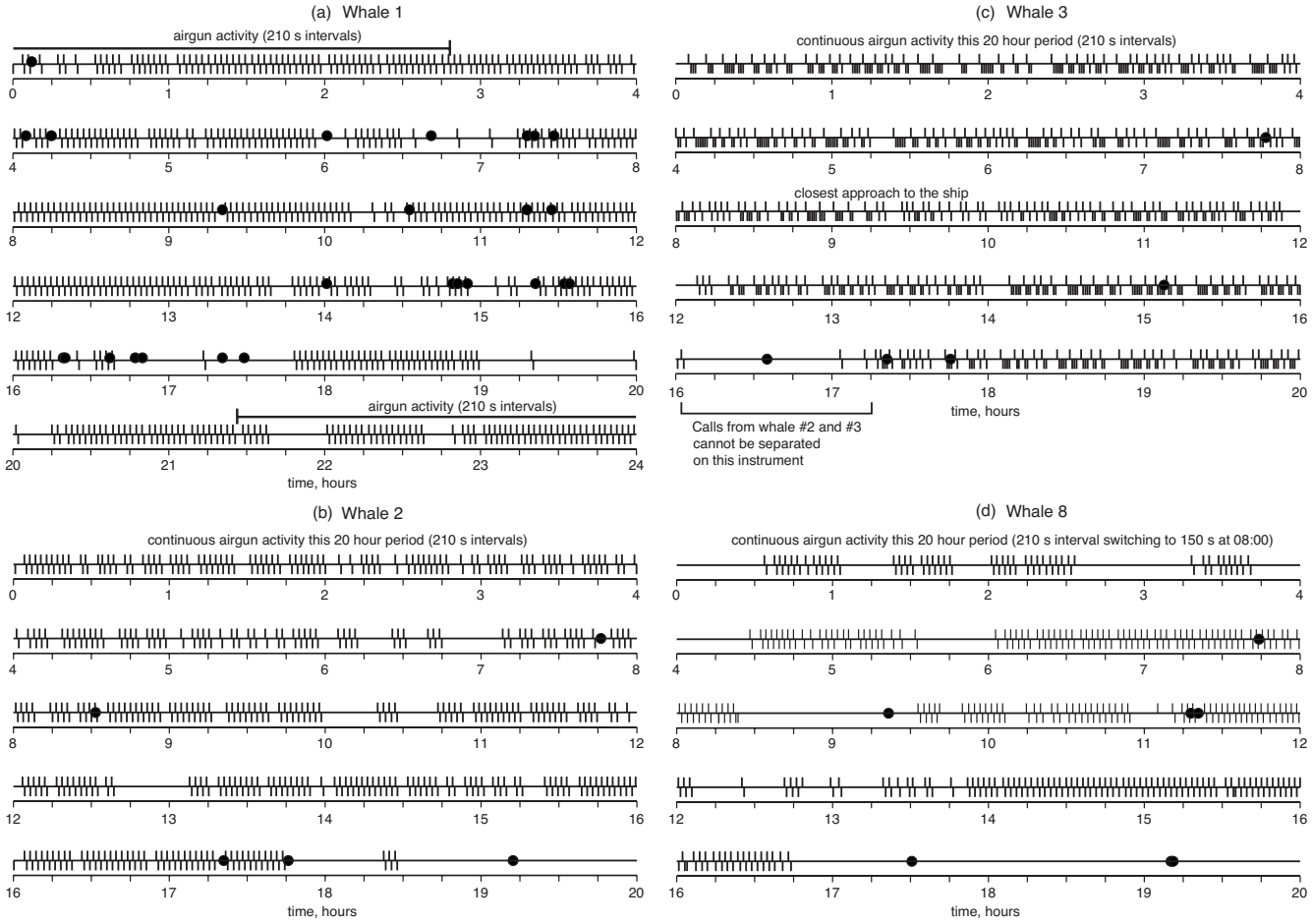


FIG. 6. Call sequences of whales 1, 2, 3, and 8 [subplots (a)–(d), respectively] for 20- to 24-h periods. Ticks above the centerline indicate A calls, and ticks below the center line indicate B calls. C and D calls are not shown. Calls were determined by using seafloor instruments that were located closest to the whale (as indicated on the plots) as well as additional instruments when calls from other whales or airgun pulses masked calls. The black dots indicate earthquake T-phase events. Whales 1, 2, and 8 vocalizations consist almost exclusively of repeated AB sequences. Whale 3 vocalizations consist of AB and A multiple-B sequences in seemingly random order. For whale 1, during the first 30 min of this 24-h period a few AB calls may have been missed, due to a low signal-to-noise ratio. Periods of airgun activity are indicated on each plot.

C. Blue whale 3

On November 24 at 0100 GMT whale 3 approached close enough to the array from the east to be recorded on multiple instruments and its location was determined [Fig. 5(b)]. This whale entered the area near where whale 1 exited the area. We know that whales 1 and 3 are distinct, because the calls of the two whales overlapped during the early hours of November 24. Furthermore, whale 3 exhibits an anomalous A call (Fig. 4) with an A-multiple-B pattern, in contrast to the generic AB calls of whales 1 and 2. The anomalous A call and distinct call sequences also help identify this whale later on November 24 when other AB whale calls (whales 2, 4, 5, and 7) were present in the data.

We reconstructed the entire call sequence for whale 3 over the period 0000 GMT to 2000 GMT on November 24 [Fig. 6(c)]. No obvious groupings or patterns occur within the sequences of calls. For example, there is no apparent pattern to the number of B calls that follow the A calls. Over the 20-h of recordings, the largest break in the sequences is only ~16 min [the large gap at 1600 GMT in Fig. 6(c) is due to masking of calls by another whale] and there are no obvious diurnal changes in the calls.

During the entire period that we tracked whale 3, the ship carried out airgun activities. Whale 3 initially traveled southwestward, passing within ~15 km of the ship, which was heading in the opposite direction; several hours later it turned north and then ceased vocalization. As it passed the ship, there was no obvious heading change that could be construed as an attempt to avoid the ship, such as reported by McDonald *et al.* (1995) for a blue whale approaching a seismic ship to within ~10 km. At 15 km distance from the ship the received sound levels are <145 dB (re 1 μPa rms) (Tolstoy *et al.* 2004; Diebold, 2004), less than what is expected to elicit a behavior response in some marine mammals (NMFS, 2005).

Beginning at about 1000 GMT the calls of whale 2 begin to overlap with whale 3's calls on stations near whale 3; whale 2 was located ~60 km to the northeast at that time. Later, around 1800 GMT, a third whale (whale 6) is detected by its repeated D calls that overlap the AB calls of whale 3; that whale's position was initially determined to be ~20 km to the north of whale 3. At ~2000 GMT whale 4 was suddenly vocally active at a position ~9 km to the north of

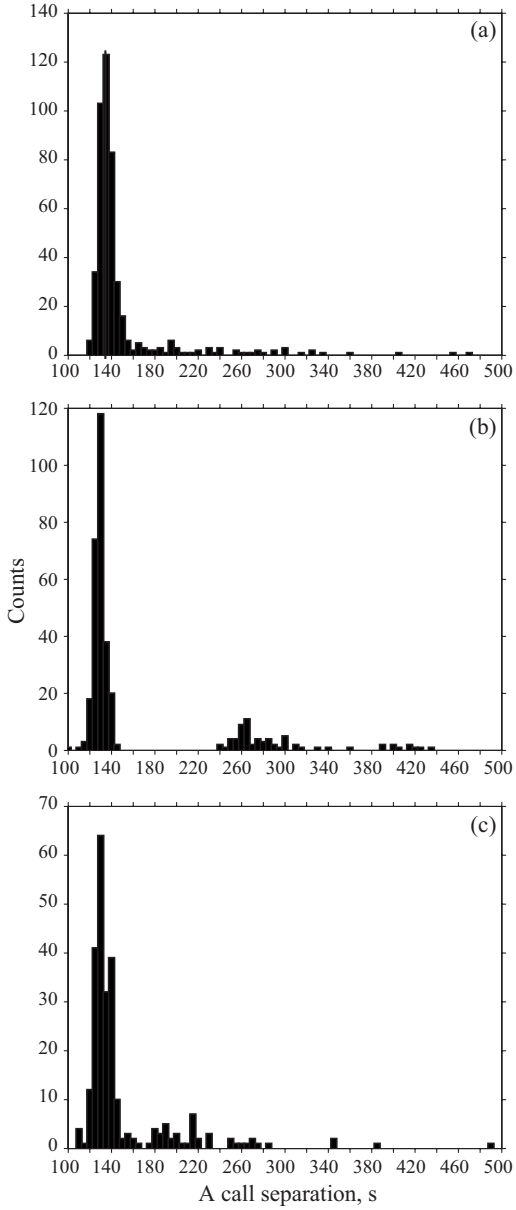


FIG. 7. Histograms (5 s bins) showing the time gap distribution between successive A calls for whales 1, 2, and 8 [subplots (a)–(c), respectively]. Data are for the calls shown in Fig. 6; a time difference was determined as the time between the onset of each A call. The histograms show that the most frequent spacing between calls is 130–135 s. While larger time gaps tend to be random, whale 2 (b) exhibits time gaps of ~260 and ~390 s, which are two and three times the size of the fundamental gap at 130 s. For clarity, time gaps greater than 500 s are not shown. Some of the scatter in the histograms is due to errors in identifying the onset of the A call.

whale 3. Once whale 4 became vocally active, whale 3 was suddenly quiet and its anomalous-A-multiple-B pattern was not detected again.

D. Blues whales 4, 5, 6, and 7

On November 24 whale 4 became vocally active at ~2000 GMT and whale 5 became vocally active after 2145 GMT. Both whales had moved silently into the southeast region of the study, but one of them may have been whale 1 whose last known position was just to the north. We were able to track their calls for a few hours, during which time whale 4 remained in one area and whale 5 moved southward [Fig. 5(b)]. Whales 4 and 5 exhibited AB call sequences; neither whale exhibited the anomalous A call of whale 3. During the time we were able to locate them, the ship (with airguns in use) operated ~60 km north of the whales' positions. Calls of whale 2 overlap in time with those of whale 4, indicating that they are distinct whales. Calls of whale 2 do not overlap with those of whale 5, which began vocalizing identifiable calls only 10 min after the last recognizable call of whale 2, but there may have been some earlier calls of whale 5 that were masked by those of whales 2 and 4. 36 km separates whale 2 from whale 5, indicating that these were also distinct whales.

D calls (whale 6) were recorded from 1800 GMT November 24 to 0700 GMT November 25 on stations in the southeast quadrant of the study. These calls overlap with those of whales 2, 3, 4, 5, and 8 on those stations, indicating the presence of another distinct whale. Owing to a poor signal-to-noise ratio, the swim track could only be determined for a short period between 1900 and 2300 GMT on November 24 as the whale moved southward.

For a short time earlier in the day (1500–2000 GMT) another blue whale, whale 7, was recorded near ORB3. The ship may have passed closely by whale 7, but since we detected and located it several hours after the ship had left the area, the proximity between the ship and whale cannot be established. The difficultly in locating this whale was caused by a low signal-to-noise ratio due to its position well outside of the main array, noise from the airgun source, and a persistent masking of its calls by those of whale 3.

E. Blue whale 8

In the early hours of November 25, at least two and possibly three blue whales were recorded on stations near ORB02 in the southeast quadrant of the study area. These calls likely were made by whales 2, 4, and 5, who were last detected in this general vicinity, but probably not whale 3, since the calls were purely of AB combinations and did not exhibit the anomalous A call of whale 3. Due to the overlapping nature of the calls and weak signals, it was not possible to locate each of these whales, but one whale did have a strong signal-to-noise ratio and we were able to track it for many hours as it moved from ORB02 eastward out of the array at a speed of ~5 km/h [Fig. 5(c)]. Although we designated it whale 8, it may be one of the whales previously identified. Throughout the day, the ship performed airgun work 72–84 km to the north of this whale.

We were able to monitor all calls from whale 8 over a 20-h period beginning at approximately 0000 GMT on November 25. Throughout the 20-h period, AB calls [Fig. 6(d)] were repeated semi-regularly; except for the rare ABB call or A-only call, no other call sequence was formed. Repeated sequences of AB calls are spaced at 130 ± 6 s time intervals, but like other whales there are larger gaps of random size [Fig. 7(c)]. During the 20-h period, whale 8 may have traveled alone, as no overlapping or out-of-sequence calls from the same location were recorded that would indicate an accompanying whale or whales. Near the end of the day of November 25, whale 8 approached ORB03 at the eastern-

most extreme of the study area. At least two other whales (one of which may have been whale 7) were vocalizing in the area (probably within 10 km to the east of ORB03 based on call amplitudes). We lost track of whale 8 as its calls overlapped with those of the other whales and as it moved far enough away from the main array to no longer be recorded on enough stations to determine its location. For the remainder of the day and throughout the next day, its calls and the calls of at least one other whale were recorded (faintly) on ORB03, indicating that these whales remained within ~40 km of the eastern edge of the experiment area.

VI. DISCUSSION AND CONCLUSIONS

The blue whales detected by this study are members of a vocally distinct population of blue whales that inhabit the northeast Pacific, ranging from the Gulf of Alaska to a region off Central America (e.g., Stafford *et al.*, 2001; Stafford, 2003; McDonald *et al.*, 2006). Individuals found in the Gulf of California are also thought to be part of this group (Calambokidis *et al.*, 1990; Thompson *et al.*, 1996). Although details of the migration routes and numbers of whales that migrate are poor, a hydrophone study detected members of the northeastern Pacific blue whale population year round in the eastern tropical Pacific (Stafford *et al.*, 1999, 2001). Therefore, it is not unexpected that the blue whales detected by our study in the month of November are members of the northeastern Pacific population.

Six out of eight whales formed closely spaced, repeated AB call sequences; the exceptions are whale 3 who formed sequences of A calls followed by up to six B calls and whale 6 who was tracked by its D calls. While we cannot be assured that an individual whale's call behavior will not change over time, some whales in this study did exhibit anomalous call components that repeated with each call: for example, the gap in the B call of whale 1 [Fig. 3(a)] and the anomalous A call of whale 3 (Fig. 4). Thode *et al.* (2000) also detected distinguishing characteristics of blue whale B calls that allowed them to identify individual whales. This suggests that in the future it may be possible to track some individual whales via fixed hydrophone arrays call-by-call for extended periods of time, even in the presence of other blue whales. Other than the AB call patterns that tended to be spaced every 130–135 s, we found no other regular call patterns or repetitions other than a tendency for whale 2 to exhibit a gap between AB calls that were two and three times its fundamental call spacing [Fig. 7(b)]. While there appears to be longer gaps in the latter parts of the 24-h periods of observation, this may be due to the presence of other whales at those times or foraging behavior (see below) rather than diurnal behavioral variations.

Four of the whales exhibited long (20–26 h or more) repetitive AB vocalization sequences (whales 1, 2, 3, and 8). A recent study by Oleson *et al.* (2007) indicates that repetitive AB calling sequences are characteristic of lone, migrating males, rather than foraging whales or whales in groups. We were able to reconstruct the swim tracks and time series of almost every call made by these whales as they passed through the area. Although the determination of detailed whale movements and dive depths was not possible, we were able to obtain general locations with an uncertainty of 1–2 km. The whales tended to travel alone or at least without other vocalizing whales; for hours on end there were no overlapping or out-of-sequence calls from closely spaced whales. The calls that are present tend to be evenly spaced, with an interval time typical of an individual whale. The whales also tended to travel long distances, 100 km or more over a 24-h period. Average swim speeds were ~3–7 km/h over the course of monitoring. These distances and swim speeds may be typical of blue whales that are migrating or cruising, but not foraging (Mate *et al.*, 1999). In summary, the long swim tracks of these four whales and their AB calling behavior indicate that these individuals were lone, migrating males.

On November 24, while airgun activity continued ~60–80 km away, several whales congregated in the southeast corner of the study area near ORB02 (whales 2, 3, 4, 5, 6, and 8). Such clustering behavior is indicative of foraging (Mate *et al.*, 1999). The presence of other whales also had an obvious correlation with changes in calling behavior: mainly a cessation of calling or long pauses between calls. For example, (1) as whales 2 and 3 moved into this area they ceased vocalizations, (2) whales 4 and 5 were vocally active for only short periods of time in this area, and (3) whale 8 became vocally active only as it left this area. Also notable is whale 6 (possibly the previously identified whale 1), who passed through this area while producing only D calls. In the study of Oleson *et al.* (2007), D calls were heard from both sexes during foraging, commonly from individuals within groups. Our observations, taken together with those from the study of Oleson *et al.* (2007), suggest that lone traveling males moved into this area, subsequently ceased most AB call sequences, and perhaps spent some time foraging, whether or not females were present is unknown.

For whales 1 and 2, the instruments recorded calls both during airgun activity and between airgun activity (Fig. 2). At times of starting or stopping airgun activity, these whales were located tens of kilometers from the airgun source (whale 1: 69 km at airgun shutdown and 90 km at airgun startup; whale 2: 42 km at airgun startup) and we did not detect corresponding changes in swim tracks or call behavior. For whale 1, a 20-min gap in calls occurred after the airguns became active, but many gaps occurred in the call sequences throughout the day—both during and not during airgun activity—so no causative relationship is supported. There is no indication that the whales attempted to time calls to fall between airgun pulses. The AB calls are generally spaced every 130–135 s, while the airgun pulses mainly occurred every 210 s so that the calls moved in and out of the spaces between airgun pulses. We also examined the call sequences for any anomalous behavior due to the presence of earthquake acoustic energy. Earthquakes produce significant water column energy in the frequency band used by blue whales and can mask whale calls for tens of seconds, but we found no obvious correlation between the many earthquake events that occurred during the monitoring and changes in calling behavior (changes in duration and timing). Presum-

ably, blue whales are accustomed to such high-amplitude sounds, as they occur frequently along mid-ocean ridges.

During airgun operations, airgun pulses were recorded across the entire seismic array and were thus presumably detectable by all eight whales. Overall we found no anomalous behavior that could be directly ascribed to the use of the airguns, though it should be reemphasized that the average distance from airgun source to the whales was tens of kilometers (Table I). For whale 3, who approached the ship to within about 15 ± 2 km (the closest of any whale), the call patterns and the whale's heading exhibit no detectable changes. Since the whales were not closer than ~ 15 km to the ship, and usually much farther away, sound levels produced by the Ewing's airguns and experienced by the whales are expected to be less than 145 dB (re 1 μPa). Under current guidelines, the National Marine Fisheries Services defines the radius about the ship with received sound levels of 160 dB as distances within which some cetaceans are likely to be subject to behavioral disturbance (NMFS, 2005). While this study found no behavioral response to the airgun activity, and hence supports these guidelines, further studies with more detailed observations are warranted.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation, Ocean Sciences Division, under Grant No. OCE0224903. We thank John Diebold for the theoretical estimates of the Ewing airgun array output levels and M. Carolina Anchieta for the earthquake locations shown in Fig. 1; Carolina located these events during an undergraduate summer internship at UH. Olga Hernandez contributed to this work during an internship at the University of Hawaii (that formed part of her Prédoctorat Programme at the Ecole Normale Supérieure de Paris). We also thank two anonymous reviewers for their careful consideration of the manuscript.

Aroyan, J. L., McDonald, M. A., Webb, S. C., Hildebrand, J. A., Clark, D., Laitman, J. T., and Reidenberg, J. S. (**2000**). "Acoustic models of sound production and propagation," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), Vol. **12**, pp. 409–469.

Calambokidis, J., Steiger, G. H., Cubbage, J. C., Balcomb, K. C., Ewald, C., Kruse, S., Wells, R., and Sears, R. (**1990**). "Sightings and movements of blue whales off central California 1986–1988 from photo-identification of individuals," Rep. Int. Whal. Comm. **12**, 343–348.

Cummings, W. C., and Thompson, P. O. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.

Diebold, J. (**2004**). "Modeling marine seismic source arrays," Lamont Doherty Earth Observatory, http://www.ldeo.columbia.edu/res/fac/oma/sss/NMFS_Lamont_airgun+modeling.pdf (Last viewed January 2009).

Dunn, R. A., Toomey, D. R., Detrick, R. S., and Wilcock, W. S. D. (**2001**). "Continuous mantle melt supply beneath an overlapping spreading center on the East Pacific Rise," Science **291**, 1955–1958.

Lagerquist, B. A., Stafford, K. M., and Mate, B. R. (**2000**). "Dive characteristics of satellite-monitored blue whales (*Balaenoptera musculus*) off the central California Coast," Marine Mammal Sci. **16**, 375–391.

Malakoff, D. A. (**2001**). "Roaring debate over ocean noise," Science **291**, 576–578.

Malakoff, D. A. (**2002**). "Suit ties whale death to research cruise," Science **298**, 722–723.

Mate, B. R., Lagerquist, B. A., and Calambokidis, J. (**1999**). "Movements of North Pacific blue whales during the feeding season off Southern California and their southern fall migration," Marine Mammal Sci. **15**, 1246–1257.

McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data," J. Acoust. Soc. Am. **109**, 1728–1735.

McDonald, M. A., Hildebrand, J. A., and Webb, S. C. (**1995**). "Blue and fin whales observed on a seafloor array in the Northeast Pacific," J. Acoust. Soc. Am. **98**, 712–721.

McDonald, M. A., Mesnick, S. L., and Hildebrand, J. A. (**2006**). "Biogeographic characterization of blue whale song worldwide: Using song to identify populations," J. Cetacean Res. Manage. **8**, 55–65.

National Research Council (**2003**). *Ocean Noise and Marine Mammals* (National Academy Press, Washington, DC).

NMFS (**2005**). "Small takes of marine mammals incidental to specified activities: Marine seismic survey of the Aleutian Islands in the North Pacific Ocean/notice of issuance of an incidental take authorization," Fed. Regist. **70**, 901–913.

Nosal, E.-M., and Frazer, L. N. (**2006**). "Track of a sperm whale from delays between direct and surface-reflected clicks," Appl. Acoust. **67**, 1187–1201.

Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.

Širović, A., Hildebrand, J. A., and Wiggins, S. M. (**2007**). "Blue and fin whale call source levels and propagation range in the Southern Ocean," J. Acoust. Soc. Am. **122**, 1208–1215.

Stafford, K. M. (**2003**). "Two types of blue whale calls recorded in the gulf of Alaska," Marine Mammal Sci. **19**, 682–693.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**1999**). "An acoustic link between blue whales in the Eastern Tropical Pacific and the Northeast Pacific," Marine Mammal Sci. **15**, 1258–1268.

Stafford, K. M., Nieukirk, S. L., and Fox, C. G. (**2001**). "Geographic and seasonal variation of blue whale calls in the North Pacific," J. Cetacean Res. Manage. **3**, 65–76.

Tarantola, A., and Valette, B. (**1982**). "Inverse problems=quest for information," J. Geophys. **50**, 159–170.

Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," J. Acoust. Soc. Am. **107**, 1286–1300.

Thompson, P. O., Findley, L. T., Vidal, O., and Cummings, W. C. (**1996**). "Underwater sounds of blue whales, *Balaenoptera musculus*, in the Gulf of California, Mexico," Marine Mammal Sci. **12**, 288–293.

Tolstoy, M., Diebold, J. B., Webb, S. C., Bohnenstiehl, D. R., Chapp, E., Holmes, R. C., and Rawson, M. (**2004**). "Broadband calibration of R/V *Ewing* seismic sources," Geophys. Res. Lett. **31**, L14310.

Watkins, W. W., Daher, M. A., George, J. E., and Rodriguez, D. (**2004**). "Twelve years of tracking 52-Hz whale calls from a unique source in the North Pacific," Deep Sea Res. Part I **51**, 1889–1901.

Yochem, P. K., and Leatherwood, S. (**1985**). "Blue whale—Balaenoptera musculus (Linnaeus, 1758)," in *Handbook of Marine Mammals*, edited by S. H. Ridgeway and R. Harrison (Academic, London), Vol. **3**, pp. 193–240.

The contrast-source stress-velocity integral-equation formulation of three-dimensional time-domain elastodynamic scattering problems: A structured approach using tensor partitioning

Adrianus T. de Hoop
Laboratory of Electromagnetic Research, Delft University of Technology, 2628 CD Delft, The Netherlands

Aria Abubakar[a] and Tarek M. Habashy
Schlumberger-Doll Research, 1 Hampshire Street, Cambridge, Massachusetts 02139

(Received 18 March 2009; revised 19 June 2009; accepted 22 June 2009)

The contrast-source stress-velocity integral-equation formulation of three-dimensional time-domain elastodynamic scattering problems is discussed. A novel feature of the formulation is a tensor partitioning of the relevant dynamic stress and the contrast source volume density of deformation rate. The partitioning highlights several features about the structure of the formulation. These can advantageously be incorporated in a computational implementation of the method. An application to the case of a scatterer composed of isotropic material and embedded in an isotropic elastic background medium shows that the corresponding newly introduced constitutive coefficients are more natural as a characterization of the media than the traditional Lamé coefficients.
 [DOI: 10.1121/1.3179672]

PACS number(s): 43.20.Fn, 43.20.Bi, 43.20.Px [RLW] Pages: 1095–1100

I. INTRODUCTION

In different regimes of wave scattering, the contrast-source integral-equation formulation with kernels resulting from the Green's functions of the background medium has proven to be a versatile tool for analyzing and computing the associated wave phenomena.[1] As compared with the scattering of acoustic waves in fluids and the scattering of electromagnetic waves, where in the formulation only scalar and vector quantities are involved, the scattering of elastic waves in solids brings in the extra feature of the presence of tensorial quantities of rank 2, viz., the dynamic stress and the volume source density of deformation rate[2] (Sec. 15.9). As is known from the theory of static elasticity[3] (pp. 91–91), such tensors admit a decomposition into their omnidirectional, their symmetrical deviatoric, and their anti-symmetrical constituents. The importance of this decomposition is that the three constituents can be shown to be mutually orthogonal in the function space they span. This orthogonality property is, in its turn, bound to be of importance both in the characterization of the elastic properties of scatterer and its background medium and in the computational implementation of the relevant field integral representations, leading to contrast-source as well as wavefield integral equations. For a wide class of elastic properties of the media involved a complete partitioning of the governing coupled wave equations and integral relations accompanies the decomposition of the dynamic stress and the contrast volume source density of deformation rate. In particular, this holds for media that are isotropic in their elastic behavior. For such media it is shown that the elastic constitutive coefficients that are associated with the partitioning turn out to be more natural as a characterization of their wave properties than the traditional Lamé coefficients.

This paper starts with recalling the generic form of the three-dimensional (3D) time-domain dynamic stress–particle velocity coupled elastodynamic wave equations and the corresponding Green's function type wavefield integral representations. Next, the scattering problem is formulated in terms of its contrast volume source distributions. Subsequently, the dynamic stress and the contrast volume source density of deformation rate (both of which are symmetric tensors of rank 2) are decomposed into their omnidirectional and (symmetric) deviatoric constituents. The consequences of this decomposition for the elastic wave equations and the Green's tensors in the relevant field integral representations are investigated. Having discussed the general aspects of the procedure, the application to scatterers consisting of isotropic material present in an isotropic elastic embedding is presented. Here, the relevant compliance coefficients turn up that, for elastic wave dynamics, prove to be more natural than the traditional Lamé coefficients. Finally, the analytic expressions for the elastodynamic Green's tensors for a homogeneous, isotropic medium are given.

The theory is presented in its full generality of linearized elastodynamics. Inhomogeneity and arbitrary anisotropy are included in the constitution of the solid. As to the inertia properties of the medium a symmetrical inertia tensor of rank 2 is introduced that generalizes the volume density of mass as it applies to isotropic solids and can accommodate the presence of preferred-direction oriented heterogeneities in a macroscopic mixture theory (spatial averaging over representative elementary domains) that can be used on a scale that the interrogating pulsed elastic waves can sense. The anisotropic elastic properties are represented in the compliance, a symmetrical tensor of rank 4. The theory shows that, in scat-

[a]Author to whom correspondence should be addressed. Electronic mail: aabubakar@slb.com

tering, the compliance is a more natural quantity to characterize inhomogeneities than the traditional stiffness tensor. It is important to notice that in the use of the contrast-source formulation in inverse scattering problems (detection of all sorts of inhomogeneities in a structure under elastodynamic interrogation) it is the values of these constitutive coefficients that one is after. The analysis presented shows which fundamental combinations of these coefficients manifest themselves already in the structure of the governing wave equations. Also, the relevant Green's tensors that occur in any of the wavefield's integral representations (which are at the heart of any wave scattering formulation) are constructed out of these combinations and explicitly show how they occur in the particle velocity and/or the dynamic stress of the generated wave motion and thus are accessible to measurement. This feature can help in the design of measurement setups in the search of particular constitutive parameters of the scattering objects. In the processing of the measured scattering data, the property of "orthogonality in function space" of the different constituents associated with the tensor partitioning is expected to behave rather "independently." This could serve as a guideline to the design of processing software as far as filtering of noise is concerned.

As the case of a homogeneous, isotropic solid already shows, the wavespeeds of the different propagating wave constituents are related to certain combinations of the constitutive coefficients associated with the tension partitioning. This implies that arrival time measurements and elastodynamic ray-tracing techniques[4] yield additional information to measured values of pulse shapes (amplitudes, rise times, time widths, and frequency of oscillation of ringing pulses) and can be used as such in inverse scattering parameter extraction methods.

Computational algorithms for modeling elastodynamic wave motion in heterogeneous and anisotropic structures are notoriously complicated and computationally, time consuming. The tensor partitioning method offers itself as a tool for breaking up any algorithm for the wave quantities as a whole into subroutines applying to the separate constituents, after which the latter's interaction is programed separately. Such a procedure would make the computer code more transparent in its relation to the underlying physics as well as more efficient as far as computation time is concerned.

II. DESCRIPTION OF THE CONFIGURATION AND FORMULATION OF THE PROBLEM

Position in the configuration is specified by the coordinates $\{x_1,x_2,x_3\}$ with respect to an orthogonal, Cartesian reference frame with the origin $\mathcal{O}$ and the three mutually perpendicular base vectors $\{\boldsymbol{i}_1,\boldsymbol{i}_2,\boldsymbol{i}_3\}$ of unit length each. In the indicated order, the base vectors form a right-handed system. The corresponding position vector is $\boldsymbol{x}=x_1\boldsymbol{i}_1+x_2\boldsymbol{i}_2+x_3\boldsymbol{i}_3$. The time coordinate is t. The subscript notation for vectors and tensors will be used and the summation convention for repeated subscripts applies. Lower-case Latin subscripts are employed for this purpose; they run through the values 1, 2, and 3. Partial differentiation with respect to x_m is denoted by ∂_m; ∂_t is a reserved symbol for differentiation with respect to t.

TABLE I. Elastodynamic wavefield, medium, and source quantities.

Symbol	Quantity
$\tau_{p,q}$	Dynamic stress
v_r	Particle velocity
$\rho_{k,r}$	Coefficient of inertia
$S_{i,j,p,q}$	Compliance
f_k	Volume source density of force
$h_{i,j}$	Volume source density of deformation rate

Both the background medium and the scatterer are assumed to consist of media that are linear, time-invariant, locally and instantaneously reacting in their elastodynamic behavior. The background medium has the entire $\mathbb{R}^3$ as its support. The support of the scatterer is the bounded domain $\mathcal{D}^{\mathrm{sc}}\subset\mathbb{R}^3$. In the scatterer the elastodynamic constitutive coefficients differ from those applied to the background medium. The physical quantities associated with elastodynamic wave motion in general are listed in Table I. The generic form of the coupled elastodynamic stress-velocity wave equations is taken as[2] (Sec. 10.7)

$$-\Delta^+_{k,m,p,q}\partial_m\tau_{p,q}+\rho_{k,r}\partial_t v_r=f_k, \tag{1}$$

$$\Delta^+_{i,j,n,r}\partial_n v_r-S_{i,j,p,q}\partial_t\tau_{p,q}=h_{i,j}. \tag{2}$$

Here, $\Delta^+_{i,j,p,q}$ denotes the symmetrical unit tensor of rank 4 defined as

$$\Delta^+_{i,j,p,q}=\tfrac{1}{2}(\delta_{i,p}\delta_{j,q}+\delta_{i,q}\delta_{j,p}), \tag{3}$$

with $\delta_{i,j}$ as the symmetrical unit tensor of rank 2 (Kronecker tensor): $\delta_{i,j}=1$ for $i=j=0$ for $i\neq j$.

To ensure the uniqueness of the solution to the initial-value problem associated with Eqs. (1) and (2) (a requirement set by the condition of causality of the physical phenomena involved), the constitutive coefficients have to satisfy the symmetry relations

$$\rho_{k,r}=\rho_{r,k}, \tag{4}$$

and

$$S_{i,j,p,q}=S_{i,j,q,p}=S_{j,i,q,p}=S_{j,i,p,q}=S_{p,q,i,j}, \tag{5}$$

while they are to be positive definite, i.e., $v_k\rho_{k,r}v_r>0$ for any $v_r\neq 0$ and $\tau_{i,j}S_{i,j,p,q}\tau_{p,q}>0$ for any $\tau_{p,q}\neq 0$.[5]

The wavefield quantities are expressed in terms of their generating source distributions through the relevant Green's tensors (point-source solutions) according to the scheme[2] (Sec. 15.8)

$$\begin{bmatrix}-\tau_{p,q}\\ v_r\end{bmatrix}(\boldsymbol{x},t)=\int_{\boldsymbol{x}'\in\mathrm{supp}(h,f)}\begin{bmatrix}G^{\tau,h}_{p,q,i,j} & G^{\tau,f}_{p,q,k}\\ G^{v,h}_{r,i,j} & G^{v,f}_{r,k}\end{bmatrix}(\boldsymbol{x},\boldsymbol{x}',t)\overset{(t)}{*}\begin{bmatrix}h_{i,j}\\ f_k\end{bmatrix}(\boldsymbol{x}',t)dV(\boldsymbol{x}') \quad \text{for } \boldsymbol{x}\in\mathbb{R}^3, \tag{6}$$

where $\overset{(t)}{*}$ denotes time convolution. The Green's tensors satisfy the symmetry relations

$$G^{\tau,h}_{p,q,i,j}(\boldsymbol{x},\boldsymbol{x}',t) = G^{\tau,h}_{p,q,j,i}(\boldsymbol{x},\boldsymbol{x}',t),$$

$$G^{\tau,h}_{q,p,j,i}(\boldsymbol{x},\boldsymbol{x}',t) = G^{\tau,h}_{q,p,i,j}(\boldsymbol{x},\boldsymbol{x}',t), \tag{7}$$

$$G^{\tau,f}_{p,q,k}(\boldsymbol{x},\boldsymbol{x}',t) = G^{\tau,f}_{q,p,k}(\boldsymbol{x},\boldsymbol{x}',t), \tag{8}$$

$$G^{v,h}_{r,i,j}(\boldsymbol{x},\boldsymbol{x}',t) = G^{v,h}_{r,j,i}(\boldsymbol{x},\boldsymbol{x}',t), \tag{9}$$

while with the aid of the elastodynamic time-domain reciprocity relation of the time-convolution type they can be shown to have the reciprocity properties[2] (pp. 471–472)

$$G^{\tau,h}_{p,q,i,j}(\boldsymbol{x},\boldsymbol{x}',t) = G^{\tau,h}_{i,j,p,q}(\boldsymbol{x}',\boldsymbol{x},t), \tag{10}$$

$$G^{\tau,f}_{p,q,k}(\boldsymbol{x},\boldsymbol{x}',t) = -G^{v,h}_{k,p,q}(\boldsymbol{x}',\boldsymbol{x},t), \tag{11}$$

$$G^{v,f}_{r,k}(\boldsymbol{x},\boldsymbol{x}',t) = G^{v,f}_{k,r}(\boldsymbol{x}',\boldsymbol{x},t). \tag{12}$$

In a homogeneous medium the arguments $\boldsymbol{x}$ and $\boldsymbol{x}'$ in the Green's tensors only occur via their difference $\boldsymbol{x}-\boldsymbol{x}'$, which makes the integral in Eq. (6) of the spatial convolution type, which property can be computationally useful.

In the evaluation of the Green's tensors, also the inverse of the compliance, the *stiffness* $C_{p,q,i,j}$, is needed. It is defined through

$$S_{i,j,m,n}C_{m,n,p,q} = \Delta^{+}_{i,j,p,q} \tag{13}$$

and shares the same symmetry and reciprocity properties as the compliance.

III. CONTRAST-SOURCE FORMULATION OF THE SCATTERING PROBLEM

The scatterer is irradiated by the action of controlled sources with volume density of force $f^{\mathrm{inc}}_k(\boldsymbol{x},t)$ and volume source density of deformation rate $h^{\mathrm{inc}}_{i,j}(\boldsymbol{x},t)$. They are present in the *background medium* with coefficient of inertia $\rho^b_{k,r}(\boldsymbol{x})$ and compliance $S^b_{i,j,p,q}(\boldsymbol{x})$ and generate the *incident wavefield* $\{\tau^{\mathrm{inc}}_{p,q}(\boldsymbol{x},t), v^{\mathrm{inc}}_r(\boldsymbol{x},t)\}$. The domain

$$\mathcal{D}^{\mathrm{inc}} = \mathrm{supp}(f^{\mathrm{inc}}_k) \cup \mathrm{supp}(h^{\mathrm{inc}}_{i,j}) \tag{14}$$

is the union of the supports of the generating source distributions.

Starting point for the contrast-source formulation of the scattering problem is writing the constitutive properties of the scatterer as a deviation from the ones of the background medium in which it is embedded. Let

$$\rho_{k,r}(\boldsymbol{x}) = \rho^b_{k,r}(\boldsymbol{x}) + \delta\rho_{k,r}(\boldsymbol{x}) \quad \text{for } \boldsymbol{x} \in \mathcal{D}^{\mathrm{sc}}, \tag{15}$$

$$S_{i,j,p,q}(\boldsymbol{x}) = S^b_{i,j,p,q}(\boldsymbol{x}) + \delta S_{i,j,p,q}(\boldsymbol{x}) \quad \text{for } \boldsymbol{x} \in \mathcal{D}^{\mathrm{sc}}, \tag{16}$$

where

$$\mathcal{D}^{\mathrm{sc}} = \mathrm{supp}(\delta\rho) \cup \mathrm{supp}(\delta S) \tag{17}$$

is the union of the supports of the contrasts.

Upon introducing the *scattered wavefield* $\{\tau^{\mathrm{sc}}_{p,q}, v^{\mathrm{sc}}_r\}$ as

$$\tau^{\mathrm{sc}}_{p,q} = \tau_{p,q} - \tau^{\mathrm{inc}}_{p,q}, \tag{18}$$

$$v^{\mathrm{sc}}_r = v_r - v^{\mathrm{inc}}_r \tag{19}$$

accounting for the presence of the scatterer via the introduction of the *contrast volume source density of deformation rate*

$$h^{\mathrm{sc}}_{i,j} = \delta S_{i,j,p,q}(\boldsymbol{x})\partial_t\tau_{p,q}, \tag{20}$$

and the *contrast volume source density of force*

$$f^{\mathrm{sc}}_k = -\delta\rho_{k,r}(\boldsymbol{x})\partial_t v_r, \tag{21}$$

the coupled wave equations for the incident and scattered wavefields can be combined to [cf. Eqs. (1) and (2)]

$$-\Delta^{+}_{k,m,p,q}\partial_m\tau^{\mathrm{inc;sc}}_{p,q} + \rho^b_{k,r}\partial_t v^{\mathrm{inc;sc}}_r = f^{\mathrm{inc;sc}}_k, \tag{22}$$

$$\Delta^{+}_{i,j,n,r}\partial_n v^{\mathrm{inc;sc}}_r - S^b_{i,j,p,q}\partial_t\tau^{\mathrm{inc;sc}}_{p,q} = h^{\mathrm{inc;sc}}_{i,j}, \tag{23}$$

and the incident and scattered wavefield integral representations to [cf. Eq. (6)]

$$\begin{bmatrix} -\tau^{\mathrm{inc;sc}}_{p,q} \\ v^{\mathrm{inc;sc}}_r \end{bmatrix}(\boldsymbol{x},t) = \int_{\boldsymbol{x}\in\mathcal{D}^{\mathrm{inc;sc}}} \begin{bmatrix} G^{\tau,h}_{p,q,i,j} & G^{r,f}_{p,q,k} \\ G^{v,h}_{r,i,j} & G^{v,f}_{r,k} \end{bmatrix}(\boldsymbol{x},\boldsymbol{x}',t) \overset{(t)}{*} \begin{bmatrix} h^{\mathrm{inc;sc}}_{i,j} \\ f^{\mathrm{inc;sc}}_k \end{bmatrix}(\boldsymbol{x}',t)\,dV(\boldsymbol{x}') \quad \text{for } \boldsymbol{x} \in \mathbb{R}^3, \tag{24}$$

in which the Green's tensors are the ones associated with the background medium.

Once the incident wavefield has been determined, Eq. (24) as it applies to the scattered wavefield can be used to arrive at integral-equation formulations for the solution of the scattering problem.

A. Wavefield integral-equation formulation

The wavefield integral-equation formulation follows upon substituting Eqs. (18) and (19) in the left-hand side and Eqs. (20) and (21) in the right-hand side and enforcing the result for $\boldsymbol{x} \in \mathcal{D}^{\mathrm{sc}}$, while considering the wavefield quantities $\tau_{p,q}(\boldsymbol{x})$ and $v_r(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{D}^{\mathrm{sc}}$ as the unknowns. A formulation of this kind has been employed in Ref. 6 to solve elastic wave problems in inhomogeneous media.

B. Contrast-source integral-equation formulation

The contrast-source integral-equation formulation follows upon substituting Eqs. (18) and (19) in the left-hand side, operating on the resulting upper equation with $\delta S_{i,j,p,q}(\boldsymbol{x})\partial_t$ and on the lower equation with $\delta\rho_{k,r}(\boldsymbol{x})\partial_t$ and using the resulting relations throughout $\mathcal{D}^{\mathrm{sc}}$, while considering $h^{\mathrm{sc}}_{i,j}(\boldsymbol{x})$ and $f^{\mathrm{sc}}_k(\boldsymbol{x})$ as the unknowns.

IV. THE DECOMPOSITION OF THE ELASTODYNAMIC TENSORS OF RANK 2 INTO THEIR OMNIDIRECTIONAL AND DEVIATORIC PARTS

The symmetrical tensors of rank 2 occurring in elastodynamics can, in a unique manner, be decomposed into their *omnidirectional parts* and their *deviatoric parts* in the function space they span. This decomposition highlights particular features of the generated wave motion and can be computationally advantageous. The decomposition is most

effectively carried out through the introduction of a collection of unit tensors of rank 4. In view of two-dimensional, next to 3D, modeling applications, the definitions are given for an arbitrary number N ($N \geqslant 2$) of spatial dimensions.[7]

The *identity tensor*:

$$\Delta_{i,j,p,q} = \delta_{i,p}\delta_{j,q}, \tag{25}$$

the *symmetrical unit tensor*:

$$\Delta^{+}_{i,j,p,q} = \tfrac{1}{2}(\delta_{i,p}\delta_{j,q} + \delta_{i,q}\delta_{j,p}), \tag{26}$$

the *omnidirectional part of the symmetrical unit tensor*:

$$\Delta^{\delta}_{i,j,p,q} = \frac{1}{N}\delta_{i,j}\delta_{p,q}, \tag{27}$$

the *deviatoric part of the symmetrical unit tensor*:

$$\Delta^{+\backslash\delta}_{i,j,p,q} = \Delta^{+}_{i,j,p,q} - \Delta^{\delta}_{i,j,p,q} = \frac{1}{2}(\delta_{i,p}\delta_{j,q} + \delta_{i,q}\delta_{j,p}) - \frac{1}{N}\delta_{i,j}\delta_{p,q}. \tag{28}$$

The unit tensors thus introduced all have the *unitary properties*

$$\Delta_{i,j,m,n}\Delta_{m,n,p,q} = \Delta_{i,j,p,q}, \tag{29}$$

$$\Delta^{+}_{i,j,m,n}\Delta^{+}_{m,n,p,q} = \Delta^{+}_{i,j,p,q}, \tag{30}$$

$$\Delta^{\delta}_{i,j,m,n}\Delta^{\delta}_{m,n,p,q} = \Delta^{\delta}_{i,j,p,q}, \tag{31}$$

$$\Delta^{+\backslash\delta}_{i,j,m,n}\Delta^{+\backslash\delta}_{m,n,p,q} = \Delta^{+\backslash\delta}_{i,j,p,q}. \tag{32}$$

Furthermore, we have the *orthogonality property*

$$\Delta^{\delta}_{i,j,m,n}\Delta^{+\backslash\delta}_{m,n,p,q} = 0. \tag{33}$$

For any tensor $T_{p,q}$ of rank 2 we then have

$$\Delta_{i,j,p,q}T_{p,q} = T_{i,j}, \tag{34}$$

while we define its *symmetrical part* by

$$T^{+}_{i,j} = \Delta^{+}_{i,j,p,q}T_{p,q}, \tag{35}$$

its *omnidirectional part* by

$$T^{\delta}_{i,j} = \Delta^{\delta}_{i,j,p,q}T_{p,q}, \tag{36}$$

and its *deviatoric part* by

$$T^{+\backslash\delta}_{i,j} = \Delta^{+\backslash\delta}_{i,j,p,q}T_{p,q}. \tag{37}$$

Evidently,

$$T^{+}_{i,j} = T^{\delta}_{i,j} + T^{+\backslash\delta}_{i,j}, \tag{38}$$

and

$$T^{\delta}_{i,j}\Delta_{i,j,p,q}T^{+\backslash\delta}_{p,q} = T^{\delta}_{i,j}T^{+\backslash\delta}_{i,j} = 0. \tag{39}$$

Equations (38) and (39) imply that $T^{\delta}_{i,j}$ and $T^{+\backslash\delta}_{i,j}$ are mutually orthogonal constituents of $T^{+}_{i,j}$ in the function space spanned by $T^{+}_{i,j}$.

With the decomposition, Eqs. (1) and (2) can be rewritten as (note that now $N=3$)

$$-\partial_m(\tau^{\delta}_{k,m} + \tau^{+\backslash\delta}_{k,m}) + \rho_{k,r}\partial_t v_r = f_k, \tag{40}$$

$$(\Delta^{\delta}_{i,j,n,r} + \Delta^{+\backslash\delta}_{i,j,n,r})\partial_n v_r - S_{i,j,p,q}\partial_t(\tau^{\delta}_{p,q} + \tau^{+\backslash\delta}_{p,q}) = h^{\delta}_{i,j} + h^{+\backslash\delta}_{i,j}. \tag{41}$$

In general, Eq. (41) does not decompose into separate equations for the separate constituents. However, such a decomposition does take place for the class of media for which (no coupling between the mutually orthogonal constituents)

$$\Delta^{\delta}_{i,j,m,n}S_{m,n,r,s}\Delta^{+\backslash\delta}_{r,s,p,q} = \Delta^{+\backslash\delta}_{i,j,m,n}S_{m,n,r,s}\Delta^{\delta}_{r,s,p,q} = 0. \tag{42}$$

With

$$S^{\delta,\delta}_{i,j,p,q} = \Delta^{\delta}_{i,j,m,n}S_{m,n,r,s}\Delta^{\delta}_{r,s,p,q} \tag{43}$$

and

$$S^{+\backslash\delta,+\backslash\delta}_{i,j,p,q} = \Delta^{+\backslash\delta}_{i,j,m,n}S_{m,n,r,s}\Delta^{+\backslash\delta}_{r,s,p,q}, \tag{44}$$

we have in such a case

$$S_{i,j,p,q} = S^{\delta,\delta}_{i,j,p,q} + S^{+\backslash\delta,+\backslash\delta}_{i,j,p,q}, \tag{45}$$

in which the two constituents on the right-hand side are mutually orthogonal. As a consequence, Eq. (41) decomposes into

$$\Delta^{\delta}_{i,j,n,r}\partial_n v_r - S^{\delta,\delta}_{i,j,p,q}\partial_t\tau^{\delta}_{p,q} = h^{\delta}_{i,j} \tag{46}$$

and

$$\Delta^{+\backslash\delta}_{i,j,n,r}\partial_n v_r - S^{+\backslash\delta,+\backslash\delta}_{i,j,p,q}\partial_t\tau^{+\backslash\delta}_{p,q} = h^{+\backslash\delta}_{i,j}. \tag{47}$$

The decomposition has also consequences for the field integral representation (6). Through the decomposition, Eq. (6) can be reformulated as

$$\begin{bmatrix} -\tau^{\delta}_{p,q} \\ -\tau^{+\backslash\delta}_{p,q} \\ v_r \end{bmatrix}(\boldsymbol{x},t) = \int_{\boldsymbol{x}\in\mathrm{supp}(h,f)} \begin{bmatrix} G^{\delta,\delta}_{p,q,i,j} & G^{\delta,+\backslash\delta}_{p,q,i,j} & G^{\delta,f}_{p,q,k} \\ G^{+\backslash\delta,\delta}_{p,q,i,j} & G^{+\backslash\delta,+\backslash\delta}_{p,q,i,j} & G^{+\backslash\delta,f}_{p,q,k} \\ G^{v,\delta}_{r,i,j} & G^{v,+\backslash\delta}_{r,i,j} & G^{v,f}_{r,k} \end{bmatrix} (\boldsymbol{x},\boldsymbol{x}',t) \overset{(t)}{*} \begin{bmatrix} h^{\delta}_{i,j} \\ h^{+\backslash\delta}_{i,j} \\ f_k \end{bmatrix}(\boldsymbol{x}',t)dV(\boldsymbol{x}') \quad \text{for } \boldsymbol{x} \in \mathbb{R}^3, \tag{48}$$

in which

$$G^{\delta,\delta}_{p,q,i,j} = \Delta^{\delta}_{p,q,r,s}G^{\tau,h}_{r,s,m,n}\Delta^{\delta}_{m,n,i,j}, \tag{49}$$

$$G^{\delta,+\backslash\delta}_{p,q,i,j} = \Delta^{\delta}_{p,q,r,s}G^{\tau,h}_{r,s,m,n}\Delta^{+\backslash\delta}_{m,n,i,j}, \tag{50}$$

$$G^{\delta,f}_{p,q,k} = \Delta^{\delta}_{p,q,r,s}G^{\tau,f}_{r,s,k}, \tag{51}$$

$$G^{+\backslash\delta,\delta}_{p,q,i,j} = \Delta^{+\backslash\delta}_{p,q,r,s}G^{\tau,h}_{r,s,m,n}\Delta^{\delta}_{m,n,i,j}, \tag{52}$$

$$G^{+\backslash\delta,+\backslash\delta}_{p,q,i,j} = \Delta^{+\backslash\delta}_{p,q,r,s}G^{\tau,h}_{r,s,m,n}\Delta^{+\backslash\delta}_{m,n,i,j}, \tag{53}$$

$$G^{+\backslash\delta,f}_{p,q,k} = \Delta^{+\backslash\delta}_{p,q,r,s}G^{\tau,f}_{r,s,k}, \tag{54}$$

$$G^{v,\delta}_{r,i,j} = G^{v,h}_{r,m,n}\Delta^{\delta}_{m,n,i,j}, \tag{55}$$

$$G^{v,+\backslash\delta}_{r,i,j} = G^{v,h}_{r,m,n}\Delta^{+\backslash\delta}_{m,n,i,j}. \tag{56}$$

The decomposition of $S_{i,j,p,q}$ holds in particular for the case of isotropic media. Here, it leads to the introduction of con-

stitutive coefficients that are related to the traditional Lamé coefficients but are particular combinations of them that show up in the contrast-source integral-equation formulation of the scattering problem.

V. BACKGROUND MEDIUM AND SCATTERER WITH ISOTROPIC CONSTITUTIVE PROPERTIES

For isotropic, lossless media, the constitutive coefficients take the form

$$\rho_{k,r}(\boldsymbol{x}) = \rho(\boldsymbol{x})\,\delta_{k,r}, \tag{57}$$

where $\rho(\boldsymbol{x})$ is the volume density of mass [$\rho(\boldsymbol{x})>0$ in view of the uniqueness conditions],

$$S_{i,j,p,q}(\boldsymbol{x}) = S^{\delta,\delta}_{i,j,p,q}(\boldsymbol{x}) + S^{+\backslash\delta,+\backslash\delta}_{i,j,p,q}(\boldsymbol{x}), \tag{58}$$

with

$$S^{\delta,\delta}_{i,j,p,q}(\boldsymbol{x}) = \Lambda(\boldsymbol{x})\Delta^{\delta}_{i,j,p,q} \tag{59}$$

and

$$S^{+\backslash\delta,+\backslash\delta}_{i,j,p,q}(\boldsymbol{x}) = M(\boldsymbol{x})\Delta^{+\backslash\delta}_{i,j,p,q}, \tag{60}$$

where $\Lambda(\boldsymbol{x})$ and $M(\boldsymbol{x})$ [with $\Lambda(\boldsymbol{x})>0$ and $M(\boldsymbol{x})>0$ in view of the uniqueness conditions] are the compliance coefficients and related to the Lamé coefficients λ and μ via

$$\frac{1}{\Lambda} = 3\lambda + 2\mu, \tag{61}$$

$$\frac{1}{M} = 2\mu. \tag{62}$$

The corresponding stiffness is given by

$$C_{p,q,i,j}(\boldsymbol{x}) = C^{\delta,\delta}_{p,q,i,j}(\boldsymbol{x}) + C^{+\backslash\delta,+\backslash\delta}_{p,q,i,j}(\boldsymbol{x}), \tag{63}$$

with

$$C^{\delta,\delta}_{p,q,i,j}(\boldsymbol{x}) = \frac{1}{\Lambda(\boldsymbol{x})}\Delta^{\delta}_{p,q,i,j} \tag{64}$$

and

$$C^{+\backslash\delta,+\backslash\delta}_{p,q,i,j}(\boldsymbol{x}) = \frac{1}{M(\boldsymbol{x})}\Delta^{+\backslash\delta}_{p,q,i,j}. \tag{65}$$

For the scattering problem, the constitutive properties of the background medium are specified by

$$\rho^{b}_{k,r}(\boldsymbol{x}) = \rho^{b}(\boldsymbol{x})\,\delta_{k,r}, \tag{66}$$

$$S^{b}_{i,j,p,q}(\boldsymbol{x}) = \Lambda^{b}(\boldsymbol{x})\Delta^{\delta}_{i,j,p,q} + M^{b}(\boldsymbol{x})\Delta^{+\backslash\delta}_{i,j,p,q}, \tag{67}$$

and the contrast constitutive properties by

$$\delta\rho_{k,r}(\boldsymbol{x}) = \delta\rho(\boldsymbol{x})\,\delta_{k,r}, \tag{68}$$

$$\delta S_{i,j,p,q}(\boldsymbol{x}) = \delta\Lambda(\boldsymbol{x})\Delta^{\delta}_{i,j,p,q} + \delta M(\boldsymbol{x})\Delta^{+\backslash\delta}_{i,j,p,q}. \tag{69}$$

With this, the contrast source densities become

$$h^{\mathrm{sc}}_{i,j} = h^{\mathrm{sc};\delta}_{i,j} + h^{\mathrm{sc};+\backslash\delta}_{i,j}, \tag{70}$$

with

$$h^{\mathrm{sc};\delta}_{i,j} = \delta\Lambda(\boldsymbol{x})\partial_t\tau^{\delta}_{i,j}, \tag{71}$$

$$h^{\mathrm{sc};+\backslash\delta}_{i,j} = \delta M(\boldsymbol{x})\partial_t\tau^{+\backslash\delta}_{i,j} \tag{72}$$

and

$$f^{\mathrm{sc}}_{k} = -\,\delta\rho(\boldsymbol{x})\partial_t v_k. \tag{73}$$

VI. THE GREEN'S TENSORS FOR A HOMOGENEOUS, ISOTROPIC, LOSSLESS BACKGROUND MEDIUM

Although the decomposition discussed in Sec. IV leads to the introduction of elastic constitutive coefficients that are directly related to the contrast source densities of deformation rate that occur in the elastodynamic scattering problem and the Green's tensors do decompose accordingly, the wave speeds occurring in them prove still to be related to combinations of the newly introduced coefficients and not to them separately. To show this, we give in this section the expressions for the Green's tensors pertaining to a homogeneous, isotropic, lossless background medium with the constitutive coefficients $\Lambda=\Lambda_0$, $M=M_0$, and $\rho=\rho_0$. In the expressions, the wavespeed c_P of compressional waves and the wavespeed c_S of shear waves occur as well as the *P*-wave and *S*-wave constituents of the Green's tensor $G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t)$ of the *elastodynamic wave equation*. The latter tensor follows from

$$(c_P^2 - c_S^2)\partial_r\partial_s G_{s,k}(\boldsymbol{x}-\boldsymbol{x}',t) + c_S^2\partial_s\partial_s G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) - \partial_t^2 G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) = -\,\delta_{r,k}\delta(\boldsymbol{x}-\boldsymbol{x}',t) \tag{74}$$

and is obtained as[2] (Sec. 13.5)

$$G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) = \delta_{r,k}\frac{1}{c_S^2}\frac{\delta(t-|\boldsymbol{x}-\boldsymbol{x}'|/c_S)}{4\pi|\boldsymbol{x}-\boldsymbol{x}'|} + \partial_r\partial_k\left[\frac{(t-|\boldsymbol{x}-\boldsymbol{x}'|/c_P)H(t-|\boldsymbol{x}-\boldsymbol{x}'|/c_P)}{4\pi|\boldsymbol{x}-\boldsymbol{x}'|} - \frac{(t-|\boldsymbol{x}-\boldsymbol{x}'|/c_S)H(t-|\boldsymbol{x}-\boldsymbol{x}'|/c_S)}{4\pi|\boldsymbol{x}-\boldsymbol{x}'|}\right] \quad \text{for } \boldsymbol{x}\neq\boldsymbol{x}', \tag{75}$$

where $H(t)$ denotes the Heaviside unit step function. Note that, in view of the homogeneity of the medium, the Green's tensors depend on $\boldsymbol{x}$ and $\boldsymbol{x}'$ only via their difference $\boldsymbol{x}-\boldsymbol{x}'$. Expressed in terms of Λ_0, M_0, and ρ_0 we further have

$$c_P = \left[\left(\frac{2}{3M_0} + \frac{1}{3\Lambda_0}\right)\frac{1}{\rho_0}\right]^{1/2} \tag{76}$$

and

$$c_S = \left(\frac{1}{2M_0\rho_0}\right)^{1/2} \tag{77}$$

The Green's tensors occurring in Eq. (48) can now be expressed in terms of $G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t)$ and its derivatives. From the particle-velocity elastodynamic wave equation with point force excitation (18) we directly obtain

$$G^{v,f}_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) = \frac{1}{\rho_0}\partial_t G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t). \tag{78}$$

Using Eq. (19), it then follows that

$$G^{\delta,f}_{p,q,k}(\boldsymbol{x}-\boldsymbol{x}',t)=\frac{1}{\rho_0\Lambda_0}\Delta^{\delta}_{p,q,n,r}\partial_n G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) \tag{79}$$

$$G^{+\backslash\delta,f}_{p,q,k}(\boldsymbol{x}-\boldsymbol{x}',t)=\frac{1}{\rho_0 M_0}\Delta^{+\backslash\delta}_{p,q,n,r}\partial_n G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t). \tag{80}$$

Using Eq. (20), we obtain, noting that $\partial'_s=-\partial_s$ for the case of homogeneous media,

$$G^{v,\delta}_{r,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=\frac{1}{\rho_0\Lambda_0}\Delta^{\delta}_{i,j,n,r}\partial_n G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t) \tag{81}$$

and

$$G^{v,+\backslash\delta}_{r,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=\frac{1}{\rho_0 M_0}\Delta^{+\backslash\delta}_{i,j,n,r}\partial_n G_{r,k}(\boldsymbol{x}-\boldsymbol{x}',t). \tag{82}$$

Finally, Eq. (21) yields

$$G^{\delta,\delta}_{p,q,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=-\frac{1}{\Lambda_0}\Delta^{\delta}_{p,q,i,j}\delta(\boldsymbol{x}-\boldsymbol{x}')H(t)-\frac{1}{\rho_0\Lambda_0^2}\partial_t^{-1}\Delta^{\delta}_{p,q,n,r}\Delta^{\delta}_{i,j,i'j'}\partial_n\partial_{i'}G_{r,j'}(\boldsymbol{x}-\boldsymbol{x}',t), \tag{83}$$

$$G^{\delta\backslash\delta}_{p,q,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=-\frac{1}{\rho_0\Lambda_0 M_0}\partial_t^{-1}\Delta^{\delta}_{p,q,n,r}\Delta^{+\backslash\delta}_{i,j,i'j'}\partial_n\partial_{i'}G_{r,j'}\times(\boldsymbol{x}-\boldsymbol{x}',t), \tag{84}$$

$$G^{\backslash\delta,\delta}_{p,q,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=-\frac{1}{\rho_0\Lambda_0 M_0}\partial_t^{-1}\Delta^{+\backslash\delta}_{p,q,n,r}\Delta^{\delta}_{i,j,i'j'}\partial_n\partial_{i'}G_{r,j'}\times(\boldsymbol{x}-\boldsymbol{x}',t), \tag{85}$$

$$G^{\backslash\delta,\backslash\delta}_{p,q,i,j}(\boldsymbol{x}-\boldsymbol{x}',t)=-\frac{1}{M_0}\Delta^{+\backslash\delta}_{p,q,i,j}\delta(\boldsymbol{x}-\boldsymbol{x}')H(t)-\frac{1}{\rho_0 M_0^2}\partial_t^{-1}\Delta^{+\backslash\delta}_{p,q,n,r}\Delta^{+\backslash\delta}_{i,j,i'j'}\partial_n\partial_{i'}G_{r,j'}(\boldsymbol{x}-\boldsymbol{x}',t) \tag{86}$$

In the right-hand sides of the expressions for the Green's tensors in terms of the tensor of rank 2 of the elastodynamic wave equation spatial derivatives up to order 4 acting on the elementary scalar Green's functions of the wave equation occur. A list of these derivatives can be found in Ref. 2 (Sec. 13.3).

VII. CONCLUSION

It is shown that for a large class of elastodynamic scattering problems the decomposition of the pertaining tensors of rank 2 (dynamic stress and volume source density of deformation rate) into their omnidirectional and deviatoric constituents has a number of advantages. First of all, the contrast in elastic properties of the scatterers, in particular, those in the case of scatterers composed of isotropic media present in a background with isotropic properties, admits a more natural representation than the traditional one in terms of the relevant Lamé coefficients. Since the newly introduced contrast coefficients directly occur in the contrast-source densities that are representative for the scattering phenomena, any sensitivity analysis in the associated inverse scattering problems becomes more transparent. The introduction of the relevant projection tensors of rank 4 also highlights the structure of the relevant (also decomposed) Green's tensors, which structure can serve as a guideline to the computational implementation of the contrast-source integral-equation method.

ACKNOWLEDGMENTS

The authors would like to thank the (anonymous) reviewers for their efforts in evaluating the manuscript and for making suggestions to more clearly elucidate the underlying physics as well as its relevance to applications.

[1] A. Abubakar and P. M. Van den Berg, "Iterative forward and inverse algorithms based on domain integral equations for three-dimensional electric and magnetic objects," J. Comput. Phys. **195**, 236–262 (2004).
[2] A. T. De Hoop, *Handbook of Radiation and Scattering of Waves* (Academic, London, 1995).
[3] L. E. Malvern, *Introduction to the Mechanics of a Continuous Medium* (Prentic-Hall, Englewood Cliffs, NJ, 1969).
[4] J. D. Achenbach, A. K. Gautesen, and H. McMaken, *Ray Methods for Waves in Elastic Solids* (Pitman, Boston, 1982).
[5] A. T. De Hoop, "A uniqueness theorem for the time-domain elastic-wave scattering in inhomogeneous, anisotropic solids with relaxation," J. Am. Chem. Soc. **115**, 2711–2715 (2004).
[6] J. Yang, A. Abubakar, P. M. Van den Berg, T. M. Habashy, and F. Reitich, "A CG-FFT approach to the solution of a stress-velocity formulation of three-dimensional elastic scattering problems," J. Comput. Phys. **227**, 10018–10039 (2008).
[7] A. T. De Hoop, "Orthogonal function-space decomposition of tensors of rank two in *N*-dimensional space," Laboratory of Electromagnetic Research, Delft University of Technology Report No. EEMCS/EM 2008-09, Delft University of Technology, Delft, The Netherlands, 2008.

Rapid thickness measurements using guided waves from a scanning laser source

Takahiro Hayashi,[a] Morimasa Murase, and Muhammad Nor Salim
Faculty of Engineering, Nagoya Institute of Technology, Gokiso, Showa, Nagoya 466-8555, Japan

(Received 6 January 2009; revised 13 June 2009; accepted 16 June 2009)

Guided waves have been effectively used for rapid inspections of plates and pipes. However, the guided-wave technique is not generally used for measuring the remaining thickness in a plate and a pipe due to the difficulties in guided-wave motion. Instead, time-consuming and costly direct contact thickness measurements are still used in practice. This study describes a thickness measurement technique using the A0 mode of a Lamb wave generated by a laser source. A finite element analysis of Lamb wave revealed that this mode propagates with small reflections and mode conversions at a rounded shallow defect and has larger amplitude at thinner regions. Using these characteristics, it is experimentally demonstrated that the distributions of plate thickness were obtained from the amplitude of A0 mode generated by a scanning laser source and received by an angle-beam transducer. The resulting distribution images were obtained at extremely high speed compared to the conventional thickness measurements.
 [DOI: 10.1121/1.3177268]

PACS number(s): 43.35.Zc, 43.35.Yb, 43.20.Mv [AJZ] Pages: 1101–1106

I. INTRODUCTION

Low-frequency ultrasonic propagation of guided waves along elongated structures such as plates and pipes has been widely used for nondestructive inspection of inaccessible regions in large structures.[1,2] Usually, in guided-wave inspection, defects are located by pulse-echo method using the arrival time of defect echoes and the group velocity of the guided waves. In pipe inspection, long-range screening has been realized by the use of the nondispersive axisymmetric torsional mode $T(0,1)$.[3,4] Many studies have also examined non-axisymmetric modes for improving pipe inspection.[5,6] Hayashi and Murase[7] and Davis and Cawley[8] further developed defect-imaging techniques by the use of guided waves in a pipe.

However, such fast and cost-effective techniques have not been developed for evaluating remaining thickness of a plate and a pipe, which is the most critical factor in the maintenance of large structures such as tanks and pipeworks. Thus, time-consuming and costly conventional techniques using a contact ultrasonic transducer are still used for thickness measurements.

The present study introduces a rapid thickness measurement technique for plates using guided waves. Many studies in the past discussed the amplitude change in pulse echoes for different defect depths.[2,9,10] Generally, the amplitude of an echo signal tends to increase as the ratio of projected area of the defect to its cross-sectional area increases. However, the amplitude is also affected by the width, length, and shape of a defect,[11] and so in many cases, the amplitude alone cannot be used as an index of the remaining thickness. Such imaging techniques[7,8] can be used to count defects, but not to evaluate the remaining thickness.

In this study, the authors use the A0 mode of a Lamb wave generated from a laser source[12,13] measured by an ultrasonic transducer located remotely from the source, rather than a pulse-echo technique. A laser beam can be emitted far from the inspected area, and the source of the laser-generated ultrasonic wave can be rastered with a galvano scanner. The Lamb wave is detected with angle-beam transducers located far away from the source region. Rapid thickness measurements can thus be carried out.

The paper is organized as follows. First, numerical analyses are carried out to determine the scope of the thickness measurement using a combined model of a semi-analytical finite element (SAFE) method[6,14–17] and a finite element (FE) method. Next, the wave propagation of an A0 mode at a defect in a plate is discussed in detail. Then, images of remaining thickness of a plate are obtained experimentally.

II. CALCULATION TECHNIQUE FOR GUIDED-WAVE PROPAGATION

In order to understand guided-wave propagation around defects, a combined SAFE[6,14–17] and FE model was used in this study. This section briefly describes the combination technique for calculating guide wave propagation and then evaluates the accuracy of the calculation code.

Recently, SAFE has been widely used in guided-wave calculations.[6,14–17] In that method, only the cross-section of the object is divided into small elements, while the longitudinal direction is expressed in terms of orthogonal functions. Consequently, a smaller number of nodes are required than in the ordinary FE method, so that less memory and shorter calculation times are required. In addition, modal analysis is easier because the wave propagation is described as a sum of guided-wave modes. However, the geometry and material properties must be constant in the longitudinal direction, and

[a]Author to whom correspondence should be addressed. Electronic mail: hayashi@nitech.ac.jp

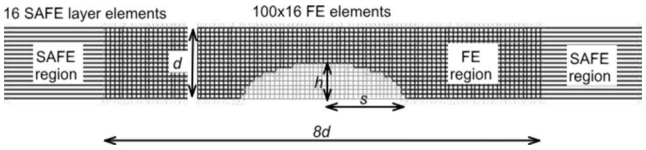


FIG. 1. SAFE and FE meshes used in the combination technique for modeling a plate having a defect.

thus a plate or a pipe with a defect of complex shape cannot be modeled only by the SAFE model. For this reason, a combination technique[18–21] is here adopted. Various combination techniques have been developed for an effective calculation of Lamb-wave propagation. For example, Al-Nassar *et al.*[18] developed the combination technique of an analytical solution of Lamb wave and FE method, Cho and Rose[19] introduced the combination technique of an analytical solution and boundary element (BE) method, and Galan and Abascal[20] used the SAFE and BE models. In this study, the authors adopt the similar combination technique, in which the SAFE is used for intact regions and the ordinary FE method is only used in defect areas.

In order to solve for Lamb-wave propagation in a plate having a defect, such a combination method for a two-dimensional elastodynamic problem is used here. As shown in Fig. 1, the defect is described by a FE region divided in both the vertical and horizontal directions into small rectangular elements. At both ends of the FE region, SAFE regions divided into layered elements are connected. In the FE region, a 16×100 rectangular mesh is used, where the elastic constants of the elements in a dented region are specified to be small. Using a uniform mesh of identical rectangular shapes, the calculations in the FE region require less calculation time and memory. The SAFE and FE regions are connected with continuous displacements and stresses at their boundaries. Displacements and stresses within the plate can be calculated at every frequency step. The frequency domain data are converted to transient wave forms in time domain using inverse fast Fourier transform (FFT).

First of all, in order to evaluate the accuracy of the SAFE and FE combination model, the reflection and transmission factors were calculated and compared with the values obtained by Cho and Rose[19] and Galan and Abascal[20] for a steel plate having an elliptical defect of horizontal radius s and vertical radius h for A0 incident mode. These factors can be calculated from the in-plane average power. The displacements and stresses in the SAFE regions determine the in-plane average power for the incident, reflected, and transmitted waves. Taking the ratio of the reflected or transmitted waves to the incident wave, the square root of the ratios becomes the reflection and transmission factors, R and T.[19,20]

Figure 2 plots R and T of the A0 and S0 modes for A0 incidence in a steel plate of thickness d, longitudinal velocity $c_L=5940$ m/s, and transverse velocity $c_T=3200$ m/s with an elliptically rounded dent of $s=1.5d$ and $h=0.5d$. The horizontal axis is the product of frequency f and thickness d. Cho and Rose[19] and Galan and Abascal[20] found the same results for a value of fd below the A1 cutoff frequency of $fd=1.63$ MHz mm. This agreement demonstrates that the combination model has sufficient accuracy.

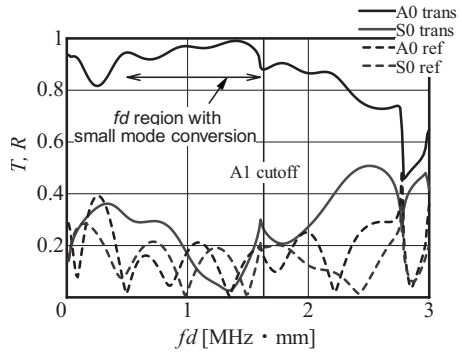


FIG. 2. Transmission and reflection factors for A0 incidence in a steel plate with $c_L=5940$ m/s and $c_T=3200$ m/s.

In Fig. 2, below the cutoff frequency, reflections and mode conversions are small, and the incident A0 mode propagates through the defect while keeping its mode shape. This phenomenon is an important characteristic for the thickness measurements described next.

III. PROPAGATION OF AN A0 MODE AT A DENT AND THICKNESS MEASUREMENT

Displacements and stresses are obtained at every frequency step, and then the reflection and transmission factors R and T are calculated. After collecting displacements at all grid points on the cross-section of a plate, time-domain wave forms are obtained by taking an inverse FFT. Then the wave forms at all grid points can be visualized.[17,21] In this section, Lamb-wave propagation for A0 incidence is calculated and visualized for explanation of the principles of the thickness measurement technique described in this study.

Wave propagation around an elliptical defect ($s=3d$ and $h=0.75d$) is calculated for A0 incidence. For the comparison with experimental results shown later, calculations after this section are carried out for an aluminum plate of thickness d ($c_L=6260$ m/s and $c_T=3080$ m/s). Figure 3 shows the group velocity dispersion curves for the aluminum plate, as well as the frequency spectrum for incident waves ranging below the A1 cutoff fd value where reflections and mode conversions are small.

Wave forms at three different time steps obtained in the calculation are shown in Fig. 4. Gray scale on the surface denotes the displacement in the vertical direction. The incident A0 mode reaches the elliptical defect, and then it propa-

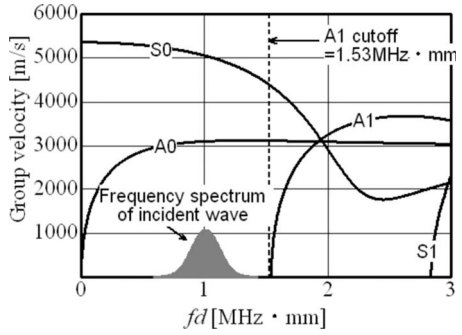


FIG. 3. Group velocity dispersion curves for an aluminum plate, along with the frequency spectrum of the incident wave used in the calculation.

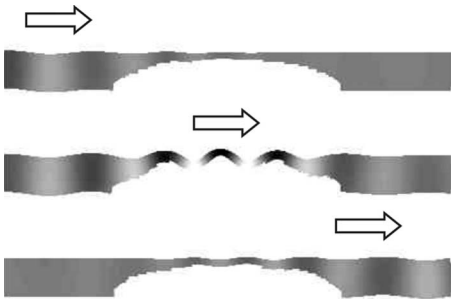


FIG. 4. Propagation of the A0 mode around a defect.

gates with keeping the mode shape of flexural vibrations. When propagating through smaller thickness, the out-of-plane displacement of the A0 mode becomes larger. Finally, the transmitted wave returns to the original amplitude in the intact thick region. This amplitude change is caused by the fact that almost all of the energy of the incident A0 mode propagates through the defect region in this frequency range.

In other words, when an A0 mode in the frequency range of small reflection and mode conversion is incident on a rounded defect, we can roughly estimate the thickness by measuring the normal amplitude at the surface of the defect. Now, assume that the A0 mode emitted from an ultrasonic transducer, such as angle-beam transducers, is detected by a laser interferometer over a surface of an object. Though the A0 mode can be emitted efficiently using the angle-wedge whose angle is determined by Snell's law, laser beam direction and focus must be controlled and adjusted very carefully for stable measurements by the laser interferometer from a distance.

On the other hand, the authors consider the case when normal loading is applied to the surface of a plate and an A0 mode is detected using an angle-beam transducer. Taking into account the reciprocal theorem, the wave form of the A0 mode is identical to the normal vibration detected at the source by the angle-beam transducer. That is, when the normal vibration is applied to a thin region of a plate and an A0 mode in the frequency range of small reflection and mode conversion is detected remotely, the amplitude of the detected wave form becomes large. By contrast, if a vibration is applied at a thick region, the amplitude of the detected signals is small.

When a laser beam is used for ultrasonic generation, it can be easily rastered by mirrors and emitted from a remote distance (e.g., 10 m away). Therefore, remote rapid thickness measurement can be realized by using a scanning laser source[12,13] (SLS) and angle-beam transducers to detect an A0 mode in the frequency range of small reflection and mode conversion.

IV. SCOPE OF THE THICKNESS MEASUREMENT

The thickness measurement technique using a SLS requires an A0 mode propagating through defects with small reflections and mode conversion. Therefore, it can be estimated that this technique has limitations in defect shape,

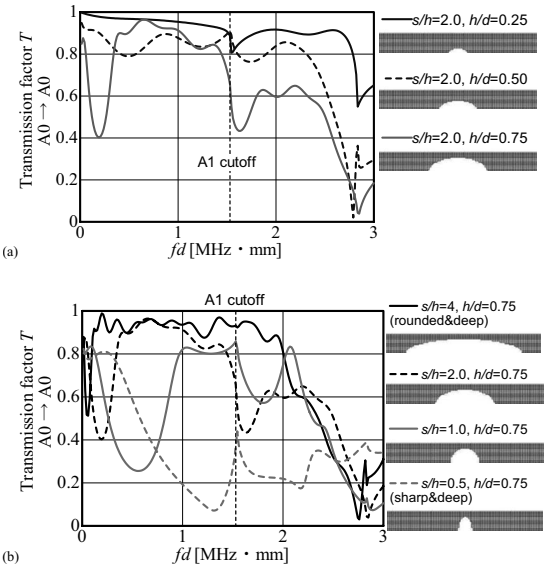


FIG. 5. Transmission factors for A0 incidence at a rounded defect. (a) Different depth but fixed shape. (b) Different shape but fixed depth.

depth, and ultrasonic frequency. In this section, the scope of the measurement is investigated for aluminum plates with defects of various shapes.

First, consider the transmission factors for A0 incidence shown in Fig. 5. Figure 5(a) shows a rounded defect of varying depth for a fixed shape $s/h=2.0$. On the other hand, Fig. 5(b) plots the transmission factors for a deep defect of fixed depth $h/d=0.75$ but variable shape. Above the A1 cutoff, the transmission factors become small as the frequency increases. On the other hand, below the cutoff, the transmission factors are over 0.8 across the fd range of about 0.5–1.4 MHz mm except sharp and deep defects $s/h=1.0$, $h/d=0.75$ and $s/h=0.5$, $h/d=0.75$.

Transmission factors for the A0 mode at $fd=1.0$ MHz mm for defects with different s/h and h/d values are plotted in Fig. 6. The general trend is that the transmission factors become smaller for sharper and deeper defects.

Generation of Lamb waves by a SLS and detection of the A0 mode at a left remote intact region are next simulated. When point loading is applied on the surface of a plate, several modes propagate in both directions. Since the amplitude of each mode can be separately calculated in SAFE

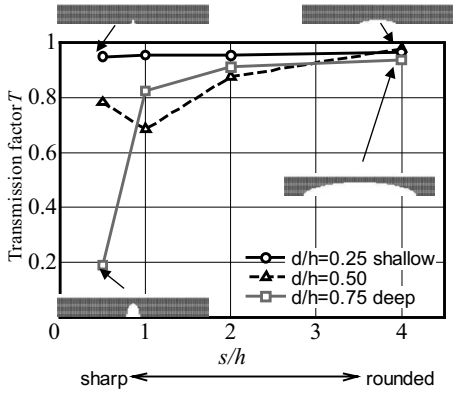


FIG. 6. Transmission factors for A0 incidence at a rounded defect of various shapes and depths.

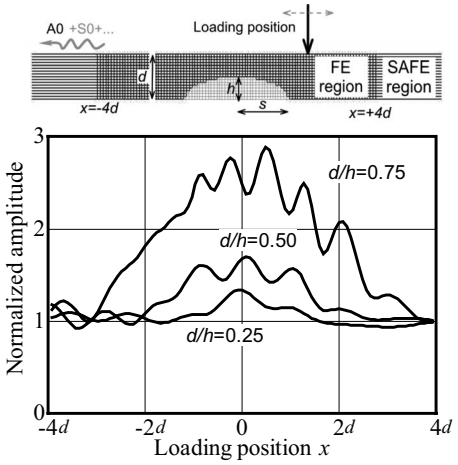


FIG. 7. Normalized amplitude for displacement in the thickness direction of an A0 mode for defects with various depths (d/h=0.25,0.5,0.75) but the same shape (s/h=4.0).

regions, the normal displacements of an A0 mode at the left-SAFE region are obtained, as shown in Fig. 7. Similar to Fig. 1, the FE region is from $x=-4d$ to $+4d$, and a harmonic vibration of fd=1.0 MHz mm is applied to the upper surface of the FE region. The center of the elliptical defect is at x=0, and the horizontal axis is in multiples of the thickness d. The normal displacement of A0 mode at the left region is normalized by the value when the normal loading is applied at the intact point $x=+4d$. The amplitude distributions for defects of three different depths (d/h=0.25,0.5,0.75) and fixed shapes s/h=4.0 are shown.

The amplitude distributions are roughly in inverse proportion to the depths d/h. However, the modulation in the defect region implies that there are some reflections and mode conversions, especially in a deep defect of d/h=0.75

The inverse amplitude of the distributions in Fig. 7 is graphed in Fig. 8 for the harmonic wave fd=1.0 MHz mm.

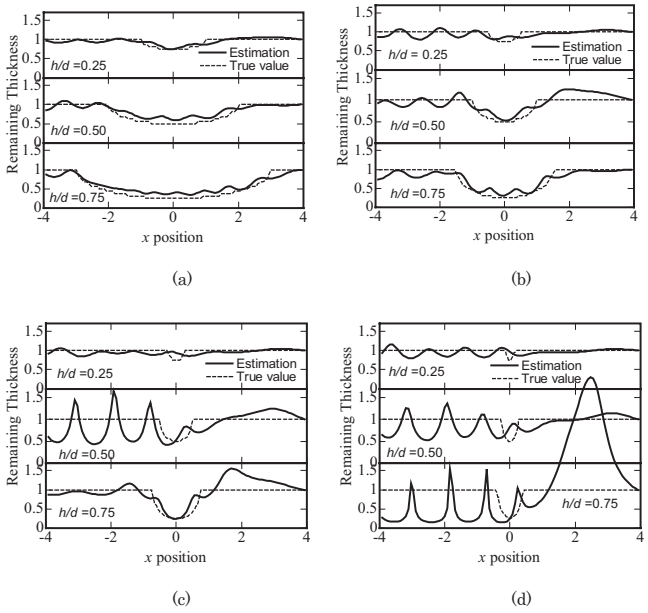


FIG. 8. Estimated remaining thickness for defects of various shapes and depths: (a) s/h=4.0, (b) s/h=2.0, (c) s/h=1.0, and (d) s/h=0.5.

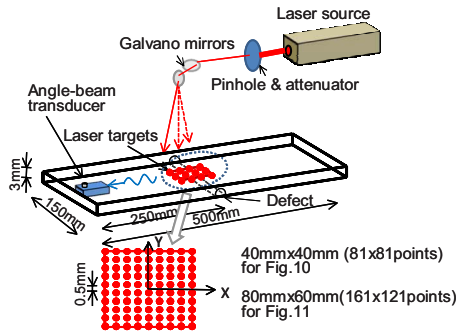


FIG. 9. (Color online) Sample and hardware setup.

The dashed lines denote the defect shape used in the calculation, while the solid lines are the estimated remaining thickness obtained from the amplitude distributions. For a rounded defect such as s/h=4 in Fig. 8(a), good agreement is found due to the small reflections and mode conversion. On the other hand, for sharper and deeper defects, which generate more reflected waves and larger mode conversions, the estimated values do not agree as well. For example, in Fig. 8(c) for s/h=1.0 and Fig. 8(d) for s/h=0.5, the peaks at the left of the defect are caused by the reflections from the defect, while the large values at the right are caused by the fact that the transmitted waves are small due to the sharp and deep defects.

V. EXPERIMENTS

In order to verify the remote thickness measurements using a SLS, experiments were carried out using aluminum plates of 3.0 mm thickness. Figure 9 shows the sample and experimental setup. A Q-switched Nd/yttrium aluminum garnet laser beam (Quantel Brilliant Ultra, 532-nm wavelength, 20-Hz rate, 7.2-ns pulse duration, and 32-mJ pulse energy) was scanned rapidly in the x and y directions by two galvano mirrors placed about 300 mm away from the plate. The laser spot diameter was adjusted to about 1.0 mm by a pinhole to improve the spatial resolution of the images. The laser energy was attenuated to prevent damaging the surface of the specimens. An A0 Lamb-wave mode was detected by an angle-beam transducer composed of a 400-kHz 1-3 piezoelastic composite and a resin wedge. The detected radio-frequency signals were amplified by 56 dB by a preamplifier. Then, the signals were filtered by a fourth-order Butterworth low-pass filter of 400 kHz and a high-pass filter of 600 kHz to extract components below and above the A1 cutoff frequency of fd=1.53 MHz mm. Only direct signals were gated to prevent the influence of edge reflections. The signals were not averaged so that they could be measured as quickly as possible. The measurement spots on the aluminum plates were 81×81 points in 0.5-mm steps over a range of 40×40 mm^2 in Fig. 10. In these cases, the distributions of estimated remaining thickness were obtained in about 10 min. Since the entire process of galvano scanner control, data acquisition, and display of results is now sequentially operated by LABVIEW software for the purpose of laboratory study, the measurements can be further speeded up by optimized acquisition done in ultrasonic microscopes.

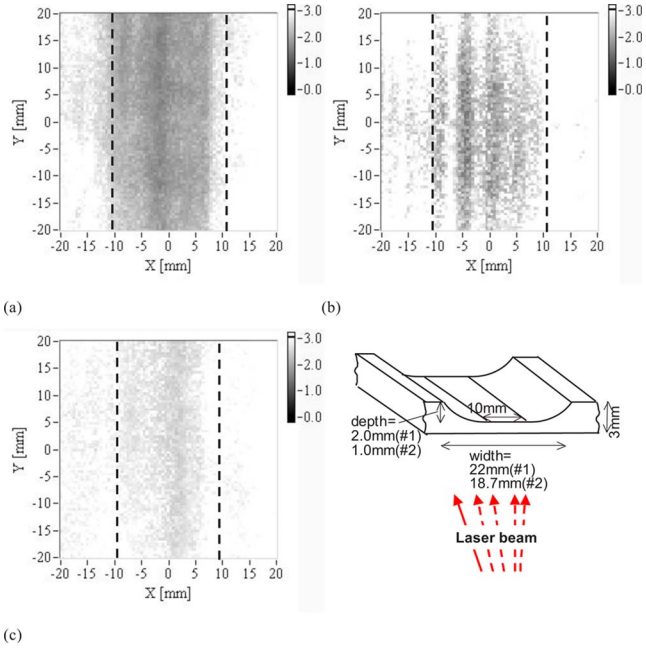


FIG. 10. (Color online) Estimated thickness distributions. (a) At defect 1, and a 400-kHz LPF was applied to the signals. (b) At defect 1, a 600-kHz HPF was applied to the signals. (c) At shallow defect 2, a 400-kHz LPF was applied to the signals.

The specimens used were aluminum plates measuring $150 \times 500 \times 3.0$ mm^3. Groove defects of 2.0-mm (1) and 1.0-mm (2) depths were machined at the center of the aluminum plates to match the calculation model. Both defects had a flat zone of 10-mm width at the bottom, and the edges were cut by a ball end mill of 20-mm diameter. The cross-section of these defects is shown in Fig. 10

Figure 10 shows the distribution of estimated remaining thickness obtained when a laser beam is incident on the flat surface of plates #1 and #2. In Fig. 10(a), the signals were filtered by a 400-kHz low-pass filter to extract the frequency range below the A1 cutoff. The thickness distributions were obtained by inverting the normalized amplitude of the signals. In this case, the average value of the amplitudes detected at 81 points on the right edge of the figure is used for normalization. The dashed lines denote the edges of the dent, which agree with the area of the signal having large amplitude.

Figure 10(b) was obtained from the same experimental setup, but the signals were filtered by a 600-kHz high-pass filter in order to show the result when unsuitable frequency range is used. In contrast to Fig. 10(a), several gray bands can be seen in the defect region due to large reflections and mode conversions in this frequency range.

In order to compare the distributions of remaining thickness for defects of different depths, Fig. 10(c) shows the distribution for a grooved defect of 1.0-mm depth (2). Similar to the case of a deep defect in Fig. 10(a), a laser beam was incident on the flat surface, and the signals were filtered by a 400-kHz low-pass filter. The difference between the intact area of 3.0-mm thickness and the defect area is smaller than for the deep defect in Fig. 10, which represents that the defect in specimen 2 is shallower than that in specimen 1.

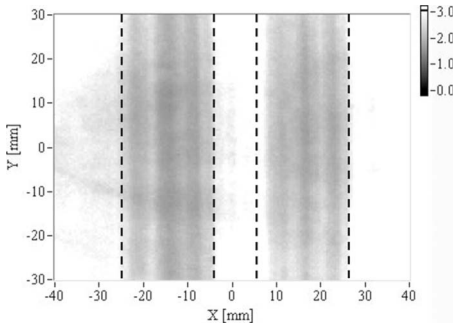


FIG. 11. Estimated thickness distribution for specimen 3 with two deep defects. The laser beams were incident on the flat surface, and a 400-kHz LPF was applied to the signals.

To simulate more complex and realistic defects, specimen 3 has two grooved defects of 2.0-mm depth. The width and depth of defects are the same as 1 in Fig. 10(a), and the centerline of the two defects are 30 mm apart. Figure 11 shows the estimated remaining thickness obtained from the amplitude. Similar to Figs. 10(a) and 10(c) for 1 and 2, a laser beam was incident on the flat surface of the plate, and a 400-kHz low-pass filter was applied to the signals. Two gray bands representing the grooved defects of 2.0-mm depth are clearly shown in Fig. 11, and the remaining thickness agrees roughly with the true value for both bands. This result demonstrates that remaining thickness can be evaluated even for the case of plural defects.

In these cases, remaining thickness in whole regions could be obtained with sufficient accuracy. However, directivity of an angle-beam transducer and attenuation and dispersion of an A0 mode may cause errors in an image of remaining thickness. In such cases, directivity of a receiving transducer can be improved by redesigning a transducer or by using multiple transducers, and variation in sensitivity due to attenuation and dispersion should be corrected from the propagation distance.

VI. CONCLUSIONS

The present study described a rapid thickness measurement technique using the amplitude of the A0 mode of Lamb waves generated by laser emission. A combination calculation technique using a SAFE method and an ordinary FE method revealed that the A0 mode in the frequency range of $fd = 1.0$ MHz mm below the A1 cutoff propagates through a shallow rounded defect with small reflections and mode conversions. The calculations also demonstrated that the A0 amplitude generated at a point source is roughly inversely proportional to the thickness at the source point. From such characteristics, the distributions of remaining thickness were experimentally obtained for aluminum plates with one and two grooved defects at high speed using a SLS and an angle-beam transducer.

Since laser-scanning using a galvano scanner is practical at high speeds, thickness measurements can be carried out much faster than in conventional contact-type thickness measurements by hand. Use of suitable equipment would provide a thickness image with the same speed as ultrasonic microscopes.

An axisymmetric $L(0,1)$ mode propagates in pipes with similar characteristics as the A0 mode of a Lamb wave. Therefore, use of the $L(0,1)$ mode would enable this technique to be applied to thickness measurements of pipes.

[1] I. A. Victorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[2] J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, New York, 1999).
[3] P. Cawley, M. J. S. Lowe, D. N. Alleyne, B. Pavlakovic, and P. Wilcox, "Practical long range guided wave testing: Application to pipes and rails," Mater. Eval. **61**, 66–74 (2003).
[4] H. Kwun, S. Y. Kim, and G. M. Light, "The magnetostrictive sensor technology for long range guided wave testing and monitoring of structures," Mater. Eval. **61**, 80–84 (2003).
[5] J. Li and J. L. Rose, "Excitation and propagation of nonaxisymmetric guided waves in a hollow cylinder," J. Acoust. Soc. Am. **109**, 457–468 (2001).
[6] T. Hayashi, K. Kawashima, Z. Sun, and J. Rose, "Analysis of flexural mode focusing by a semi-analytical finite element method," J. Acoust. Soc. Am. **113**, 1241–1248 (2003).
[7] T. Hayashi and M. Murase, "Defect imaging with guided waves in a pipe," J. Acoust. Soc. Am. **117**, 2134–2149 (2005).
[8] J. Davis and P. Cawley, "The application of synthetically focused imaging techniques for high resolution guided wave pipe inspection," in *Review of Progress in Quantitative NDE*, edited by D. Thompson and D. Chimenti (Plenum, New York, 2007), Vol. **26**, pp. 681–688.
[9] D. N. Alleyne, M. J. S. Lowe, and P. Cawley, "The reflection of guided waves from circumferential notches in pipes," J. Appl. Mech. **65**, 635–641 (1998).
[10] A. Demma, P. Cawley, M. Lowe, and A. G. Roosenbrand, "The reflection of the fundamental torsional mode from cracks and notches in pipes," J. Acoust. Soc. Am. **114**, 611–625 (2003).
[11] F. Benmeddour, S. Grondel, J. Assaad, and E. Moulin, "Study of the fundamental Lamb modes interaction with symmetrical notches," NDT & E Int. **41**, 1–9 (2008).
[12] P. A. Fomitchov, A. K. Kromin, S. Krishnaswamy, and J. D. Achenbach, "Imaging of damage in sandwich composite structures using a scanning laser source technique," Composites, Part B **35**, 557–562 (2004).
[13] J. Takatsubo, B. Wang, H. Tsuda, and N. Tooyama, "Generation laser scanning method for the visualization of ultrasounds propagating on a 3-D object with an arbitrary shape," J. Solid Mechanics and Materials Eng. **1**, 1405–1411 (2007).
[14] G. R. Liu and J. D. Achenbach, "Strip element method for stress analysis of anisotropic linearly elastic solids," J. Appl. Mech. **61**, 270–277 (1994).
[15] S. K. Datta, A. H. Shah, R. L. Bratton, and T. Chakraborty, "Wave propagation in laminated composite plates," J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[16] T. Hayashi, W.-J. Song, and J. L. Rose, "Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example," Ultrasonics **41**, 175–183 (2003).
[17] T. Hayashi and J. L. Rose, "Guided wave simulation and visualization by a semi-analytical finite element method," Mater. Eval. **61**, 75–79 (2003).
[18] Y. A. Al-Nassar, S. K. Datta, and H. Shah, "Scattering of Lamb waves by a normal rectangular strip weldment," Ultrasonics **29**, 125–132 (1991).
[19] Y. Cho and J. L. Rose, "A boundary element solution for a mode conversion study on the edge reflection of Lamb waves," J. Acoust. Soc. Am. **99**, 2097–2109 (1996).
[20] J. M. Galan and R. Abascal, "Numerical simulation of Lamb wave scattering in semi-infinite plates," Int. J. Numer. Methods Eng. **53**, 1145–1173 (2002).
[21] T. Hayashi, "Guided wave animation using semi-analytical finite element method," in Proceedings of the 16th World Congress on Nondestructive Testing (2004), pp. 786–817.

Contribution of crosstalk to the uncertainty of electrostatic actuator calibrations

Qamar A. Shams and Hector L. Soto
NASA Langley Research Center, Mail Stop 238, Hampton, Virginia 23681

Allan J. Zuckerwar
Analytical Services and Materials, 107 Research Drive, Hampton, Virginia 23666

(Received 19 March 2009; revised 10 June 2009; accepted 11 June 2009)

Crosstalk in electrostatic actuator calibrations is defined as the ratio of the microphone response to the actuator excitation voltage at a given frequency with the actuator polarization voltage turned off to the response, at the excitation frequency, with the polarization voltage turned on. It consequently contributes to the uncertainty of electrostatic actuator calibrations. Two sources of crosstalk are analyzed: the first attributed to the stray capacitance between the actuator electrode and the microphone backplate, and the second to the ground resistance appearing as a common element in the actuator excitation and microphone input loops. Measurements conducted on 1/4, 1/2, and 1 in. air condenser microphones reveal that the crosstalk has no frequency dependence up to the membrane resonance frequency and that the level of crosstalk lies at about −60 dB for all three microphones—conclusions that are consistent with theory. The measurements support the stray capacitance model. The contribution of crosstalk to the measurement standard uncertainty of an electrostatic actuator calibration is therewith 0.01 dB. [DOI: 10.1121/1.3167483]

PACS number(s): 43.38.Kb, 43.58.Vb [NHF] Pages: 1107–1110

I. INTRODUCTION

Crosstalk in electrostatic actuator calibrations is defined as the ratio of the microphone response to the actuator excitation voltage at a given frequency with the actuator polarization voltage turned off to the response, at the excitation frequency, with the polarization voltage turned on. Because crosstalk produces microphone output signals unrelated to the membrane displacement, it is a source of uncertainty in electrostatic actuator calibrations. This paper will describe circuit models representing the respective crosstalk mechanisms, an analysis yielding the magnitude and frequency dependence of the crosstalk, the results of an experiment to measure the crosstalk in condenser microphones of three different sizes (1, 1/2, and 1/4 in.), and guidelines for estimating the calibration uncertainty.

II. CIRCUIT MODELS AND ANALYSIS

A. Origin of the crosstalk

The hardware and circuit elements of the electrostatic actuator setup are identified in Fig. 1. A description of the principle of operation can be found elsewhere[1] and is not repeated here. For the present purpose, the electrostatic pressure generated by a harmonic excitation voltage has three components: an increment to the existing static pressure, a harmonic component, and a second harmonic component. In the absence of the actuator polarization voltage E_a the harmonic component would disappear if it were not for the crosstalk. Thus the crosstalk produces a harmonic output signal even when the membrane is not excited at the harmonic frequency. There are two physical mechanisms by which crosstalk can take place: electric field leakage through the stray capacitance between the actuator electrode and the microphone backplate, and an inadvertent feedback voltage generated in the resistance in the ground connection to the microphone membrane.

The stray capacitance mechanism is based on the fact that the actuator polarization voltage E_a (typically 800 V) is different from the microphone polarization voltage E_0 (typically 200 V). As a result, electric flux lines originating at the actuator electrode EA (and lead wire) bypass the membrane M and reach the backplate BP; they are thus unproductive in exciting the membrane into motion but provide a contribution to the output voltage e_0. The flux lines easily penetrate the walls of the microphone preamplifier (to which the microphone cartridge is closely attached) because their penetration depth is very large compared to the wall thickness. For example, using the known formula for the penetration depth,[2]

$$\delta = (\pi f \mu \sigma)^{-1/2}, \tag{1}$$

and values for frequency $f = 100$ kHz, magnetic permeability $\mu = 1.26 \times 10^{-6}$ V s/(A m), and electrical conductivity $\sigma = 1.4 \times 10^{6}$ $(\Omega\ \mathrm{m})^{-1}$ (stainless steel), one finds $\delta = 0.001\ 34$ m (0.0528 in.), compared to a typical wall thickness of 0.000 325 m (0.0128 in.). Thus a fraction

$$\exp(-0.000\ 325/0.001\ 34) = 0.78$$

of the flux lines succeeds to penetrate through the walls of the preamplifier.

The resistance R_f in the ground connection to the membrane appears in both the EA and microphone loops. The voltage across R_f generated in the actuator loop appears as an input voltage in the microphone loop, even when the membrane is unexcited.

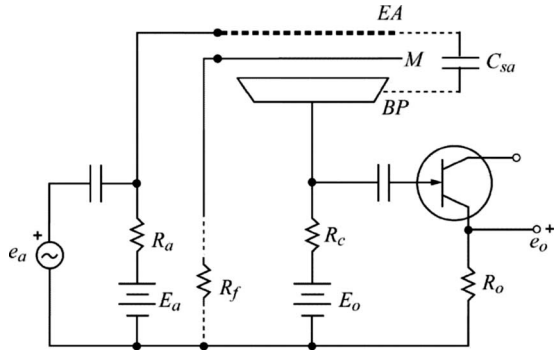


FIG. 1. Hardware and circuit elements of the electrostatic actuator setup. Legend: EA slotted or perforated actuator electrode, M and BP microphone membrane and backplate, E_a and R_a actuator polarization voltage and resistance, E_0 and R_c microphone polarization voltage and resistance, e_a actuator excitation voltage, e_0 and R_0 the microphone output voltage and source resistance of the follower, and C_{sa} and R_f the hypothesized stray capacitance and ground resistance.

In the following circuit models, each crosstalk mechanism is considered independently of the other.

B. The stray capacitance model

The circuit is defined in Fig. 2 and the circuit elements in the caption. The mechanical impedance of the membrane and the electromechanical coupling factors for the actuator and microphone are, respectively,

$$Z_m = j\omega M_M + R_A + \frac{1}{j\omega C_M} \approx \frac{1}{j\omega C_M}, \tag{2}$$

$$\psi_a = \frac{C_a E_a}{\pi a_M^2 h_a}, \tag{3}$$

$$\psi_t = \frac{C_{to} E_o}{\pi a^2 h_t}, \tag{4}$$

where M_M, R_A, and C_M are the membrane mass, damping resistance, and compliance, C_a is the actuator-membrane capacitance, a_M is the membrane radius, a is the effective membrane radius, ω is the angular frequency, and h_a and h_t are the actuator-membrane and backplate-membrane gaps. In the approximation shown in Eq. (2), the membrane mechanical impedance can be approximated by that of its compliance

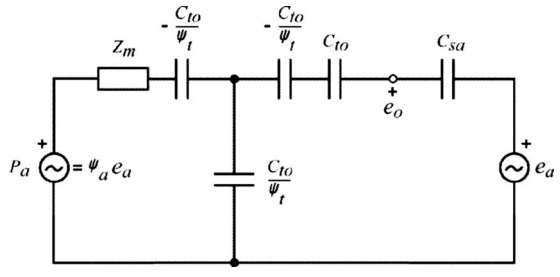


FIG. 2. Stray capacitance model. Legend: Z_m mechanical impedance of the membrane, C_{to} polarized membrane-backplate capacitance, C_{sa} stray actuator electrode-backplate capacitance, ψ_a and ψ_t electromechanical coupling factors for the actuator and microphone (see text), e_a and e_o input voltage to the actuator and output voltage of the microphone, and p_a electrostatic pressure generated by e_a.

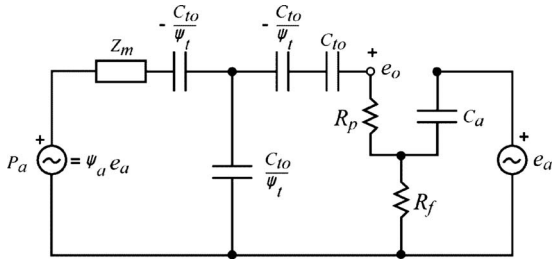


FIG. 3. Ground resistance model. Legend: R_p the parallel combination of R_c and the input resistance (not shown) of the source follower, and C_a actuator electrode-membrane capacitance.

at frequencies sufficiently below the membrane resonance frequency.

The output voltage e_o is generated by two sources: a direct source e_a and a dependent source p_a (proportional to e_a). The direct source is coupled to the output through the stray actuator-backplate capacitance C_{sa}. As can be seen from Eq. (3), when the actuator polarization voltage is turned off, the dependent source vanishes and the output depends on the direct source alone. Let the crosstalk X_a be defined as

$$X_a = 20 \log[(e_o/e_a)_0/(e_o/e_a)_P], \tag{5}$$

where $(e_o/e_a)_0$ is the transfer function with the actuator polarization voltage turned off and $(e_o/e_a)_P$ is the transfer function with the actuator polarization voltage turned on. With the approximation given in Eq. (2), analysis of the circuit of Fig. 2 yields the following expression for the crosstalk in decibels:

$$X_a = -20 \log\left[1 + \left(\frac{\psi_a \psi_t C_M}{C_{sa}}\right)\left(\frac{1}{1 - \frac{\psi_t^2 C_M}{C_{to}}}\right)\right]. \tag{6}$$

Noteworthy is the fact that the crosstalk is independent of frequency, at least over the range where approximation (2) holds.

C. The ground resistance model

In the circuit of Fig. 3 the stray capacitance C_{sa} is replaced with the combination of circuit elements C_a, R_f, and R_p, where R_p is defined in the figure. Analysis of the circuit yields the following expression for the crosstalk in decibels:

$$X_a = -20 \log\left[1 + \left(\frac{\psi_a \psi_t R_P C_M}{R_f C_a}\right)\left(\frac{1}{1 - \frac{\psi_t^2 C_M}{C_{to}}}\right)\right]. \tag{7}$$

The crosstalk again is independent of the frequency. Hence evaluation of C_{sa} and R_f from measured values of crosstalk will permit determination of the prevailing mechanism of the crosstalk.

III. EXPERIMENTAL METHOD

The test objects were three air condenser microphones: B&K types 4144 (1 in.), 4192 (1/2 in.), and 4939 (1/4 in.) connected to a type 2669 preamplifier.

The electrostatic actuator tests took place in an S&V Solutions type 4298 test chamber, where the microphones

were installed in a B&K type UA1284 microphone stand and their protective grids removed. The actuator electrodes were types UA0023 (1 in.), UA0033 (1/2 in.), and UA0033 with DB0264 adapter (1/4 in.). A microphone calibration module type 5001 provided an 800 V polarization voltage to the actuator electrode and 26 dB gain to the applied ac signal, provided by an HP model 3314A function generator. The output of the preamplifier was applied to a data acquisition system (see below). The ambient temperature was measured on a mercury thermometer, the pressure on a B&K type UZ004 barometer, and the humidity on a Vaisala HMP454 hygrometer.

The data acquisition system was based on a B&K Pulse 24-bit analyzer, comprising an input/output module 3110, control module 753L, and a power supply module 2826, from which power to the preamplifier was drawn (including a 200 V microphone polarization voltage). The data were processed in fast Fourier transform format, comprising a block size of 800 lines, flat-top filter, 100 averages, and maximum input settings of 22.36 mV for the 1/4 and 1/2 in. microphones, and 70.7 mV for the 1 in. microphone. The analyzer bandwidth was set to 6.4 kHz for measurements up to 4000 Hz and to 12.8 kHz for the 8000 Hz measurements.

The first step in the procedure was to calibrate a test microphone with a B&K type 4228 pistonphone. The microphone sensitivity on the Pulse was adjusted until the response read 124 dB at 251 Hz. The sensitivities were measured to be 3.94, 11.5, and 43.6 mV/Pa for the 1/4, 1/2, and 1 in. microphones, respectively.

After microphone installation, the actuator was excited at single tones in octave intervals from 250 to 8000 Hz. The input level was adjusted until the output signal reached 94 dB at 250 Hz (but not readjusted thereafter). This was found to be 2.240, 3.407, and 1.475 V for the 1/4, 1/2, and 1 in. microphones, respectively. At each frequency three measurements of the output voltage were taken: with the actuator polarization voltage turned on, with the actuator polarization voltage turned off, and with no applied signal (background). The microphone polarization voltage, however, remained at 200 V for all the measurements. After each measurement set the temperature (24.7–25.5 °C), pressure (1008–1009 hPa), and humidity (53%) readings were recorded.

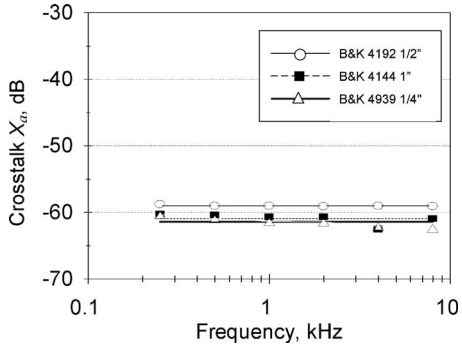


FIG. 4. Measured crosstalk, as defined in Eq. (5), versus frequency.

IV. RESULTS

Measured values of the crosstalk, expressed as the logarithmic difference implied in Eq. (5), are shown in Fig. 4 for the three microphones. The measured values for the microphone response with the actuator polarization voltage turned off are corrected for background noise. This is significant only at 250 Hz, the corrections being 1.59, 0.30, and 0.14 dB for the 1/4, 1/2, and 1 in. microphones, respectively. The

TABLE I. Microphone and electrostatic actuator parameters.

		Microphone (B&K type)			
Quantity	Units	4939 (1/4 in.)	4192 (1/2 in.)	4144 (1 in.)	Notes
a_M	m	0.003 175	0.006 35	0.012 7	[a]
a	m	0.001 97	0.003 78	0.007 10	[b]
h_t	μm	18	20	25	[c]
C_{to}	pF	6.0	19.9	56.1	[d]
V	mm^3	0.6	8.8	148	[e]
h_a	μm	400	400	500	[c]
C_a	pF	0.701	2.80	8.97	[f]
E_o	V	200	200	200	[c]
E_a	V	800	800	800	[c]
R_a	MΩ	10	10	10	[c]
R_p	GΩ	4	4	4	[c]
C_M	m^5/N	4.23×10^{-15}	6.20×10^{-14}	1.04×10^{-12}	[g]
ψ_t	C/m^3	5.46	4.43	2.83	[h]
ψ_a	C/m^3	0.044 3	0.044 3	0.028 3	[h]

[a]Half of nominal cartridge diameter.
[b]Calculated from parallel plate capacitance formula, approximately equal to backplate radius.
[c]Technical Documentation: Microphone Handbook: Vol. 1, Bruel & Kjaer A/S, Naerum (1996).
[d]Microphone calibration chart.
[e]Equivalent volume, Bruel & Kjaer product data sheet.
[f]Calculated from parallel plate capacitance formula.
[g]Calculated from equivalent volume: $C_M=V/(1.4\times101\,325)$.
[h]See Eqs. (3) and (4).

TABLE II. Crosstalk and derived circuit elements.

Quantity	Units	Microphone (B&K type) 4939 (1/4 in.)	4192 (1/2 in.)	4144 (1 in.)
X_a	dB	−61.4	−59.0	−60.9
	⋯	0.000 85	0.001 1	0.000 90
C_{sa}	pF	8.89×10^{-7}	1.454×10^{-5}	8.864×10^{-5}
R_f	Ω	5 076	20 757	39 532

level of crosstalk is found to be close to –60 dB for all three microphones, and to have no systematic dependence on the frequency as predicted by both theoretical models. The estimated standard uncertainty of the measurements is ±0.5 dB, related primarily to the microphone calibration, the data acquisition system, and the background noise.

For the purpose of determining the prevailing crosstalk mechanism, the values of the parameters appearing in Eqs. (6) and (7) are listed in Table I. Several of these are based on simplifying assumptions. In evaluations using the parallel plate capacitance formula, the stray capacitance and the input capacitance to the preamplifier are ignored. The membrane radius is taken to be half the nominal diameter, although the actual radius is somewhat smaller. The membrane-backplate gap is that of the polarized cartridge, but the "pull" of the electrostatic actuator polarization voltage is ignored. The effective actuator electrode-membrane gap is actually larger than the physical gap due to the influence of the actuator openings on the electric field,[3] but this effect is ignored. The equivalent volume, used to compute the membrane compliance, is a nominal value and not specific to a tested microphone. The frequencies of the measurement are assumed sufficiently low that acoustical loading by the actuator electrode is ignored.

Mean values of the crosstalk measurements, both in logarithmic and linear representations, are listed in Table II, along with evaluations of the stray capacitance C_{sa} and ground resistance R_f from Eqs. (6) and (7). The values of the ground resistance are unreasonably high. A more physically viable value of the ground resistance equal to 0.1 Ω leads to crosstalk values of −155 to −172 dB for the three microphones. Thus the measurements support the stray capacitance model for the prevailing mechanism of crosstalk.

The calibration standard uncertainty due to crosstalk for the cases considered is

$$U_X = 20\log(1+10^{X_a/20}) = 20\log(1+10^{-60/20})$$
$$= 20\log(1.001) = 0.009 \text{ dB}.$$

Coincidentally, the example illustrated in Ref. 4, Table C.1, lists a standard uncertainty due to crosstalk of 0.01 dB in close agreement to the above. A tenfold increase in the stray capacitance C_{sa} will lead to a calibration standard uncertainty of 0.086 dB. Consequently, the microphone mounting and lead wire to the actuator electrode must be arranged very carefully in order to suppress this source of uncertainty.

V. CONCLUSIONS

The experimental results support the stray capacitance model as the source of crosstalk in electrostatic actuator calibrations and reveal no systematic dependence on the frequency, as predicted by both theoretical models. The level of crosstalk is found to be close to −60 dB for all three microphone sizes, and the contribution to the calibration standard uncertainty near 0.01 dB.

[1]E. Fredericksen, "Electrostatic actuator," in *AIP Handbook of Condenser Microphones*, edited by G. S. K. Wong and T. F. W. Embleton (AIP, New York, 1995), Chap. 15.
[2]J. D. Kraus, *Electromagnetics* (McGraw-Hill, New York, 1953).
[3]P. V. Bruel, "The accuracy of condenser microphone calibration methods: Part II," B&K Technical Review, 1-1965.
[4]IEC, "Measurement microphones—Part 6: Electrostatic actuators for determination of frequency response," 61094-6, IEC, Geneva, 2004.

Uncertainty model for contact instability prediction

Antonio Culla[a)]
Department of Mechanics and Aeronautics, University of Rome "La Sapienza," Rome 00184, Italy

Francesco Massi
Université de Lyon, CNRS, INSA-Lyon, LaMCoS UMR5259, F-69621 Villeurbanne, France

(Received 17 July 2008; revised 16 June 2009; accepted 26 June 2009)

Contact between sliding bodies can cause vibrations leading to instability. The problem of squeal due to high frequency noise from brake systems is due to unstable vibrations generated at the contact interface between the pad and disk. Squeal noise is characterized by extreme unpredictability due to large uncertainties on the values of parameters of the system. Parametrical complex eigenvalue analysis is a common tool used to predict squeal instability. In this paper a substructured linear finite element model of a simplified brake system is studied. A parametrical analysis is focused on a test case and compared to experimental results. The analysis is developed as a function of the parameters assumed to be the most influential but also the most uncertain: friction coefficient and the parameters driving the dynamics of the system. The uncertainties are accounted for by considering parameters such as random variables. A Monte Carlo simulation and a probabilistic technique are performed simultaneously to study the probability of squeal occurrence. Finally, a reduced model based on the transfer function calculated at the contact is developed to perform the analysis with reduced computational effort.
 [DOI: 10.1121/1.3183376]

PACS number(s): 43.40.At, 43.50.Lj, 43.20.Ks, 43.40.Qi [JGM] Pages: 1111–1119

I. INTRODUCTION

Friction-induced vibrations are a common issue in mechanical design.[1] The coupling between system vibration and local contact behavior leads to several wear problems and sound emissions.

The problem of high frequency vibrations in automotive brakes, ranging from 1 to 20 kHz, is still unresolved despite several decades of research. Squeal is a difficult subject, partly because of its strong dependence on multiple parameters and also because the mechanical interactions in brake systems are very complicated, encompassing nonlinear contact problems at the friction interface. General agreement exists[2–4] regarding the origin of such vibrations, i.e., unstable coupling between two modes of the brake system. The literature includes several works on the numerical[5] and experimental[2] investigation of squeal instability which highlight the modal lock-in[3,6–8] between two specific modes of the system, one of which becomes unstable.

Complex eigenvalue analysis (CEA) is a numerical tool often used for predicting squeal instability.[5] This instability is associated with the positive real part of the eigenvalues of the system. Over the last two decades different works have attempted to predict squeal by correlating CEA with experimental squeal.[9] Nevertheless, the high complexity of brake systems does not permit the successful prediction of squeal instability by modeling their real dynamics. In fact, the brake device is characterized by high modal density in the frequency range of interest and its dynamics are very sensitive to small parametric variations. Moreover, a single system can be subject to uncertainties due to mass production (tolerances, assembly, etc.) and to service conditions (brake pressure, wear, temperature, etc.). Since it is impossible to build an accurate deterministic model it is convenient to consider an uncertain model whose imprecisely known physical parameters are taken into account as random variables.

In order to develop this analysis a Monte Carlo simulation is performed by varying the model's friction coefficient and driving parameters known to affect the system's dynamics. Proper probability density functions (pdfs) are imposed to determine the samples of the considered random variables, and the pdfs of the real parts of the system eigenvalues are determined. Consequently, the probability of squeal occurrence is estimated.

A simplified model is developed[10] to analytically assess the instability of the system in the case where the friction coefficient is a random variable. This approach is useful because of the reduction in the computational effort required to predict instability.

The work presented here uses a model of a simplified brake apparatus designed specifically to reduce the number of influential parameters. Moreover, a substructured finite element model (FEM) of the system is developed beforehand to reduce computational effort.

II. THE FEM

A. Linear model

The model considered in this paper [Fig. 1(b)] represents a simplified brake [Fig. 1(a)]. The geometry of the model is related to the geometry of the experimental set-up designed for squeal reproduction and analysis.[11,12] The small contact

[a)]Author to whom correspondence should be addressed. Electronic mail: antonio.culla@ uniroma1.it

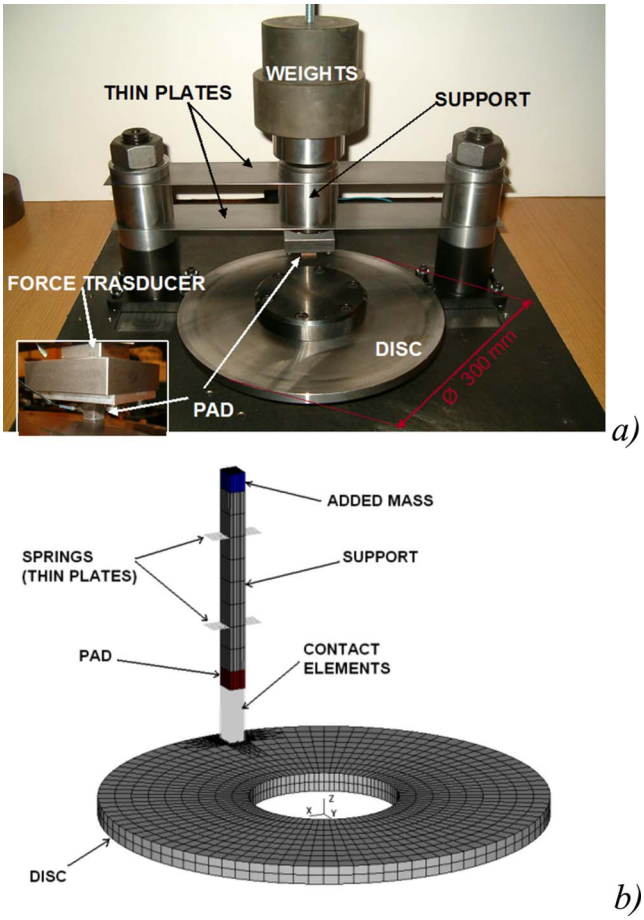


FIG. 1. (Color online) (a) Experimental set-up and (b) FEM of the experimental set-up.

area between pad and disk allows for low coupling between the bending modes of the pad or support (tangential direction) and bending vibration of the disk (normal direction). In particular, the modes are characterized by localization of the energy in the disk, the support (caliper), and the pad and are referred to here as disk, support, and pad modes, respectively.[11,12] ANSYS, a commercial FEM software, is used to investigate the dynamics of the brake system and calculate its complex eigenvalues as a function of different parameters. The SOLID45 element is used to mesh all the solid components of the system. The brake disk consists of an annular volume that is clamped on the internal radius and free on the external one. The brake pad consists of a cube 10×10 mm^2. The pad support is modeled by a beam $10\times10\times100$ mm^3, and a thin, high density layer (5 mm) is attached to the top of the support to simulate the mass added by the weights placed in the experimental set-up. Four rows of springs, two on each side, hold the beam in the horizontal (friction force) direction and model the thin aluminum plates that hold the pad support in the experimental set-up [Fig. 1(a)]. Proportional damping has been introduced in the model with $\alpha=0.02$ and $\beta=10^{-7}$. In order to perform a CEA of the brake assembly, linear elements are introduced to simulate the contact. This approximation is used to properly linearize the contact in the vicinity of the equilibrium point, i.e., the steady sliding position. The aim of this model is to capture the onset of instability by assuming it increases under linear conditions. The most commonly used linear schematization of the contact surface for modal analysis is the penalization technique. This can be obtained by a stiffness element (MATRIX27) that links each pair of nodes at the contact surface between the disk and the pad. The resulting element stiffness matrix is asymmetric, i.e.,

$$\begin{bmatrix} F^p \\ N^p \\ F^d \\ N^d \end{bmatrix} = \begin{bmatrix} 0 & \mu k & 0 & -\mu k \\ 0 & k & 0 & -k \\ 0 & -\mu k & 0 & \mu k \\ 0 & -k & 0 & k \end{bmatrix} \begin{bmatrix} x^p \\ z^p \\ x^d \\ z^d \end{bmatrix}, \tag{1}$$

where F and N are the tangential and normal contact forces, x and z are the tangential and normal displacements of the nodes of disk d and pad p, and k is the normal stiffness of the contact element. This type of schematization introduces tangential and normal contact forces proportional to the normal relative displacement between the disk and the pad. By introducing these contact elements, the stiffness matrix of the system becomes asymmetric, and a CEA can result in unstable eigenvalues. The simplified modeling of the contact by matrix (1) necessary for modal analysis does not account for friction damping and real contact distribution during sliding, neither of which are investigated in this work.

B. The substructured model

The FEM has been substructured due to the large number of solutions needed to perform the parametrical analysis and the considerable computational effort necessary to obtain each solution of the asymmetric problem. Substructuring is a procedure that condenses a group of finite elements into one element represented as a matrix. The substructure routine of ANSYS is used. Three main substructures of the brake system are considered here: the support, the pad, and the disk. The single-matrix element is known as a superelement. First, the single substructures are modeled separately, and the ANSYS substructuring routine calculates the superelement matrix. Then, the matrix is assembled and the superelements are connected to the complete model by master nodes defined beforehand. Finally, the assembled system is solved. When calculating the superelement matrix, the same elements and nodes as in the complete model are used.

The disk is characterized by a mapped mesh of 2412 elements. The nodes at the contact surface and at the inner radius, where the stresses are placed, are defined as master nodes. The solution of the substructured model efficiently approximates the solution of the complete model when the distribution of the master nodes allows describing the mode shapes in the frequency range of interest. Thus two other sets of master nodes are defined to include the nodes at the external diameter of the disk, which makes it possible to calculate the bending modes of the disk precisely.

The superelement representing the support includes the beam and the lateral spring modeling the thin plates. The master nodes are defined at the connection area with the pad and at the end of the lateral spring, where the support is constrained. Four lines of nodes corresponding to the four edges of the beam are also included in the master nodes to obtain the corrected bending modes of the support.

The pad is modeled by 1000 brick elements and master nodes are defined at its edges and at the two opposite surfaces that connect the pad with the support and the disk.

The substructures are then connected and constrained in the final model; 121 MATRIX27 contact stiffness elements are included to connect the disk and pad contact surfaces. The substructured model was validated by comparison with the complete FEM. The error between the calculated eigenvalues is less than 2% and the same instabilities are predicted.

III. CEA

The CEA provides the tool for tracing the instability regions of the system. The numerical eigenvalue extraction is performed by the damped method by ANSYS and repeated as a function of the driving parameters (friction coefficient, the Young modulus of the pad, and the stiffness of the springs holding the support). The analysis is first performed in the frequency range from 900 to 20000 Hz, and several unstable frequencies are identified. As shown in Ref. 11 squeal occurs only when a coincidence in frequency of two suitable system modes is obtained (lock-in), i.e., when a disk mode characterized by bending vibrations couples with a mode of the pad or the support characterized by vibration in the tangential direction. The squeal modal coupling (lock-in[12]) is identified from the CEA, if the parameters of the system vary, when the two eigenvalues start sharing the same imaginary part (frequency) and opposite real part (see Fig. 4). One of the two eigenvalues takes on a positive real part and becomes unstable. The two eigenvalues have the same frequencies until the modal coupling disappears (lock-out), and the real parts return to the original values due to system damping. It has to be pointed out that when structural damping is introduced into the numerical model, the two eigenvalues do not exactly coincide during the unstable coupling. This observation agrees with analytical[10] and numerical[13] works that are available in the literature.

A. Experimental validation

The simple dynamics of the system under investigation [Fig. 1(a)] allow reproducing several squeal conditions, and the modes of the systems involved in the unstable coupling can be identified.[11,12] The squeal events are identified by the harmonic sound emission, and the spectra of the system vibration exhibit a peak in the squeal frequencies over 40 dB (Fig. 2). The dynamics of the system were monitored during the tests to link their variation to increased instability, and experiments[11] showed that disk dynamics can couple either with the dynamics of the pad or with the dynamics of the support, leading to squeal instabilities in both cases. Five different squeal frequencies were found: 1566, 2467, 3767, 7850, and 10 150 Hz. These squeal conditions were obtained for defined values of the driving parameters, and all of them are easily reproducible. The FEM was used to numerically predict the squeal frequencies obtained experimentally. Table I summarizes the squeal frequencies obtained experimentally and the respective unstable range of frequencies predicted by the numerical model,[9] i.e., the frequency range covered by the unstable eigenvalue (with a positive real part) for the corresponding squeal instability. Good agreement was found, thereby validating the FEM.[9]

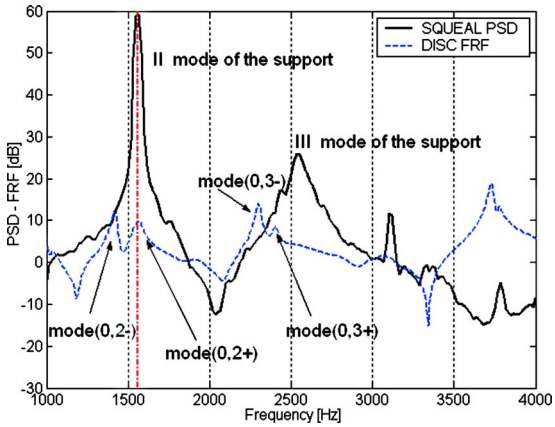


FIG. 2. (Color online) Tuning between the second mode of the support and the disk mode (0,2+) at 1566 Hz.

B. Test case

Here the analysis focuses on squeal instability at 1566 Hz, occurring when the second mode of the support couples with the bending mode of the disk characterized by two nodal diameters. Figure 2 shows the vibrations due to the unstable coupling.[11] The continuous line is the Power Spectral Density (PSD) of support acceleration in the tangential direction during squeal at 1566 Hz, while the dashed line is the measured frequency response function (FRF) of the disk. The coincidence in frequency between the modes leads to squeal. Figure 3 shows the parametrical analysis of the instability computed with the substructured FEM as a function of two main parameters.

- μ is the Coulomb friction coefficient between the disk and the pad. The uncertainty of this parameter is due to the variability of the contact conditions: surface topography, environmental conditions, etc.
- γ is the ratio between the Young modulus of the pad and the stiffness of the springs holding the support. This parameter is chosen because its variation allows reproducing the experimental variation in the system's natural frequencies.[9] It is chosen to represent the variations in several parameters capable of modifying the dynamics of a

TABLE I. Squeal frequencies: lock-in between the support modes, pad modes, and disk modes. Comparison between experimental squeal frequencies and FEM prediction.

Experimental squeal frequency (lock-in)	Predicted squeal (Hz)	Error (%)
II support mode and (0,2+) disk mode: squeal at 1566 Hz	[1535, 1550]	1.03
III support mode and (0,3+) disk mode: squeal at 2467 Hz	[2405, 2427]	1.6
Pad mode and (0,6+) disk mode: squeal at 7850 Hz	[7690, 7720]	1.65
Pad mode and (0,6+) disk mode: squeal at 10 150 Hz	[10190, 10195]	0.4

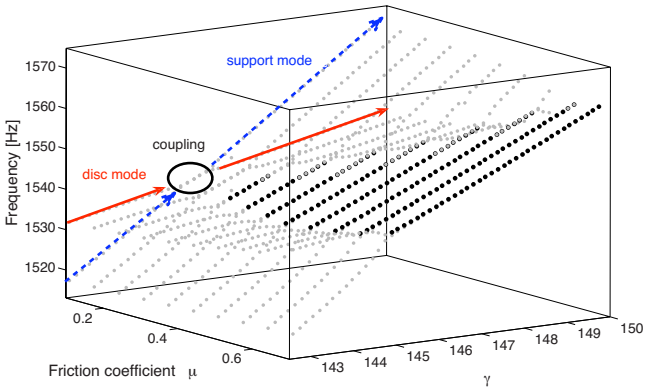


FIG. 3. (Color online) Parametrical eigenvalue analysis as a function of the friction coefficient and the system dynamics.

real brake system.

The gray dots indicate eigenvalues characterized by a negative real part (stable), while the black dots refer to eigenvalues characterized by a positive real part (unstable). For a fixed value of the friction coefficient the parametrical analysis highlights the lock-in phenomenon, i.e., the coalescence between two eigenvalues that start to have the same imaginary part (frequency) and opposing real part, until one of the two becomes unstable. When the friction coefficient is equal to 0.1, the two eigenvalues coalesce, but the real part does not become positive and the system is still stable [Fig. 4(a)]. When the friction coefficient is increased, the range of the parameters where the two eigenvalues coalesce becomes wider and one of the eigenvalues becomes unstable [Figs. 4(b) and 4(c)]. In fact, the friction forces couple the normal vibrations of the disk and the tangential vibrations of the support, allowing for self-excited vibrations of the system (instability).

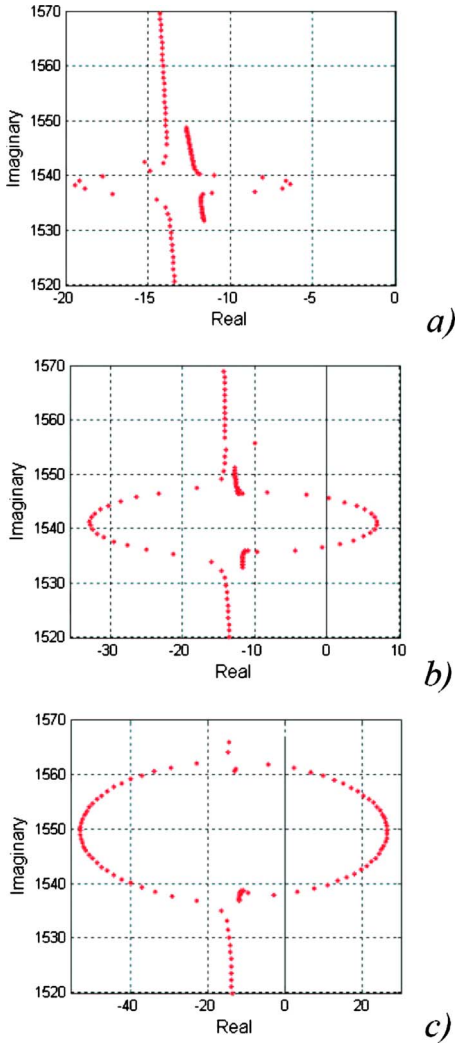


FIG. 4. (Color online) Locus plot highlighting the coalescence between the two modes [second support mode and (0,2+) disk mode] for three values of the friction coefficient: (a) μ=0.1, (b) μ=0.3, and (c) μ=0.57. For each curve, the variable parameter is γ and its range is the same as that shown in Fig. 3.

IV. RANDOM UNCERTAINTIES

The parametrical analysis shows that the squeal phenomenon is highly dependent on the values of the physical parameters of the system. In general these values are not known exactly, and thus an uncertainty relative to the nominal value of these parameters should be taken into account to predict the occurrence of squeal.

In this work the uncertainties are considered by assuming that a set of parameters of the system can be related to certain random variables. Then two studies are developed: a Monte Carlo simulation is performed to provide the probability distribution of the real part of the system eigenvalues when the friction coefficient, the pad's Young modulus, and the stiffness of the springs are random variables. A second analysis is developed by studying a simplified though meaningful and general model representing the contact between two sliding bodies (disk and pad). This model is based on the study of the transfer functions of the system calculated at the contact surface. They can be obtained either numerically or experimentally.

The model is proposed as a general tool to study the dynamic behavior of bodies in sliding contact with reduced computational effort and by using data (the transfer functions) easily available from the single bodies.

An analytical solution of the normal force between the two subsystems (disk and pad plus support) is provided and a parametric analysis performed to investigate the ability of the model to predict the instability. A probabilistic technique is then used to calculate the probability of squeal occurrence.

A. Monte Carlo simulation

In order to carry out a Monte Carlo simulation, a log-normal pdf is assumed for the friction coefficient:[14]

$$p_M(\mu) = \frac{1}{\mu\sigma_\mu\sqrt{2\pi}} e^{-(\log(\mu) - m_\mu)^2/2\sigma_\mu^2}, \tag{2}$$

where

$$m_\mu = \log(E[\mu]) - \frac{1}{2}\log\left(1 + \frac{\mathrm{Var}[\mu]}{E[\mu]^2}\right),$$

TABLE II. Statistical moments of the physical properties for the Monte Carlo simulation.

	Mean	Standard deviation
μ	0.25	0.084 (32.84%)
E (Pa)	7.76×10^8	2.54×10^7 (3.27%)
k (N/m)	5.35×10^6	1.05×10^5 (1.96%)

$$\sigma_\mu = \sqrt{\log\left(1+\frac{\mathrm{Var}[\mu]}{E[\mu]^2}\right)}. \tag{3}$$

The nominal value of the friction coefficient is its mean value, whereas the variation around it due to the real conditions at the contact (pad wear, temperature, humidity, etc.) is represented by its standard deviation. The suitability of the shape chosen for the pdf is decided by considering the nature of μ. One of the most important justifications for using the Gaussian distribution is the central limit theorem. In fact, the friction phenomenon can be considered as the sum of many independent factors summarized by the friction coefficient. Also, μ is positive in nature. The chosen log-normal pdf guarantees the respect of these two conditions.

Moreover a uniform statistical distribution is assumed for the γ coefficient influencing the dynamics of the system. Table II shows the significant statistical moments of the three parameters (friction coefficient, Young modulus of the pad, and stiffness of the spring holding the support) that are considered in the Monte Carlo simulation.

The choice of the uniform pdf depends on the impossibility of defining one value of E or k as being greater than another one.

Since the uniform pdf is bounded and bordered between a maximum and a minimum, the range of variation in the Young modulus E and stiffness k can be calculated by their statistical moments. The variation in E is 5.67% around its mean value, while the variation in k is 3.4%.

Using a random number generator, 4800 samples, which are combinations of the three parameters, are calculated by the substructured FEM to perform the Monte Carlo procedure. The pdf of the system eigenvalues is calculated, and the pdf of the real part of these eigenvalues is shown in Fig. 5.

The pdf obtained agrees with the results shown by the locus plots in Fig. 4. In Fig. 5, the two central peaks are the real parts of the eigenvalues related (under stable condition) to the modal damping imposed on the model. The symmetric distribution around the central peaks is due to the coalescence of the eigenvalues when the system is unstable. In such cases the real parts are opposite with respect to the starting values, and one of them can reach positive values (see Fig. 5). Figure 6 shows the pdf of the real parts of the two coupling eigenvalues. The smaller real part (in absolute value) corresponds to the smaller modal damping. When coupling occurs the mode with smaller damping moves toward positive values and becomes unstable: a part of the sample set takes on positive values. Figure 4 also shows that the eigenvalue with the smaller absolute value of the real part (lower damping) becomes unstable. It is known[10,15] that the less damped mode becomes unstable during squeal.

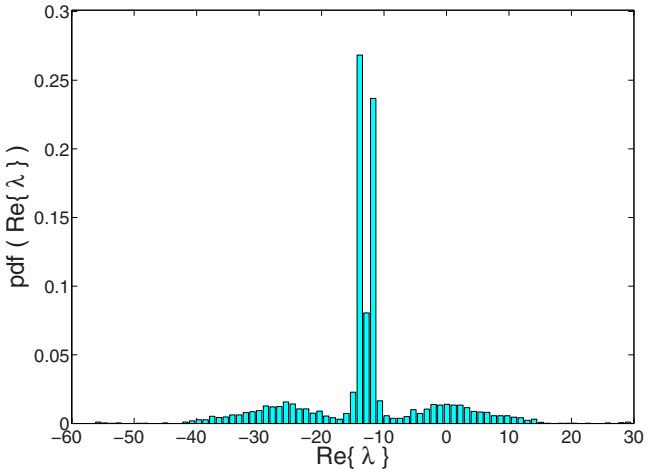


FIG. 5. (Color online) Histogram of the pdf of the system eigenvalue's real part.

For the specific probability distribution shape, the most meaningful statistical parameters are the median and the probability that the instability occurs, i.e., the probability of having eigenfrequencies with a positive real part.

Since only the real parts of the disk eigenvalues becomes positive (see Fig. 6), the statistical parameters are calculated for this pdf:

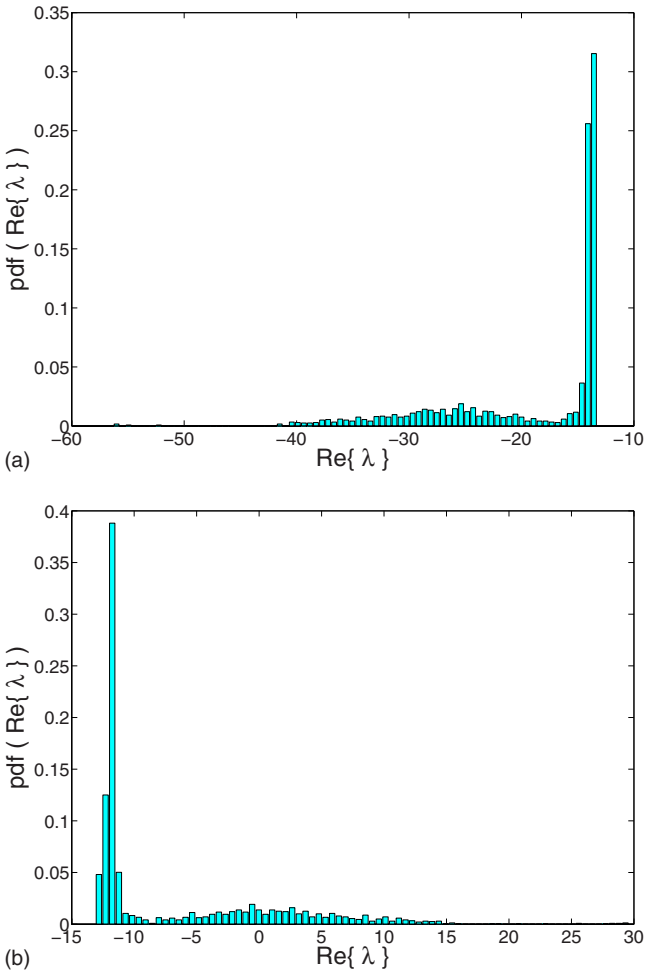


FIG. 6. (Color online) Histograms of the pdfs of the eigenvalue's real part: (a) support mode and (b) disk mode.

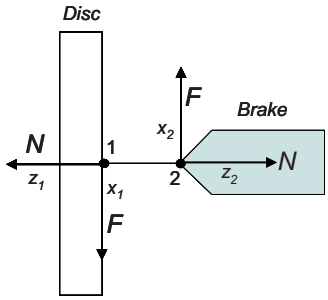


FIG. 7. (Color online) Diagram of the simplified coupling between disk and brake.

$$\text{Median}[\text{Re}\{\lambda\}] = -13.2,$$

$$P(\text{Re}\{\lambda\} > 0) = 0.22.$$

Since the median is negative the most probable value of $\text{Re}\{\lambda\}$ corresponds to a stable configuration. Since the pdf of $\text{Re}\{\lambda\}$ is normalized in the case studied, the probability that the instability occurs is 22%.

V. THEORETICAL APPROACH

In this section a simplified model is used to describe the contact between disk and pad.[6,7,10] One contact point is considered and Fig. 7 shows the scheme of the model considered. Here a distinction between the pad and support is neglected: these two subsystems together are known as "brake."

The total friction force F_T and the total normal force N_T are the sum of steady components F_0 and N_0 and fluctuating components F and N, respectively:

$$F_T = F_0 + F, \quad N_T = N_0 + N, \tag{4}$$

with $F \ll F_0$ and $N \ll N_0$. x and z are the displacements along the directions corresponding to F and N, respectively. The displacements at the contact points of the two subsystems are related to fluctuating forces F and N at the same points of the FRFs of each body (not in contact), as shown by the following equations:

$$z_1 = G_{11}N + G_{12}F,$$

$$x_1 = G_{21}N + G_{22}F,$$

$$z_2 = H_{11}N + H_{12}F,$$

$$x_2 = H_{21}N + H_{22}F, \tag{5}$$

where G_{ij} are the FRFs of the disk and H_{ij} are the FRFs of the brake. For the analysis presented the FRFs are calculated by the FEM of the single substructures. In order to solve the dynamic problem, two further relationships have to be considered. The first one links the tangential and normal forces to the friction coefficient:

$$F = \mu N. \tag{6}$$

The second equation is the kinematic relationship between the displacement of the contact points:

$$z_2 + z_1 = r, \tag{7}$$

where r is the roughness of the surface.

By solving the set of Eqs. (5)–(7), the normal force between the two subsystems is calculated and highlighted by the following equation:

$$N = \frac{r}{D(\omega)}, \tag{8}$$

with

$$D(\omega) = H_{11} + G_{11} + \mu(H_{12} + G_{12}).$$

Since G_{ij} and H_{ij} are the FRFs of stable systems, it follows that the coupled whole system is unstable if and only if D has at least one zero, i.e., $1/D$ has at least one pole, in the lower Fourier half-plane.[10]

The FRFs of the subsystems considered not in contact can be expressed by a linear combination of complex modes by the following equations:

$$G_{nm} = \sum_i \frac{\varphi_{di}(\xi_n)\varphi_{di}(\xi_m)}{\omega_{di}^2 + 2j\omega\omega_{di}\delta_{di} - \omega^2},$$

$$H_{nm} = \sum_i \frac{\varphi_{bi}(\xi_n)\varphi_{bi}(\xi_m)}{\omega_{bi}^2 + 2j\omega\omega_{bi}\delta_{bi} - \omega^2}. \tag{9}$$

Indices b and d remain for the brake and the disk, respectively, i is the mode number, φ_b and φ_d are the mode shapes, ω_b and ω_d are the natural frequencies, and δ_b and δ_d are the modal dampings. ξ_n and ξ_m are the coordinates of points n and m, respectively. Thus D can be rewritten as follows:

$$D(\omega) = \sum_p \frac{\varphi_{dp}(\xi_1)(\varphi_{dp}(\xi_1) + \mu\varphi_{dp}(\xi_2))}{\omega_{bp}^2 + 2j\omega\omega_{bp}\delta_{bp} - \omega^2} + \sum_q \frac{\varphi_{bq}(\xi_1)(\varphi_{bq}(\xi_1) + \mu\varphi_{bq}(\xi_2))}{\omega_{dq}^2 + 2j\omega\omega_{dq}\delta_{dq} - \omega^2}, \tag{10}$$

where ξ_1 and ξ_2 indicate the coordinates of points 1 and 2. These points correspond to the central nodes on the contact surface of the FEM of the pad. Therefore, the eigenvectors φ_b and φ_d are calculated at points 1 and 2 by FEM, while the natural frequencies ω_b and ω_d are provided to calculate $D(\omega)$. Equation (8) shows a relationship between normal force N and roughness r. Function $1/D$ is not the transfer function of any physical system; it provides only an input-output relationship. Therefore, the poles of $1/D$ determine the stability of the system:[10,16–19] the analysis of the instability is performed as a function of the system dynamics, considering surface roughness as the disturbance.

A two-mode approximation is performed to simplify the problem. This approach is valid and useful if the squeal involves two relatively isolated modes of the uncoupled subsystems.[10] Therefore, the denominator D can be written as follows:

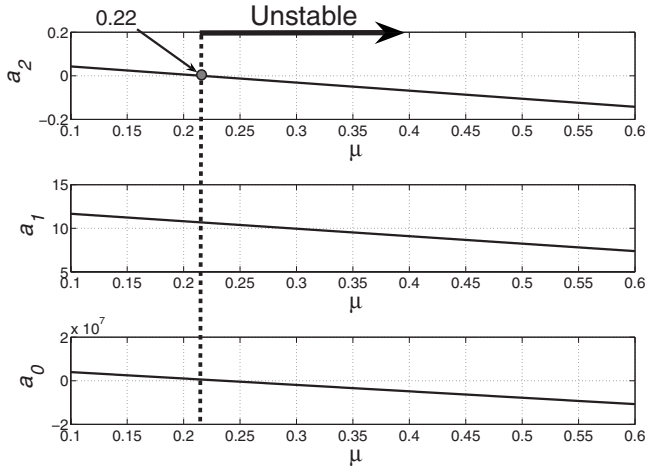


FIG. 8. Coefficients of Eq. (12) versus μ.

$$D(\omega) \cong \frac{\varphi_{b1}^2(\xi_1) + \mu\varphi_{b1}(\xi_1)\varphi_{b1}(\xi_2)}{\omega_{b1}^2 + 2j\omega\omega_{b1}\delta_{b1} - \omega^2} + \frac{\varphi_{d1}^2(\xi_1) + \mu\varphi_{d1}(\xi_1)\varphi_{d1}(\xi_2)}{\omega_{d1}^2 + 2j\omega\omega_{d1}\delta_{d1} - \omega^2}, \tag{11}$$

and its zeros are the roots of the following equation:

$$a_2(j\omega)^2 + a_1 j\omega + a_0 = 0, \tag{12}$$

where

$$a_2 = a_{20} + \mu a_{21},$$

$$a_{20} = -\varphi_{b1}^2(\xi_1) - \varphi_{d1}^2(\xi_1),$$

$$a_{21} = -(\varphi_{b1}(\xi_1)\varphi_{b1}(\xi_2) + \varphi_{d1}(\xi_1)\varphi_{d1}(\xi_2)),$$

$$a_1 = a_{10} + \mu a_{11},$$

$$a_{10} = 2(\varphi_{b1}^2(\xi_1)\delta_{d1}\omega_{d1} + \varphi_{d1}^2(\xi_1)\delta_{b1}\omega_{b1}),$$

$$a_{11} = 2(\varphi_{b1}(\xi_1)\varphi_{b1}(\xi_2)\delta_{d1}\omega_{d1} + \varphi_{d1}(\xi_1)\varphi_{d1}(\xi_2)\delta_{b1}\omega_{b1}),$$

$$a_0 = a_{00} + \mu a_{01},$$

$$a_{00} = \varphi_{d1}^2(\xi_1)\omega_{b1}^2 + \varphi_{b1}^2(\xi_1)\omega_{d1}^2,$$

$$a_{01} = \varphi_{b1}(\xi_1)\varphi_{b1}(\xi_2)\omega_{d1}^2 + \varphi_{d1}(\xi_1)\varphi_{d1}(\xi_2)\omega_{b1}^2. \tag{13}$$

The condition for system stability implies that the three coefficients of Eq. (12) have the same sign.

A. Parametrical analysis

As for the FEM a preliminary parametrical analysis of the instability is carried out to verify the validity of the reduced model.

By varying the friction coefficient between 0.1 and 0.6 it is shown (Fig. 8) that the system is stable up to μ=0.22; over this point the coefficient a_2 becomes negative, while the coefficient a_1 is positive over the whole range of μ.

This result agrees with the threshold value of the friction coefficient obtained by the CEA on the FEM (Fig. 3).

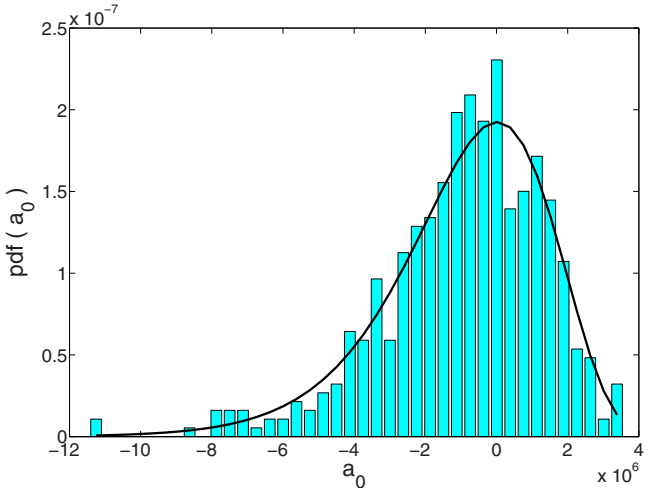


FIG. 9. (Color online) pdf of the coefficient a_0 of Eq. (12).

B. Probabilistic approach

The coefficients of Eq. (12) depend on the physical parameters of the systems (E and k) and on the friction coefficient μ. Their dependence on the Young modulus of the pad and on the stiffness of the support is not shown explicitly. In fact, a_2, a_1, and a_0 depend directly on the modal parameters of the system (normal modes and normal frequencies) which in turn depend on E and k. On the contrary, the dependence of the coefficients on μ is explicit.

In Sec. IV A μ, E, and k are considered random variables. Consequently the coefficients of Eq. (12) become stochastic variables. The randomness of the friction coefficient is described by imposing the log-normal pdf of Eq. (2) and that of E and k is described by the uniform distribution. The pdfs of coefficients a_2, a_1, and a_0 can be calculated analytically[20] if the pdfs of the random variables on which the coefficients depend are known. This work focuses only on the dependence on μ, as it appears explicitly in Eq. (13).[21]

By defining the following variables:

$$a_{2-20} = a_2 - a_{20},$$

$$a_{1-10} = a_1 - a_{10},$$

$$a_{0-00} = a_0 - a_{00}, \tag{14}$$

the pdfs sought are

$$p_{A_2}(a_2) = \frac{e^{[\log|a_{2-20}| - (\log|a_{21}| + m_\mu)]^2/-2\sigma_\mu^2}}{(a_{2-20})(\mathrm{sgn}(a_{21})\sigma_\mu)\sqrt{2\pi}},$$

$$p_{A_1}(a_1) = \frac{e^{[\log|a_{1-10}| - (\log|a_{11}| + m_\mu)]^2/-2\sigma_\mu^2}}{(a_{1-10})(\mathrm{sgn}(a_{11})\sigma_\mu)\sqrt{2\pi}},$$

$$p_{A_0}(a_0) = \frac{e^{[\log|a_{0-00}| - (\log|a_{01}| + m_\mu)]^2/-2\sigma_\mu^2}}{(a_{0-00})(\mathrm{sgn}(a_{01})\sigma_\mu)\sqrt{2\pi}}. \tag{15}$$

Figures 9–11 show the curves of the pdf coefficients compared with the corresponding histograms. These histo-

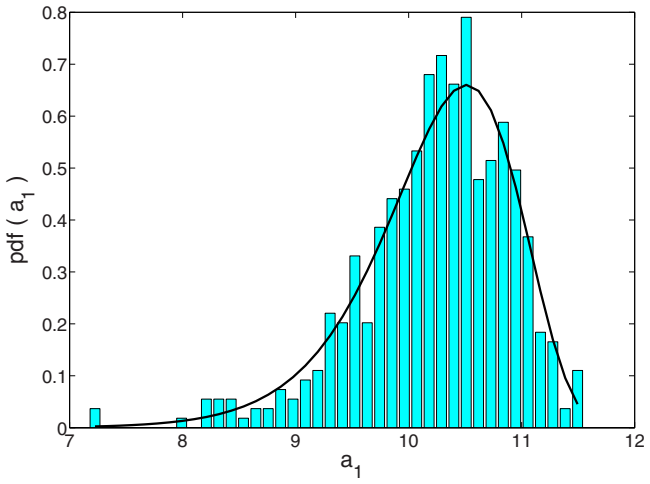


FIG. 10. (Color online) pdf of the coefficient a_1 Eq. (12).

gram are obtained by introducing into Eq. (13) the random values of μ calculated by Eq. (3). This comparison shows the accuracy of the analytical result.

The probability of the occurrence of instability can be calculated by integrating the pdf of a_2 in the negative range of its values. The result is

$$P(a_2 < 0) = \int_{-\infty}^{0} p_{A_2}(a_2) da_2 = 0.64.$$

This result disagrees with the result of the Monte Carlo simulation by the FEM.

Obviously, since the function mapping μ into a_2 is linear, agreement between the μ pdf and the a_2 pdf is good. On the contrary, the Monte Carlo simulation is performed by varying not only the friction but also the dynamics of the system so that the two eigenvalues are coincident only for a subset of the samples. The randomness of these parameters is not considered in the simplified system, and consequently the probability of squeal occurrence is higher than for the Monte Carlo result.

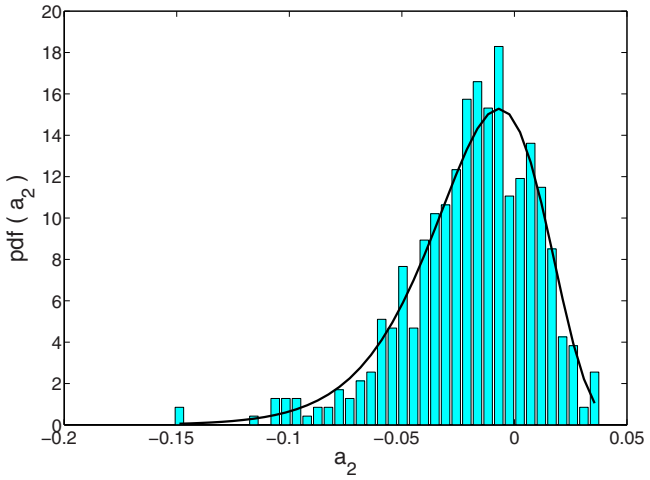


FIG. 11. (Color online) pdf of the coefficient a_2 Eq. (12).

VI. CONCLUDING REMARKS

This paper presented the results obtained by a study on the occurrence of squeal instability. Because of the sensitivity of squeal to the governing parameters and the uncertainties on the latter, a probabilistic approach is necessary for squeal prediction by CEA. Moreover, because of the complexity of brake systems, a simplified model developed for bodies in sliding contact was proposed to reduce computational effort. The inputs of this model are the transfer functions that can be measured or calculated on the single bodies not yet in contact. The dynamics of the single substructures were then coupled by the model of the contact, and the stability of the system was studied.

Initially a substructured FEM was built to simulate the system and a parametric analysis was developed in the case of variable friction coefficients and system dynamics. The occurrence of instability was determined by studying the sign of the complex eigenvalues.

A second analysis was performed on the substructured model by considering the randomness of the same parameters used in the parametric study. The probability of instability occurrence was calculated by a Monte Carlo simulation. Finally a simplified model was investigated to obtain and study the analytical relationship linking the parameters of the subsystems. This simple model contains the basic requisite of the complete model by introducing the transfer functions of the single substructures (without contact). They can be obtained either numerically or experimentally. The model was proposed as a general tool for studying the dynamic behavior of bodies in sliding contact that requires reduced computational effort and which uses data easily recoverable from single bodies (transfer functions). Parametrical and probabilistic approaches were provided to understand whether the simplified model can predict instability in comparison with the FEM and good agreement was obtained. Future work will focus on developing the model by introducing several contact points.

[1] A. Akay, "Acoustic of friction," J. Acoust. Soc. Am. **111**, 1525–1548 (2002).
[2] N. M. Kinkaid, O. M. O'Reilly, and P. Papadopoulos, "Automotive disc brake squeal," J. Sound Vib. **267**, 105–166 (2003).
[3] A. Akay, J. Wickert, and Z. Xu, "Investigation of mode lock-in and friction interface," Final Report, Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, PA (2000).
[4] M. R. North, "Disk brake squeal, a theoretical model," Technical Report 1972/5, Motor Industry Research Association, Warwickshire, England (1972).
[5] H. Ouyang, W. Nack, Y. Yuan, and F. Chen, "Numerical analysis of automotive disc brake squeal: A review," International Journal of Vehicle Noise and Vibrations **1**, 207–231 (2005).
[6] M. R. North, "Disk brake squeal," in *Braking of Road Vehicles*, edited by M. E. P. Limited (Automobile Division of the Institution of Mechanical Engineers, London, England, 1976), pp. 169–176.
[7] J. Flint and J. Hultn, "Lining-deformation-induced modal coupling as squeal generator in a distributed parameter disc brake model," J. Sound Vib. **254**, 1–21 (2002).
[8] Q. Cao, H. Ouyang, M. I. Friswell, and J. E. Mottershead, "Linear eigenvalue analysis of the disc-brake squeal problem," Int. J. Numer. Methods Eng. **61**, 1546–1563 (2004).
[9] F. Massi, L. Baillet, and O. Giannini, "Squeal prediction on a simplified brake system by complex eigenvalue analysis," in *Proceedings of International Conference on Noise and Vibration Engineering ISMA* (2006) (K.U. Leuven Department of Mechanical Engineering, PMA, Leuven, Belgium).

[10]P. Duffour and J. Woodhouse, "Instability of systems with a frictional point contact. Part 1: Basic modelling," J. Sound Vib. **271**, 365–390 (2004).
[11]F. Massi, O. Giannini, and L. Baillet, "Brake squeal as dynamic instability: An experimental investigation," J. Acoust. Soc. Am. **120**, 1388–1399 (2006).
[12]F. Massi, L. Baillet, O. Giannini, and A. Sestieri, "Brake squeal: Linear and nonlinear numerical approaches," Mech. Syst. Signal Process. **21**, 2374–2393 (2007).
[13]B. Herv, J. J. Sinou, H. Mah, and L. Jzquel, "Analysis of squeal noise in clutches and mode coupling instabilities including damping and gyroscopic effects," Eur. J. Mech. A/Solids **27**, 141–160 (2008).
[14]M. K. Ochi, *Applied Probability and Stochastic Process* (Wiley, New York, 1990).
[15]F. Massi and O. Giannini, "Effect of damping on the propensity of squeal instability: An experimental investigation," J. Acoust. Soc. Am. **123**, 2017–2023 (2008).
[16]R. H. Lyon, "Progressive phase trands in multi-degree-of-freedom systems," J. Acoust. Soc. Am. **73**, 1223–1228 (1983).
[17]R. H. Lyon, "Range and frequency dependence of transfer function phase," J. Acoust. Soc. Am. **76**, 1433–1437 (1984).
[18]M. Tohyama and R. H. Lyon, "Zeros of a transfer function in a multi-degree-of-freedom vibrating system," J. Acoust. Soc. Am. **86**, 1854–1862 (1989).
[19]M. Tohyama and R. H. Lyon, "Transfer function phase and truncated impulse response," J. Acoust. Soc. Am. **86**, 2025–2029 (1989).
[20]The pdf of a random variable Y depending on another random variable X is calculated by the following equation: $p_Y(y)=p_X(g^{-1}(y))|dg^{-1}/dy|$. $p_X(x)$ is the pdf of a random variable X, g is a functional on X, such that $Y=g(X)$, and ad g^{-1} is the inverse function of g (Ref. 14).
[21]I. Elishakoff, *Probabilistic Theory of Structures*, 2nd ed. (Dover, Mineola, NY, 1999).

Coupling of axial and transverse displacement fields in a straight beam due to boundary conditions

Jerry H. Ginsberg
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 6 January 2009; revised 19 June 2009; accepted 26 June 2009)

A classical simply-supported beam is modified by tilting the pad constraining the roller. The consequence is that the axial and transverse displacements of the beam are coupled by the boundary conditions. The eigenanalysis methodology is extended to this unusual situation where displacement components are coupled, even though each displacement has a different associated wavenumber. The dependence of the natural frequencies on the tilt angle is evaluated and typical results for the eigenfunctions are presented. The response of the beam to harmonic point force excitation is synthesized by constructing a modal series. Frequency response functions at the drive point for cases where the force is applied in the axial and transverse directions are computed. The results indicate that deviations of the tilt angle from zero affect the displacement in the direction that is orthogonal to the excitation much more than the driven displacement.
 [DOI: 10.1121/1.3183368]

PACS number(s): 43.40.Cw, 43.40.At [JGM] Pages: 1120–1124

I. INTRODUCTION

Almost every modern textbook in vibrations introduces the analysis of continua by studying the behavior of straight bars having constant cross-sectional properties. This is so because the field equations describing the cross-sectional displacement components associated with extension and flexure are uncoupled. Furthermore, if the cross-section is circular, torsion is uncoupled from extension and flexure. The consequence of this decoupling is that a vibration analysis reduces to the study of a single displacement variable that depends on one spatial coordinate. Reality differs greatly from this idealized model. If the transverse displacement is sufficiently large, the axial force resultant affects the flexural rigidity.[1] Even within the limited scope of the present work, which is linear vibration phenomena associated with very small displacements, coupling of displacement fields can occur for a variety of reasons. For example, a spring that is neither parallel nor orthogonal to the axis of the bar induces both axial and transverse forces when it is deformed, and the rotation due to flexural effects induces an axial acceleration if there is an attached mass whose center is not situated on the bar's axis.[2] More significantly, one seldom encounters a straight bar that is isolated; real systems use networks of bars interconnected at arbitrary orientation, with the consequence that flexural and extensional effects are always coupled.[3]

The fact that extensional waves in a bar are not dispersive, whereas flexural waves are, leads to some interesting issues when one considers their interaction. Detailed exploration of the phenomena is prohibitive for structural networks, and the role of any attachments to an isolated bar obscures one's ability to understand the physical implications of the coupling effects. The system that is explored here is quite simple: a beam that would be typed as "simply-supported" in its nominal configuration of a pin at one end and a roller at the other. The coupling of flexure and extension arises because the surface that guides the roller is not parallel to the beam's axis. Consequently, the normal force exerted by the roller on the beam is not perpendicular to the beam's axis. The corollary of the fact that this force always has both axial and transverse components is that neither axial nor transverse displacement can occur individually. The analysis will provide physical understanding of the interaction process. It also will provide insight to the degree to which deviations from an idealized configuration affect the vibrational properties of a system. The analysis begins with derivation of the coupled modes, which are used to analyze a forced response.

As depicted in Fig. 1, the system of interest is a beam having uniform cross-sectional properties that is pinned at its left end, designated $x=0$, and is supported by a roller at its right end. The roller rides on a plane that is tilted at angle θ from the horizontal. The configuration where $\theta=0$ corresponds to a simply-supported beam for flexure and a fixed-free bar for extension, whereas $\theta=90°$ corresponds to a cantilever beam for flexure and a fixed-fixed bar for extension. The excitation is taken to be a concentrated force applied at position x_F because the response to other types of excitation can be synthesized from this case. If the frequency range is sufficiently low in comparison to the fundamental thickness shear frequency, the displacement field is well described by the transverse displacement w and axial displacement u of the cross-section's centroid.

Analysis of the response of the bar in Fig. 1 could be carried out by deriving transfer functions in a frequency domain formulation. Such an analysis does not provide the physical insight afforded by modal analysis, which is the approach followed here. The eigensolutions for this system, which constitute a mathematical basis for constructing the response to any excitation, are derived first. They enable one to assess at a fundamental level the degree to which the displacement components couple. The derived modal prop-

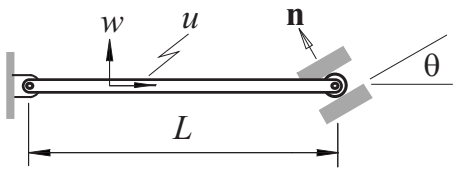


FIG. 1. System configuration and definition of the displacement components.

erties are then used to construct the response of the system to a harmonic concentrated force. An important aspect is evaluation of the degree to which a small deviation from the nominal horizontal roller ($\theta=0$) affects the modal and forced response properties.

II. MODAL PROPERTIES

The field equations governing the displacement components are

$$\rho A\ddot{w} + EIw^{iv} = F_x\delta(x-x_F),$$

$$\rho A\ddot{u} - EAu'' = F_x\delta(x-x_F), \tag{1}$$

where F_x and F_z are the components of the applied force, E and ρ are Young's modulus and the material density, and A and I are the cross-sectional area and area moment of inertia, respectively.

The boundary conditions for the left end, $x=0$, require that both displacement components and the bending moment be zero. At the right end, $x=L$, it is necessary that the displacement in the direction of the normal **n** to the roller pad be zero and that the resultant force perpendicular to **n** as well as the bending moment vanish. Thus, the displacement fields must satisfy

$$w = w'' = u = 0 \quad \text{at } x=0,$$

$$\left.\begin{array}{c} -u\sin\theta + w\cos\theta = 0, \\ EAu'\cos\theta - EIw'''\sin\theta = 0 \\ w'' = 0. \end{array}\right\} \quad \text{at } x=L, \tag{2}$$

To determine the eigenmodes that are the foundation for a modal analysis of forced response the displacement field for free vibration is described as a time harmonic response,

$$\left\{\begin{array}{c} u \\ w \end{array}\right\} = \left\{\begin{array}{c} \psi_u(x) \\ \psi_w(x) \end{array}\right\} e^{i\omega t}. \tag{3}$$

Substitution of this ansatz into the homogeneous version of Eq. (1) leads to ordinary differential equations governing the spatial functions:

$$\psi_u'' + \left(\frac{\alpha}{L}\right)^2\psi_u = 0, \quad \psi_w^{iv} - \left(\frac{\beta}{L}\right)^4\psi_w = 0. \tag{4}$$

The wavenumbers α and β depend on ω according to

$$\alpha = \frac{\omega L}{c}, \quad \beta = \left(\frac{\omega L}{c}\right)^{1/2}\left(\frac{L}{\kappa}\right)^{1/2}, \tag{5}$$

where c is the bar wave speed, $c=(E/\rho)^{1/2}$, and κ is the cross-section's radius of gyration, $\kappa=(I/A)^{1/2}$. The frequency ω is the same for both displacements with the consequence that the wavenumbers are related by

$$\alpha = \frac{\kappa}{L}\beta^2. \tag{6}$$

The characteristic equation associated with ψ_u in Eq. (4) is quadratic in β with two imaginary roots, whereas the corresponding equation associated with ψ_w is quartic with a pair of imaginary roots accompanied by a pair of real roots. The imaginary roots lead to sinusoidal functions, whereas the real roots are most conveniently represented in terms of hyperbolic functions. The resulting spatial functions have the same form as those for conventional beams,

$$\psi_u = a_1\sin\left(\alpha\frac{x}{L}\right) + a_2\cos\left(\alpha\frac{x}{L}\right),$$

$$\psi_w = b_1\sin\left(\beta\frac{x}{L}\right) + b_2\cos\left(\beta\frac{x}{L}\right) + b_3\sinh\left(\beta\frac{x}{L}\right) + b_4\cosh\left(\beta\frac{x}{L}\right). \tag{7}$$

Satisfaction of the boundary conditions at $x=0$ requires that

$$a_2 = b_2 = b_4 = 0. \tag{8}$$

The bending moment at $x=L$ vanishes if

$$b_3 = b_1\frac{\sin(\beta)}{\sinh(\beta)}. \tag{9}$$

At this juncture there are two undetermined coefficients, a_1 and b_1, and the displacement and resultant force boundary conditions at $x=L$ remain to be satisfied. Doing so with the simplified forms of ψ_u and ψ_w resulting from Eq. (6) and the preceding coefficient relations leads to a pair of linear equations for the coefficients,

$$[D(\beta,\kappa/L)]\left\{\begin{array}{c} a_1 \\ b_1 \end{array}\right\} = \{0\}, \tag{10}$$

where the elements of the coefficient matrix are

$$D_{1,1} = -\sin(\theta)C_{(u=0)}(\beta,\kappa/L),$$

$$D_{1,2} = \cos(\theta)C_{(w=0)}(\beta,\kappa/L),$$

$$D_{2,1} = \cos(\theta)C_{(F=0)}(\beta,\kappa/L),$$

$$D_{2,2} = \sin(\theta)C_{(S=0)}(\beta,\kappa/L). \tag{11}$$

The various $C_{(\,)}$ functions form the characteristic equations for the canonical problems corresponding to $\theta=0°$ or $\theta=90°$, as designated by their subscripts,

$$C_{(u=0)}(\beta,\kappa/L) = \sin\left(\frac{\kappa}{L}\beta^2\right),$$

$$C_{(w=0)}(\beta,\kappa/L) = 2\sin(\beta),$$

$$C_{(F=0)}(\beta,\kappa/L) = \cos\left(\frac{\kappa}{L}\beta^2\right),$$

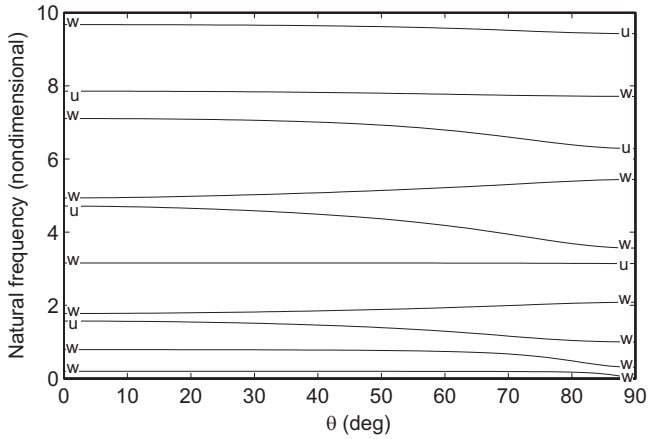


FIG. 2. Dependence of the natural frequencies on θ when $\kappa=L/50$.

$$C_{(S=0)}(\beta,\kappa/L)=\frac{\kappa}{L}\beta^2\left[\cos(\beta)-\frac{\sin(\beta)}{\sinh(\beta)}\cosh(\beta)\right]. \quad (12)$$

The consequence of this representation is that the characteristic equation from which the eigenvalues β are obtained combines the four equations for the canonical problems according to

$$|[D(\beta,\kappa/L)]|=\sin(\theta)^2C_{(u=0)}(\beta,\kappa/L)C_{(S=0)}(\beta,\kappa/L)+\cos(\theta)^2C_{(w=0)}(\beta,\kappa/L)C_{(F=0)}(\beta,\kappa/L)=0. \quad (13)$$

The appearance of sinusoidal functions in the constituent functions forming the characteristic equation means that the characteristic equation for specified κ/L has an infinite number of eigenvalues β_j that may be found by numerical methods. The corresponding natural frequencies are found from Eq. (5) to be

$$\omega_j=\frac{\kappa c}{L^2}\beta_j^2. \quad (14)$$

The dependence of the natural frequencies on the tilt angle θ is depicted in Fig. 2. The notation u or w annotating the curves at $\theta=0°$ and $\theta=90°$ denotes whether the mode for these special angles entails solely axial or transverse displacement. The fifth natural frequency has the interesting aspect that it is nearly independent of θ. Similar behavior is also observed for selected higher frequencies. This situation arises if an eigenvalue β_j is such that $C_{(u=0)}$ and $C_{(w=0)}$ are close to zero so that the characteristic equation is nearly satisfied for any θ.

The eigenfunction corresponding to a specific β_j is obtained by applying Eqs. (8) and (9) to Eq. (7),

$$\{\psi_j(x)\}=\begin{Bmatrix} a_{1,j}\sin\left(\frac{\kappa}{L}\beta_j^2\frac{x}{L}\right) \\ b_{1,j}\left[\sin\left(\beta_j\frac{x}{L}\right)+\frac{\sin(\beta_j)}{\sinh(\beta_j)}\sinh\left(\beta_j\frac{x}{L}\right)\right]\end{Bmatrix}. \quad (15)$$

Equations (10), only one of which is independent, govern the ratio $a_{1,j}/b_{1,j}$. If $\theta=0°$ or 90°, one of these equations is trivial, and one of the coefficients vanishes. A general set of equations governing any θ is obtained by treating the coefficient equations as an overdetermined set. Thus, it is required that

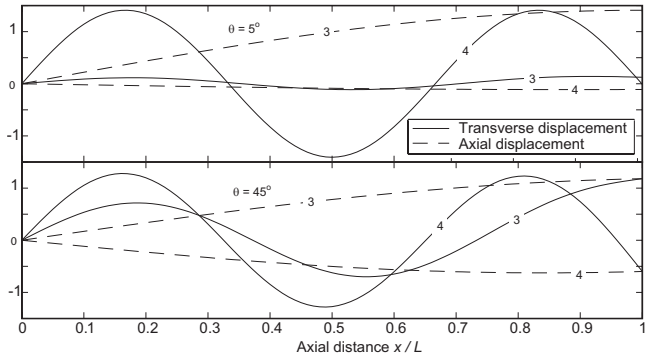


FIG. 3. Spatial dependence of the third and fourth modes when $\kappa=L/50$.

$$a_{1,j}\sin\theta\sin\left(\frac{\kappa}{L}\beta_j^2\right)=2b_{1,j}\cos\theta\sin\beta_j,$$

$$a_{1,j}\cos\theta\cos\left(\frac{\kappa}{L}\beta_j^2\right)=-b_{1,j}\frac{\kappa}{L}\beta_j\sin\theta\cos\beta_j\left[1+\frac{\tan(\beta_j)}{\tanh(\beta_j)}\right]. \quad (16)$$

Figure 3 compares the third and fourth modes, which are representative of the general behavior, for $\theta=5°$ and $\theta=45°$. Regardless of the mode number or value of θ the flexural displacement varies over a shorter scale than the axial displacement. Small deviations of θ from zero tend to affect the nominal axial modes more than the nominal flexural modes, and small deviations of θ from 90° have the same tendency. The overall ratio of displacement components becomes increasingly sensitive to changes in θ as the mode number increases.

III. FORCED RESPONSE

A modal series representation of the displacement fields is a linear sum whose terms consist of a product of each eigenfunction multiplied by a modal coordinate ξ_j that scales the mode's contribution,

$$\begin{Bmatrix} w \\ u \end{Bmatrix}=\sum_j\{\Psi_j(x)\}\eta_j(t). \quad (17)$$

In the preceding $\{\Psi_j(x)\}$ denotes normalized modes, which are obtained by selecting the as yet undefined coefficients, $a_{1,j}$ or $b_{1,j}$, such that the modal masses have a specified value. The definition used here is that all modal masses should equal the mass of the bar, ρAL. An expression for the modal mass is derived from the kinetic energy functional, which sums the contributions of the axial and transverse displacements. Substitution of Eq. (17) into that functional leads to a quadratic sum in the modal velocities,

$$T = \frac{1}{2}\int_0^L \rho A \begin{Bmatrix} \dot{w} \\ \dot{u} \end{Bmatrix}^T \begin{Bmatrix} \dot{w} \\ \dot{u} \end{Bmatrix} dx$$

$$= \frac{1}{2}\int_0^L \rho A \sum_j \{\Psi_j(x)\}^T \dot{\eta}_j(t) \sum_n \{\Psi_n(x)\} \dot{\eta}_n(t)$$

$$= \frac{1}{2}\rho AL \sum_j \dot{\eta}_j^2. \tag{18}$$

The decoupling of different modes in the preceding is a corollary of the orthogonality condition, while the absence of a factor inside the summation results from defining the modal mass to be ρAL. These properties are described by

$$\int_0^L \rho A \{\Psi_j(x)\}^T \{\Psi_n(x)\}^T dx = \rho AL(\mu_j a_{1,j}^2 + \epsilon_j b_{1,j}^2)\delta_{jn},$$

$$\mu_j = \int_0^L \psi_{u,j}^2 dx, \quad \epsilon_j = \int_0^L \psi_{w,j}^2 dx, \tag{19}$$

where $\psi_{u,j}$ and $\psi_{w,j}$ are the functions multiplying $a_{1,j}$ and $b_{1,j}$, respectively, in Eq. (15). Setting the modal mass to its designated value leads to

$$(\mu_j a_{1,j}^2 + \epsilon_j b_{1,j}^2) = 1, \tag{20}$$

which in combination with Eq. (16) fully defines the $a_{1,j}$ and $b_{1,j}$ values.

The normalized mode functions satisfy a second orthogonality condition, whose consequence is that the modes are both elastically and inertially decoupled. The corresponding potential energy function is

$$V = \frac{1}{2}\rho AL \sum_j \omega_j^2 \eta_j^2. \tag{21}$$

The set of generalized forces associated with the modal coordinates is derived from an evaluation of the virtual work done by the harmonically varying force, which is the product of the force and the virtual displacement of the point x_f where the force is applied, with the latter characterized by the modal series,

$$\delta W = \bar{F} \cdot \delta \bar{w} = \{\mathrm{Re}[F_z \; F_x](e^{i\omega t})\} \begin{Bmatrix} \delta w \\ \delta u \end{Bmatrix}$$

$$= \mathrm{Re}[F_z \; F_x] e^{i\omega t} \{\Psi_j(x_F)\} \delta \eta_j. \tag{22}$$

The coefficient of each virtual increment of a generalized coordinate is the associated generalized force, so

$$Q_j = \mathrm{Re}\,[F_z \; F_x]\{\Psi_j(x_F)\} e^{i\omega t}. \tag{23}$$

The equations of motion are obtained by forming Lagrange's equations using Eqs. (18), (21), and (23). The steady-state response induced by the time harmonic excitation is found by substituting $\eta_j = \mathrm{Re}(X_j e^{i\omega t})$ into those equations. The decoupled nature of the energy expressions leads to a simple expression for the X_j, which then is substituted into Eq. (17) to synthesize the displacement field. The result is

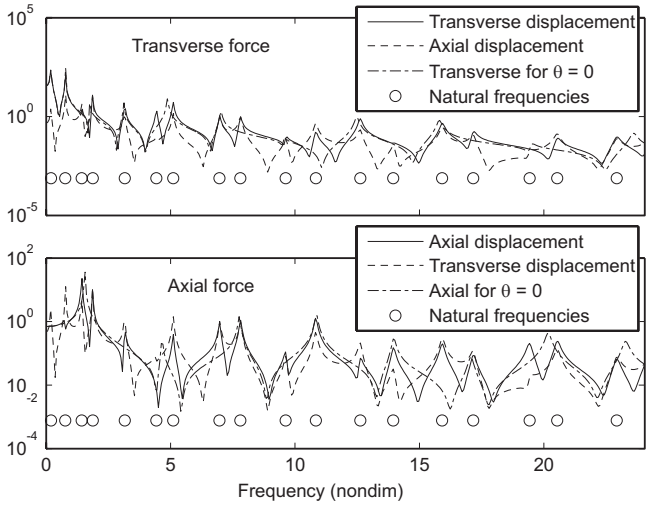


FIG. 4. Frequency response functions for a point force at $x_f = L/\sqrt{2}$ when $\theta = 45°$; $\kappa = L/50$, all $\sigma_j = 0.01$.

$$\begin{Bmatrix} w \\ u \end{Bmatrix} = \mathrm{Re} \sum_j \frac{[F_z \; F_x]\{\Psi_j(x_F)\}\{\Psi_j(x)\}}{\rho AL[\omega_j^2(1+\sigma_j) - \omega^2]} e^{i\omega t}, \tag{24}$$

where σ_j are modal loss factors that may be frequency dependent.

If one scales the natural frequencies ω_j and excitation frequency ω relative to c/L, the parameters that affect the displacement in Eq. (24) when it is scaled relative to $FL/\rho c^2 A$ are the nondimensional radius of gyration κ/L, the tilt angle θ, x_f/L, and the σ_j values. Results presented here consist of the displacement components at the drive point $x_f = L/\sqrt{2}$ for forces that are applied transversely, $F_x = 0$, $F_z = F$, and axially, $F_x = F$, $F_z = 0$. In all cases the bar is quite slender, $\kappa = L/50$. In Fig. 4, for $\theta = 45°$, both displacement components have comparable amplitudes at resonances, which occur at the natural frequencies. Away from the peaks, especially at the anti-resonances, there are substantial differences between the two displacement components. An interesting aspect is that the transverse displacements for $\theta = 0°$ and $\theta = 45°$ for the case of transverse excitation show much similarity around the lower resonances. In contrast, the axial displacements in the case of axial excitation for $\theta = 0°$ and $\theta = 45°$ are quite dissimilar. This is consistent with the previous observation that the axial displacement in a mode is generally more affected by θ than is the transverse displacement.

Figure 5 describes the situation when the support is slightly tilted, $\theta = 5°$. There is no significant difference between the transverse displacements in the direction of excitation for small θ and $\theta = 0°$. In contrast, the displacement in the direction orthogonal to the excitation is significant. Although it generally is smaller than the driven displacement, there are a few resonances where the undriven displacement exceeds the driven one. This is obviously different from $\theta = 0°$ case, where the displacement in the direction that is not driven is zero. However, when such situations are encountered, the displacement is substantially lower than the maximum amplitude observed at other resonances.

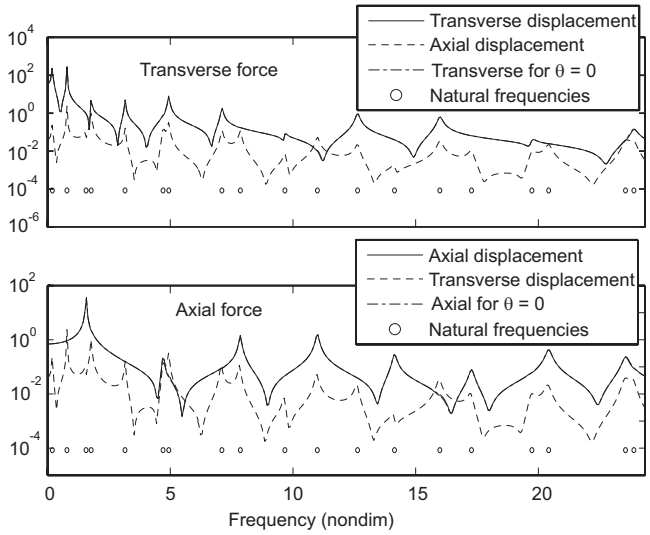


FIG. 5. Frequency response functions for a point force at $x_f=L/\sqrt{2}$ when $\theta=5°$; $\kappa=L/50$, all $\sigma_j=0.01$.

IV. DISCUSSION

Coupling of displacement components occurs in a wide variety of situations. Such coupling can occur as a result of the field equations, as in the case of shells,[4] or as a result of the boundary conditions, as in the case of oblique reflection of elastic waves.[5] If one limits consideration to situations in which displacement depends on a single spatial coordinate, as in the present study, two systems that are relevant are a ring[4] (transverse and axial displacements) and Timoshenko beam theory[2] (transverse displacement and the shear angle). In both cases, the variables are coupled by the field equations. This has the consequence that the wavenumbers for eigenfunctions are the same for both displacement components. The beam studied here is unlike those well studied systems in that the displacement components are not coupled by the field equations. The corollary of this feature is that the wavenumbers, and therefore the spatial patterns of the displacements constituting an eigenfunction, are dissimilar. Nevertheless, neither displacement can exist without the other. In this respect the vibration of a beam with a tilted roller resembles the way in which *P* and *S* waves are coupled by the boundary conditions when either is incident at a stress-free boundary in a half-space.[5]

The present analysis demonstrated that if the tilt angle for the roller is large, there are significant differences between the vibration of such a bar and that of a bar with a conventional roller support. A small tilt angle leads to a situation where the displacement in the direction of excitation (axial or transverse) shows little difference in comparison to the case where the roller is not tilted. The surprising aspect is that at certain resonances the undriven displacement component might exceed the driven one.

As an instructional exercise, the system considered here serves to introduce complexity in a readily accessible manner. The system has little direct relevance to engineering practice where elastic bars are seldom used as isolated elements. Nevertheless, the analysis extends the modal analysis methodology to systems where displacement components have dissimilar wavenumbers, and the results highlight the need to describe boundary conditions consistently with the manner in which a system actually is supported. This is especially the case when one seeks to correlate analytical models to experimental measurements, which is often the objective in experimental modal analysis.

[1]G. J. Simitses and D. H. Hodges, *Fundamentals of Structural Stability* (Elsevier, New York, 2006).
[2]J. H. Ginsberg, *Mechanical and Structural Vibration* (Wiley, New York, 2001).
[3]R. R. Craig, Jr., *Structural Dynamics* (Wiley, New York, 1981).
[4]W. Soedel, *Vibrations of Shells and Plates* (CRC, Boca Raton, FL, 2004).
[5]K. F. Graff, *Wave Motion in Elastic Solids* (Dover, Mineola, NY, 1991).

Two perspectives on equipartition in diffuse elastic fields in three dimensions

M. Perton and F. J. Sánchez-Sesma
Instituto de Ingeniería, Universidad Nacional Autónoma de México, Ciudad Universitaria, Coyoacán 04510, México D. F., Mexico

A. Rodríguez-Castellanos
Instituto Mexicano del Petróleo, Eje Central Lázaro Cárdenas 152, CP 07730, México Distrito Federal, Mexico

M. Campillo
LGIT, Observatoire de Grenoble, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex, France

R. L. Weaver
Department of Physics, University of Illinois, Urbana, Illinois 61801

(Received 21 October 2008; revised 19 May 2009; accepted 15 June 2009)

The elastodynamic Green function can be retrieved from the cross correlations of the motions of a diffuse field. To extract the *exact* Green function, perfect diffuseness of the illuminating field is required. However, the diffuseness of a field relies on the equipartition of energy, which is usually described in terms of the distribution of wave intensity in direction and polarization. In a full three dimensional (3D) elastic space, the transverse and longitudinal waves have energy densities in fixed proportions. On the other hand, there is an alternative point of view that associates equal energies with the independent modes of vibration. These two approaches are equivalent and describe at least two ways in which equipartition occurs. The authors gather theoretical results for diffuse elastic fields in a 3D full-space and extend them to the half-space problem. In that case, the energies undergo conspicuous fluctuations as a function of depth within about one Rayleigh wavelength. The authors derive diffuse energy densities from both approaches and find they are equal. The results derived here are benchmarks, where perfect diffuseness of the illuminating field was assumed. Some practical implications for the normalization of correlations for Green function retrieval arise and they have some bearing for medium imaging. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3177262]

PACS number(s): 43.40.Ga, 43.20.Bi, 43.20.Fn, 43.40.Fz [ADP] Pages: 1125–1130

I. INTRODUCTION

There is growing interest in Green function retrieval from cross correlations of motions within diffuse fields in various areas[1–4] including ultrasound,[5–7] ocean acoustics,[8,9] crustal seismology,[10–12] hazard monitoring,[13–15] earthquake engineering,[16–19] exploration seismology,[20–24] and even medical diagnostics.[25] However, the degree of diffuseness of a field cannot always be established[26] due to lack of isotropy of illumination or uneven distribution of scatterers. In seismological applications, a remarkable passive imaging capability is *de facto* present even in seismic noise wavefields,[10,11,27] The site imaging from the inversion of locally recorded *SH* wave microearthquake seismograms was accomplished in a pioneering early work.[28,29] In order to overcome the lack of coverage taking advantage of the heterogeneity of Earth's crust, an ingenious method has been proposed: the correlations of the coda of correlations (C^3).[30]

The authors propose an intuitive notion and associate the concept of diffuseness in the context of a radiating elastic field with the concept of energy equipartition. They attach the term "diffuse" to an equipartitioned radiating elastic field. The implied equivalence is useful, because measuring the degree of diffuseness of a particular radiation field is not as easily accomplished as is measuring the degree to which the energy is partitioned. In fact, by means of carefully designed experiments (in which dilatational and shear waves were identified), it was possible to recognize the presence of equipartition in the coda of several earthquakes recorded in Chilpancingo, Mexico[31] and at Pinyon Flats Observatory in California,[32] where theoretical equipartition in a layered medium was also discussed. Sometimes what can be measured or assessed are only the implications of these facts. For instance, the *exact* retrieval of Green function within an elastic medium requires both isotropic illumination of elastic plane waves and the proper partition of energy among the wave types.[33–35]

The equipartition of energy is usually described for elastic wave propagation in terms of wave polarizations and propagation velocities.[36] Equipartition implies that the available energy, in the phase space, is distributed in fixed proportions among the possible "states." There are at least two distinct and complementary descriptions of such states in an elastic diffuse wave field $\boldsymbol{u}(\boldsymbol{r},t)$.

One asserts that $\boldsymbol{u}$ may be represented as an incoherent superposition of incident plane waves of different polarizations. Theory predicts that each wave type has a share of the

available energy density: $E^S/E^P=2\alpha^3/\beta^3$, where E^P =energy density of longitudinal, or P, waves, and E^S =energy density of shear, or S, waves; and α and β are the longitudinal and transverse wave propagation velocities, respectively. This definition[36,37] is similar to the room acoustics notion of a diffuse field. It allows prediction of field correlations.

The other description establishes that the degrees of freedom of the system, the modes of vibration, are all incoherently excited with equal expected energy. This definition, introduced by Maxwell,[38] is also familiar from room acoustics.

In this communication it is demonstrated that there may be more than one way to partition the energy in diffuse elastic field. The equivalence between two possible ways is established for an infinite elastic space. Then, these results are connected with those obtained for the half-space's free surface[37] and the treatment is generalized for the half-space interior. Partition factors were obtained using these two points of view. Significant effects of the free surface are identified. As a test of consistency, the authors also compute the relationship of energy density with the imaginary part of Green function.

Within a diffuse elastic field, the components of energy density (each one associated with a given direction) are proportional to the correspondent components of the imaginary part of the Green function *at the source*.[33] In reality it can be interesting to see whether the normalized average autocorrelation stabilizes leading to the imaginary part of Green function. Both should match the energy density of a diffuse field. This fact can be applied to the imaging of the medium using backscattered fields.[39]

II. TWO POINTS OF VIEW ON EQUIPARTITION IN AN UNBOUNDED 3D SPACE

Within an unbounded isotropic elastic medium illuminated by a diffuse field, any orthogonal direction i of the three dimensional (3D) space has the same average energy density ζ_i being one-third of the total energy density ζ of bulk waves as discussed by Weaver,[37]

$$\zeta_i = \rho\omega^2 u_i^2(\boldsymbol{x}) = \zeta/3, \tag{1}$$

where the partition factor is one-third, ρ=mass density, ω =circular frequency, and $u_i(\boldsymbol{x})$=ith displacement component measured at $\boldsymbol{x}=(x_1,x_2,x_3)$. The authors' treatment is in frequency domain. In Eq. (1) the usual factor $\frac{1}{2}$ of the kinetic energy was dropped because the Virial theorem[37] indicates that the average energy density is twice the kinetic energy density.

On the other hand, by counting wave propagation modes, Weaver[36,37] established that the energy densities related to longitudinal and shear waves are, respectively, given by

$$\zeta^P = \zeta/(1+2R^3) \quad \text{and} \quad \zeta^S = \zeta 2R^3/(1+2R^3). \tag{2}$$

Here $R=\alpha/\beta$ is the velocity ratio of P and S waves. Within a diffuse field the energy carried by P waves is relatively small (for instance, for a Poisson solid in which $\alpha/\beta=\sqrt{3}$, the partition factor is 8.78%). The factor 2 inside the ζ^S expression accounts for the two polarizations of the shear waves, denoted here *SH* and *SV*, since the first one is taken with a polarization belonging only to the horizontal and the other to vertical planes. Indeed, the energy densities for these last ones are not distinguishable in unbounded space and satisfy $\zeta^{SH}=\zeta^{SV}$; therefore, $\zeta^S=\zeta^{SH}+\zeta^{SV}$. Without loss of generality, the orthogonal axes x_1 and x_2 are assumed to be horizontal. Since the *SH* wave is polarized in the horizontal plane its energy densities are

$$\zeta_1^{SH} = \zeta_2^{SH} = \zeta R^3(1+2R^3)^{-1}/2 \quad \text{and} \quad \zeta_3^{SH} = 0. \tag{3}$$

From the average of cross correlations of harmonic plane P, *SV*, and *SH* waves[34] the authors can establish the energy densities associated with *SV* and P waves. In fact, the densities ζ_i^{SV} and ζ_i^P, for direction i, are given by

$$\zeta_1^{SV} = \zeta_2^{SV} = \zeta R^3(1+2R^3)^{-1}/6,$$

$$\zeta_3^{SV} = \zeta 2R^3(1+2R^3)^{-1}/3,$$

and

$$\zeta_i^P = \zeta(1+2R^3)^{-1}/3 \quad (\text{for } i=1,2, \text{ or } 3). \tag{4}$$

This means that the energy densities in each direction can be expressed by

$$\zeta_i = \zeta_i^P + \zeta_i^{SV} + \zeta_i^{SH}. \tag{5}$$

The equipartition principle could either be seen from the points of view of the energy densities distributed in each direction or the partition determined for each kind of waves. Therefore the authors have

$$\zeta_1 + \zeta_2 + \zeta_3 = \zeta^P + \zeta^{SV} + \zeta^{SH} = \zeta, \tag{6}$$

as it should be.

III. TWO POINTS OF VIEW ON EQUIPARTITION IN A HALF-SPACE

When all body dimensions in the vicinity of the observation point are large compared to a wavelength all frequency scales are lost and there is no significant frequency dependence on the proportionalities. In fact, as it happens in the full-space case, at the free surface the partitions are frequency-independent. The corresponding partition factors have been computed from autocorrelations of an elastic diffuse field.[37]

In the presence of the free surface the energy partitions should be functions of frequency and depth. To calculate this the authors generalize Weaver's approach[37] and compute the autocorrelation of field components for the incoming elastic waves and their reflected consequences. This technique allows then to distinguish the partition factors among *incoming* wave types.

Equation (6) is extended here to the case of an elastic half-space by adding the surface energy density of the Rayleigh waves ζ^R. Now the energy densities are depth-dependent and are calculated from mean square displacements at point $\boldsymbol{x}$ as follows:

$$E_m(\boldsymbol{x}) = \sum_{\text{wave}} E_m^{\text{wave}}(\boldsymbol{x})$$

$$= \rho\omega^2 \sum_{\text{wave}} \langle u_m^{\text{wave}}(\boldsymbol{x}) u_m^{\text{wave}^*}(\boldsymbol{x})\rangle \quad \text{no sum over } m. \tag{7}$$

In this equation $E_m(\boldsymbol{x})=m$th component of the total energy density at point $\boldsymbol{x}$, the index *wave* represents all the incident waves, that is to say, *P*, *SV*, *SH*, and Rayleigh, the last one being denoted by *R*. Moreover, $E_m^{\text{wave}}(\boldsymbol{x})=$energy density at point $\boldsymbol{x}$ associated with the incoming wave and its reflected consequences and to the field component m as well. The symbol * represents the complex conjugate, and the field $u_m^{\text{wave}}(\boldsymbol{x})=$displacement in the m direction due to a given incident harmonic plane wave and its reflections, if any.

The authors assumed the incident wave fields to be isotropically distributed with unit amplitude. The angular brackets $\langle\ \rangle$ below account for an average over all incoming directions from the half-space. It is defined by means of $\langle\bullet\rangle = (2\pi)^{-1}\int_0^{2\pi}\int_0^{\pi/2}\sin\theta d\theta d\varphi$. For example, for the incident *P* waves (denoted by P_i), the authors can write, for x_1 or x_2 directions

$$E_j^P = \rho\omega^2\langle u_j^P, u_j^{P^*}\rangle = \frac{\zeta^P/2}{2\pi}\int_0^{2\pi}\left(\int_0^{\pi/2}|(u_j^{Pi}+u_j^r)|^2\sin\theta d\theta\right)d\varphi$$

$$= \frac{1}{2}\frac{\zeta^P}{2k_P}\int_0^{k_P}\left(|u_h^{Pi}+u_h^r|^2\frac{1}{\gamma}\right)k_h dk_h, \tag{8}$$

where $k_P=\omega/\alpha$ is the *P* wave number. In Eq. (8) no summation is implied over the j index; therefore, k_h is the projection of any wave vector $\boldsymbol{k}$ over an arbitrary horizontal component $\boldsymbol{e}_h$. Here, j can be 1 or 2 without any distinction and u_h verifies $u_j=\cos(\varphi-\varphi_j^0)u_h$, where $\varphi_j^0=(\boldsymbol{e}_j,\boldsymbol{e}_h)$. The superindex r states the reflected consequences due to the incident waves. Moreover, $\gamma=\sqrt{k_P^2-k_1^2-k_2^2}$ is the vertical (x_3 direction) *P* wave number. This has to fulfill the condition Im $\gamma\leq 0$. Since the horizontal directions are not distinguishable, Eq. (8) shows that the 3D isotropic case is very close to the 2D case.[33] The incident and reflected fields of the bulk displacement waves are then developed by using the unit polarization vector of the incident field: $\boldsymbol{n_P}$, of the *P* waves the corresponding reflection coefficients R_{PP} and R_{PSV} and polarization vectors below the plane $(0,\boldsymbol{e_h},\boldsymbol{e}_3)$. Since the medium is isotropic, these polarization vector and reflection coefficients are merely expressed from the corresponding wave vector. Equation (8) can then be written as

$$E_j^P(\boldsymbol{x},\omega) = \frac{1}{2}\frac{\zeta^P}{2k_P}\int_0^{k_P}|n_h^P e^{\iota\gamma x_3} + n_h^P R_{PP}e^{-\iota\gamma x_3}$$

$$- n_h^{SV}R_{PSV}e^{-\iota\nu x_3}|^2\frac{k_h}{\gamma}dk_h, \quad j=1,2$$

$$= \frac{1}{2}\frac{\zeta^P}{2k_P}\int_0^{k_P}\left|\frac{1}{k_P}(k_h + k_h r_0 e^{-2\iota\gamma x_3}\right.$$

$$\left. - \nu r_1\gamma e^{-\iota(\nu+\gamma)x_3}\right|^2\frac{k_h}{\gamma}dk_h, \tag{9}$$

where ι is the imaginary unit, $r_0=(4k_h^2\gamma\nu-(k_h^2-\nu^2)^2)/F(k_h)$, $r_1=(k_h^2-\nu^2)4k_h/F(k_h)$, and $F(k_h)=4k_h^2\gamma\nu+(k_h^2-\nu^2)^2$ is the Rayleigh function. The same approach is used for the following equations:

$$E_j^{SV}(\boldsymbol{x},\omega) = \frac{1}{2}\frac{\zeta^{SV}}{2k_{SV}}\int_0^{k_{SV}}\left|\frac{1}{k_{SV}}(\nu - \nu r_0 e^{-2\iota\nu x_3}\right.$$

$$\left. - k_h r_1 \nu e^{-\iota(\nu+\gamma)x_3}\right|^2\frac{k_h}{\nu}dk_h, \quad j=1,2, \tag{10}$$

$$E_j^{SH}(\boldsymbol{x},\omega) = \frac{1}{2}\frac{\zeta^{SH}}{2k_{SH}}\int_0^{k_{SH}}|1+e^{-2\iota\nu x_3}|^2\frac{k_h}{\nu}dk_h, \quad j=1,2, \tag{11}$$

$$E_3^P(\boldsymbol{x},\omega) = \frac{\zeta^P}{2k_P}\int_0^{k_P}\left|\frac{1}{k_P}(\gamma - \gamma r_0 e^{-2\iota\gamma x_3}\right.$$

$$\left. - k_h r_1 \gamma e^{-\iota(\nu+\gamma)x_3}\right|^2\frac{k_h}{\gamma}dk_h, \tag{12}$$

$$E_3^{SV}(\boldsymbol{x},\omega) = \frac{\zeta^{SV}}{2k_{SV}}\int_0^{k_{SV}}\left|\frac{1}{k_{SV}}(-k_h - k_h r_0 e^{-2\iota\nu x_3}\right.$$

$$\left. + \gamma r_1 \nu e^{-\iota(\nu+\gamma)x_3}\right|^2\frac{k_h}{\nu}dk_h, \tag{13}$$

where $k_{SV}=k_{SH}=\omega/\beta=S$ wave numbers for *SV* and *SH* waves, respectively. Moreover, $\nu=\sqrt{k_{SV}^2-k_1^2-k_2^2}$ is the vertical (x_3 direction) *S* wave number. This has to fulfill the conditions Im $\nu\leq 0$. The incident and reflected fields of the bulk displacement waves are then developed by using the unit polarization vectors of the incident fields: $\boldsymbol{n}_{SV}$ and $\boldsymbol{n}_{SH}$ of the *SV* and *SH* waves, respectively, and the reflection coefficients R_{SVP}, R_{SVSV}, and R_{SHSH}.

Reference 37 gives the value of the Rayleigh energy density per area of the free surface as

$$\zeta^R = \zeta\frac{\pi}{\omega}\frac{\alpha^3}{c_R^2}(1+2R^3)^{-1}, \tag{14}$$

where c_R is the Rayleigh wave speed. The motion of the Rayleigh wave is taken from 40 and the energy densities per volume of the Rayleigh wave components are

$$E_j^R(\boldsymbol{x},\omega) = \tfrac{1}{2}\zeta^R A(e^{-k_R\chi_1 x_3} - \sqrt{\chi_1\chi_2}e^{-k_R\chi_2 x_3})^2, \quad j=1,2, \tag{15}$$

$$E_3^R(\boldsymbol{x},\omega) = \zeta^R A\chi_1/\chi_2(e^{-k_R\chi_2 x_3} - \sqrt{\chi_1\chi_2}e^{-k_R\chi_1 x_3})^2, \tag{16}$$

where $A=1/\int_0^\infty(e^{-k_R\chi_1 x_3}-\sqrt{\chi_1\chi_2}e^{-k_R\chi_2 x_3})^2+\chi_1/\chi_2(e^{-k_R\chi_2 x_3}-\sqrt{\chi_1\chi_2}e^{-k_R\chi_1 x_3})^2 dx_3$, $k_R=\omega/c_R$, $\chi_1=\sqrt{1-c_R^2/\alpha^2}$, and $\chi_2=\sqrt{1-c_R^2/\beta^2}$. The constant A, which has the dimension of inverse of length, assures that the total energy density injected is the given amount. The factor $\frac{1}{2}$ in Eqs. (8)–(11) and (15) comes from the azimuthal integration of the horizontal components.

Now that all terms $E_m^{\text{wave}}(\boldsymbol{x})$ have been written explicitly, the equipartition theorem can be expressed as the sum of partial contributions of either the Cartesian components of motion or the different types of waves

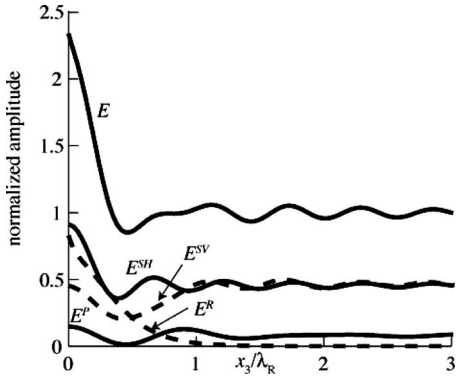


FIG. 1. Energy densities for incoming *P*, *SV*, *SH*, and *R* waves against normalized depth x_3/λ_R using Eqs. (17) and (18). The cases of SV and Rayleigh waves are drawn with discontinuous lines.

$$E = E_1 + E_2 + E_3 = E^P + E^{SV} + E^{SH} + E^R, \tag{17}$$

where the total energy in each direction i can be evaluated by means of

$$E_m(\boldsymbol{x}) = \sum_{\text{wave}} E_m^{\text{wave}}(\boldsymbol{x}). \tag{18}$$

Equations (17) and (18) are independent of Poisson ratio but to illustrate graphically these results the energy densities were evaluated for a Poisson solid (i.e., the Lamé constants verify $\lambda=\mu$ or, equivalently, $\nu=0.25$ and thus $\alpha/\beta=\sqrt{3}$) and depicted in Fig. 1, which shows energy densities for each incident wave type and their sum against normalized depth x_3/λ_R, where λ_R=Rayleigh wavelength. In this figure it can be seen that in a diffuse field the *SH* and Rayleigh waves take the largest share whereas *P* and *SV* waves even lose some energy near the boundary. The effect of the free surface tends to disappear at a depth equal to the Rayleigh wavelength.

Similarly, energy densities can be associated with the three orthogonal directions, and their sums are displayed against normalized depth in Fig. 2. Again, for illustration purposes, a Poisson solid is assumed (e.g., $\lambda=\mu$ or, equivalently, $\nu=0.25$). All the curves are normalized by the reference energy (ζ) in deep space. Besides, since the energy is conserved, the averaged integral of the sum of the energies over depth is equal to $\zeta+E^R$. Near the surface, the energy balance is strongly dependent on the vertical coordinate and presents a peak at the free surface, reaching a value of about 2.3 times the energy density in the deep space.

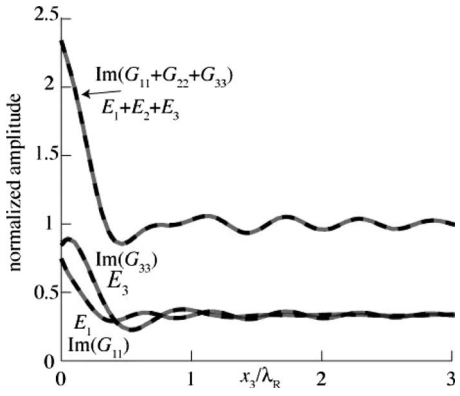


FIG. 2. Energy densities for the three orthogonal directions (black discontinuous lines) against normalized depth x_3/λ_R using Eqs. (17) and (18). Amplitudes are normalized in terms of the energy density in deep space. The gray continuous lines come from the imaginary part of the Green function components *at the source*, which are computed from Eqs. (22) and (23) given in Sec. IV.

In Fig. 2 the E_2 curve is equal to E_1 curve. At the free surface the values for E_1 and E_3 are very close. While E_1 and E_2 have their maximum at the free surface, E_3 have a peak just below at a depth of about $\lambda_R/6$. This result exhibits that at the free surface of the Poissonian half-space, the horizontal to vertical energy ratio is $E_H/E_V=(E_1+E_2)/E_3\approx 1.76$.

It is possible, from Eqs. (9)–(16), to finely study the energy share among the wave types and the components. Although this is beyond the scope of this work, the obtained results agree with the partitions at the free surface obtained by Weaver[37] and generalize them for depth dependence. Moreover, the studied relationships have close connections with experimental and theoretical work.[31–33]

IV. APPLICATION TO MEDIUM IMAGING

The energy density can be obtained as the sum of the average autocorrelations of diffuse field motions and is proportional to the imaginary part of the trace of the Green function at the source.[33] In particular, for the isotropic 3D case, the authors have the energy density in the direction m given by

$$E_m(\boldsymbol{x}) = \sum_{\text{wave}} E_m^{\text{wave}}(\boldsymbol{x}) = -\rho\omega^2 E^S k_S^{-1} 2\pi\mu\ \text{Im}(G_{mm}(\boldsymbol{x},\boldsymbol{x})). \tag{19}$$

In this equation μ=shear modulus and $\text{Im}(G_{mm}(\boldsymbol{x},\boldsymbol{x}))$ =imaginary part of component (m,m) of the Green tensor (no sum is here invoked, despite the repetition of subscripts) for source and receiver being at the same point. The singularity of Green function is restricted to its real part.

In what follows the calculation of the elastodynamic Green function at the source will be sketched. The half-space Green function $G_{ij}(\boldsymbol{x},\boldsymbol{\xi},\omega)$, which represents the displacement in the ith direction at point $\boldsymbol{x}$ generated by a harmonic unit force oriented in the jth direction at the point $\boldsymbol{\xi}$, is calculated here from the expression of the Green function $G_{ij}^0(\boldsymbol{x},\boldsymbol{\xi},\omega)$ of the infinite space. The presence of an infinite plane surface is taken into account by adding the reflected field, which is denoted as $G_{ij}^r(\boldsymbol{x},\boldsymbol{\xi},\omega)$ as follows:

$$G_{ij}(\boldsymbol{x},\boldsymbol{\xi},\omega) = G_{ij}^0(\boldsymbol{x},\boldsymbol{\xi},\omega) + G_{ij}^r(\boldsymbol{x},\boldsymbol{\xi},\omega). \tag{20}$$

The Green function $G_{ij}^0(\boldsymbol{x},\boldsymbol{\xi},\omega)$ is the Stokes solution[41] and is given by

$$G_{ij}^0(\boldsymbol{x},\boldsymbol{\xi},\omega)\mu = G(k_S r)\delta_{ij} + \frac{1}{k_S^2}\frac{\partial^2}{\partial x_i\,\partial x_j}(G(k_S r) - G(k_P r)), \tag{21}$$

where $r=|\boldsymbol{x}-\boldsymbol{\xi}|$ and $G(kr)=e^{-\imath kr}/(4\pi r)$. This last expression is then expressed in functions of k_1 and k_2, the horizontal components of wave vectors by using the Weyl integrals[42]

$$G(k_s r) = \frac{\iota}{8\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{e^{-\iota(k_1x_1+k_2x_2+\nu|x_3|)}}{\nu}dk_1dk_2 \quad \text{and}$$

$$G(k_P r) = \frac{\iota}{8\pi^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{e^{-\iota(k_1x_1+k_2x_2+\gamma|x_3|)}}{\gamma}dk_1dk_2. \quad (22)$$

The reflected field is constructed by the superposition of homogeneous and inhomogeneous plane P, SV, and SH waves

$$G^r_{ij}(\boldsymbol{x},\boldsymbol{\xi},\omega) = (A_P\boldsymbol{n}_P e^{-\iota\boldsymbol{k}_P\boldsymbol{x}} + A_{SV}\boldsymbol{n}_{SV}e^{-\iota\boldsymbol{k}_{SV}\boldsymbol{x}} + A_{SH}\boldsymbol{n}_{SH}e^{-\iota\boldsymbol{k}_{SH}\boldsymbol{x}})e^{\iota\omega t}. \quad (23)$$

The coefficients A_P, A_{SV}, and A_{SH} are determined from a linear system by enforcing the null normal stresses at the free surface. The Green functions are then computed by making a double adaptive Simpson quadrature integration over the axis k_1 and the contour c_2. This latter is the axis $k_2+\iota\varepsilon$ with $\varepsilon = 1/|k_2-p_0+\varepsilon_0|$ taken proportional to the distance of the Rayleigh pole p_0. The parameter ε_0 is used to adjust the minimal distance to the pole. The calculations were performed with $\varepsilon_0=0.01$.

Results are superposed in Fig. 2 with gray continuous lines to the energies E_1, E_2, and E_3 and their sum, which coincides with their corresponding components of the imaginary part of Green tensor at the source itself and their sum. This result verifies the equivalence between the two approaches to deal with energy density partitions in a half-space generalizing the result for a full-space. Moreover, the use of the imaginary part of Green function at the source serves as a test of concept, a quality control.

On the other hand, the relationship between the autocorrelation of the field fluctuations recorded by a single sensor (energy density) and the imaginary part of the Green's function can be used to infer backscattered waves, which can be used to image the medium.[39]

V. DISCUSSION AND CONCLUSIONS

The authors demonstrated that there may be at least two ways to understand equipartition in a diffuse elastic field, and showed their equivalence. They related this with results from previous research, restricted to the free surface[37] and to the new developments regarding Green function retrieval from field fluctuations.[1–4] Then, they extended the analysis inside the half-space checking the consistency of results. As in the full-space, the sum of energy densities is the same whether the wave types or the degrees of freedom are considered. The conservation of energy implicit in the authors' results gives us a glimpse of the two faces of equipartition.

The authors identified the conspicuous fluctuations of energy densities close to the surface due to the emergence of surface waves. These effects prevail only within a depth of about one Rayleigh wavelength. Normalized energy densities can be seen as partition coefficients.

Partition coefficients versus normalized depth here presented are *analytical benchmarks* in which perfect equipartition was assumed for the illumination. The authors' results provide tests for synthetic normalization schemes aimed to deal with real data.

The identity of the energy density and the imaginary part of Green function at the source allowed independent test of results. This relationship can be useful for medium imaging.

ACKNOWLEDGMENTS

Thanks are given to P. Gouédard, A. Pierce, G. Prieto, M. Rodríguez-González, O. Sánchez, R. Snieder, J. H. Spurlin, and K. Wapenaar for their constructive remarks. The comments from an anonymous reviewer were crucial to improve this work. The authors thank G. Sánchez and her team of Unidad de Servicios de Información (USI) of Instituto de Ingeniería, UNAM, for their help locating useful references. Support from DGAPA-UNAM, Project Nos. IN114706 and IN121709, Mexico; from project DyETI of INSU-CNRS, France; and from the Instituto Mexicano del Petróleo is greatly appreciated.

[1] R. L. Weaver, "Information from seismic noise," Science **307**, 1568–1569 (2005).
[2] E. Larose, L. Margerin, A. Derode, B. van Tiggelen, M. Campillo, N. Shapiro, A. Paul, L. Stehly, and M. Tanter, "Correlation of random wavefields: An interdisciplinary review," Geophysics **71**, SI11–SI21 (2006).
[3] A. Curtis, P. Gerstoft, H. Sato, R. Snieder, and K. Wapenaar, "Seismic interferometry—Turning noise into signal," The Leading Edge **25**, 1082–1092 (2006).
[4] P. Gouédard, L. Stehly, F. Brenguier, M. Campillo, Y. Colin de Verdiére, E. Larose, L. Margerin, P. Roux, F. J. Sánchez-Sesma, N. M. Shapiro, and R. L. Weaver, "Cross-correlation of random fields: Mathematical approach and applications," Geophys. Prospect. **56**, 375–393 (2008).
[5] R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[6] A. Malcolm, J. Scales, and B. A. van Tiggelen, "Extracting the Green's function from diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601 (2004).
[7] K. van Wijk, "On estimating the impulse response between receivers in a controlled ultrasonic experiment," Geophysics **71**, SI79–SI84 (2006).
[8] P. Roux, W. A. Kuperman, and NPAL Group, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[9] K. G. Sabra, P. Roux, A. M. Thode, G. L. D'Spain, and W. S. Hodgkiss, "Using ocean ambient noise for array self-localization and self-synchronization," IEEE J. Ocean. Eng. **30**, 338–347 (2005).
[10] N. M. Shapiro, M. Campillo, L. Stehly, and M. H. Ritzwoller, "High-resolution surface-wave tomography from ambient seismic noise," Science **307**, 1615–1618 (2005).
[11] K. G. Sabra, P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, "Surface wave tomography from microseisms in Southern California," Geophys. Res. Lett. **32**, L14311 (2005).
[12] S. Ma, G. A. Prieto, and G. C. Beroza, "Testing community velocity models for Southern California using the ambient seismic field," Bull. Seismol. Soc. Am. **98**, 2694–2714 (2008).
[13] C. Sens-Schönfelder and U. Wegler, "Passive image interferometry and seasonal variations at Merapi volcano, Indonesia," Geophys. Res. Lett. **33**, L21302 (2006).
[14] K. G. Sabra, P. Roux, P. Gerstoft, W. A. Kuperman, and M. C. Fehler, "Extracting coherent coda arrivals from cross-correlations of long period seismic waves during the Mount St. Helens 2004 eruption," Geophys. Res. Lett. **33**, L06313 (2006).
[15] U. Wegler and C. Sens-Schönfelder, "Fault zone monitoring with passive image interferometry," Geophys. J. Int. **168**, 1029–1033 (2007).
[16] R. Snieder and E. Safak, "Extracting the building response using seismic interferometry; theory and application to the Millikan library in Pasadena, California," Bull. Seismol. Soc. Am. **96**, 586–598 (2006).
[17] D. Thompson and R. Snieder, "Seismic anisotropy of a building," The Leading Edge **25**, 1093 (2006).
[18] M. D. Kohler, T. H. Heaton, and S. C. Bradford, "Propagating waves in the steel, moment-frame factor building recorded during earthquakes," Bull. Seismol. Soc. Am. **97**, 1334–1345 (2007).

[19]K. G. Sabra, A. Srivastava, F. L. di Scalea, I. Bartoli, P. Rizzo, and S. Conti, "Structural health monitoring by extraction of coherent guided waves from diffuse fields," J. Acoust. Soc. Am. **123**, EL8–EL13 (2008).
[20]K. Aki, "Space and time spectra of stationary stochastic waves with special reference to microtremors," Bull. Earthquake Res. Inst., Univ. Tokyo **35**, 415–456 (1957).
[21]J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269 (1968).
[22]J. N. Louie, "Faster, better: Shear-wave velocity to 100 meters depth from refraction microtremor analysis," Bull. Seismol. Soc. Am. **91**, 347–364 (2001).
[23]D. Halliday, A. Curtis, and E. Kragh, "Seismic surface waves in a suburban envirоment: Active and passive interferometric methods," The Leading Edge **27**, 210–218 (2008).
[24]M. Miyazawa, R. Snieder, and A. Venkataraman, "Application of seismic interferometry to extract P and S wave propagation and observation of shear wave splitting from noise data at Cold Lake, Canada," Geophysics **73**, D35–D40 (2008).
[25]K. G. Sabra, S. Conti, P. Roux, and W. A. Kuperman, "Passive in-vivo elastography from skeletal muscle noise," Appl. Phys. Lett. **90**, 194101 (2007).
[26]F. Mulargia and S. Castelaro, "Passive imaging in nondiffuse acoustic wavefields," Phys. Rev. Lett. **100**, 218501 (2008).
[27]N. M. Shapiro and M. Campillo, "Emergence of broad band Rayleigh waves from correlations of the ambient seismic noise," Geophys. Res. Lett. **31**, L07614 (2004).
[28]F. Scherbaum, "Seismic imaging of the site response using microearthquake recordings. Part I. Method," Bull. Seismol. Soc. Am. **77**, 1905–1923 (1987).
[29]F. Scherbaum, "Seismic imaging of the site response using microearthquake recordings. Part II. Application to the Swabian Jura, Southest Germany, seismic network," Bull. Seismol. Soc. Am. **77**, 1924–1944 (1987).
[30]L. Stehly, M. Campillo, B. Froment, and R. L. Weaver, "Reconstructing Green's function by correlation of the coda of the correlation (C3) of ambient seismic noise," J. Geophys. Res. **113**, B11306 (2008).
[31]R. Hennino, N. Trégourès, N. M. Shapiro, L. Margerin, M. Campillo, B. A. van Tiggelen, and R. L. Weaver, "Observation of equipartition of seismic waves," Phys. Rev. Lett. **86**, 3447–3450 (2001).
[32]L. Margerin, M. Campillo, B. A. van Tiggelen, and R. Hennino, "Energy partition of seismic coda waves in layered media: Theory and application to Pinyon Flats Observatory," Geophys. J. Int. **177**, 571–585 (2009).
[33]F. J. Sánchez-Sesma, J. A. Pérez-Ruiz, F. Luzón, M. Campillo, and A. Rodríguez-Castellanos, "Diffuse fields in dynamic elasticity," Wave Motion **45**, 641–654 (2008).
[34]F. J. Sánchez-Sesma and M. Campillo, "Retrieval of the Green's function from cross correlation: The canonical elastic problem," Bull. Seismol. Soc. Am. **96**, 1182–1191 (2006).
[35]F. J. Sánchez-Sesma, J. A. Pérez-Ruiz, M. Campillo, and F. Luzón, "Elastodynamic 2D Green function retrieval from cross-correlation: Canonical inclusion problem," Geophys. Res. Lett. **33**, L13305 (2006).
[36]R. L. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[37]R. L. Weaver, "Diffuse elastic waves at a free surface," J. Acoust. Soc. Am. **78**, 131–136 (1985).
[38]A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, translated by G. Gamow (Dover, New York, 1949), pp. 93–98.
[39]R. Snieder, F. J. Sánchez-Sesma, and K. Wapenaar, "Field fluctuations, imaging with backscattered waves, a generalized energy theorem, and the optical theorem," SIAM J. Imaging Sci. **2**, 763–776 (2009).
[40]D. Royer and E. Dieulesaint, *Ondes Élastiques Dans les Solides (Elastic Waves in Solids)* (Masson, Paris, 1996).
[41]J. D. Achenbach and Y. Xu, "Wave motion in an isotropic elastic layer generated by a time-harmonic point load of arbitrary direction," J. Acoust. Soc. Am. **106**, 83–90 (1999).
[42]K. Aki and P. G. Richards, *Quantitative Seismology*, 2nd ed. (University Science Book, Sausalito, CA, 2002).

A procedure for the assessment of low frequency noise complaints[a)]

Andy T. Moorhouse, David C. Waddington,[b)] and Mags D. Adams
Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom

(Received 2 August 2007; revised 30 March 2009; accepted 25 June 2009)

The development and application of a procedure for the assessment of low frequency noise (LFN) complaints are described. The development of the assessment method included laboratory tests addressing low frequency hearing threshold and the effect on acceptability of fluctuation, and field measurements complemented with interview-based questionnaires. Environmental health departments then conducted a series of six trials with genuine "live" LFN complaints to test the workability and usefulness of the procedure. The procedure includes guidance notes and a pro-forma report with step-by-step instructions. It does not provide a prescriptive indicator of nuisance but rather gives a systematic procedure to help environmental health practitioners to form their own opinion. Examples of field measurements and application of the procedure are presented. The procedure and examples are likely to be of particular interest to environmental health practitioners involved in the assessment of LFN complaints.
 [DOI: 10.1121/1.3180695]

PACS number(s): 43.50.Ba, 43.50.Rq, 43.66.Lj [BSF] Pages: 1131–1141

I. INTRODUCTION

Many environmental health practitioners will be familiar with complaints about low frequency noise (LFN) in the range 20–160 Hz. The vocabulary used by complainants to describe the noise they experience is highly consistent, and invariably they describe a noise that is intense, even deafening to them while at the same time visitors to their home may hear nothing. This discrepancy between how the sufferer perceives the sound and how it is experienced by others is one of the most perplexing aspects of LFN, and can leave the sufferer feeling increasingly isolated and confused. LFN is now a recognized problem in many countries in the world, as detailed in the review by Leventhall *et al.*[1] This does not mean that the causes of such suffering are fully understood and many cases still go unexplained. Further, these cases usually take up disproportionately more time and effort than other noise complaints. This adds to the stress on the LFN sufferers and the officers investigating the complaints.

Fundamental to the problem of the assessment of LFN complaints is the question of how it may be that one person can describe a sound as loud that few others can even hear. One possible explanation that may explain some, but by no means all cases, is based on the physiology of the human hearing system for low frequencies. The perceived loudness of low frequency sounds increases rapidly with increasing acoustic energy, and so low frequency sounds just above the threshold of hearing can be perceived as loud, even uncomfortably loud. Furthermore, individual hearing thresholds vary such that people with more sensitive hearing can hear low frequency sounds that are inaudible to others.

This situation does not often arise with higher frequency sounds because their perceived loudness increases much more slowly with increased acoustic energy. A compounding factor is that "sensitization" to low frequency sound may occur over time, leaving the sufferer more aware of the sound and unable to shut it out or get used to it.[2] This means that a short visit to a property affected by LFN does not always give an adequate impression of what it is like to actually live with the sound, making evaluation even more difficult. An appreciation of these subtleties is important, because the counterintuitive nature of low frequency sound makes it difficult to base accurate judgments on personal experience.

This paper summarizes work performed recently by Moorhouse *et al.*[3–5] to develop a procedure for the assessment of a LFN complaint produced as part of a Defra-funded project in the United Kingdom. It was not the intention of this work to provide guidance in locating the source of a LFN. Rather, the procedure aims to help environmental health practitioners to distinguish cases where an environmental sound is responsible for a disturbance, in which case they may be able to take some action, from those where no such action is possible. However, it is usually found that the most difficult part of an assessment is in determining the existence or otherwise of a sound that correlates with the disturbance, and if this can be established then the source can usually be found.

The structure of the paper is as follows. Section II describes the development of the procedure, involving field measurements complemented by interview-based questionnaires, and laboratory measurements comprising audiometric and subjective tests. The procedure itself is outlined in Sec. III, presenting both a criterion curve and the assessment

[a)]Aspects of this work were presented at the 12th International Meeting on Low Frequency Noise and Vibration and Its Control, Ramada Plaza Hotel, Bristol, United Kingdom, 18–20 September 2006.
[b)]Author to whom correspondence should be addressed. Electronic mail: d.c.waddington@salford.ac.uk

method. Field trials of the procedure by the environmental health practitioners are described in Sec. IV, together with a summary of their feedback regarding practical application of the method and responses of the complainants. Limitations of the procedure and implications for environmental health practitioners and audiologists are discussed in Sec. V, before the conclusions in Sec. VI.

II. DEVELOPMENT OF THE PROCEDURE

A complementary set of field and laboratory studies was conducted in order to establish the best form for an assessment method. In the field studies, 11 cases of reported LFN were investigated, as well as 5 control cases where no complaints had been received. In addition to making physical recordings of the sounds within complainants' residences, it was necessary to obtain a significant amount of personal data about the individuals using a comprehensive one-to-one semi-structured interview schedule. In the laboratory tests, a set of "thresholds of acceptability" was established by asking 18 subjects to set the level of various low frequency sounds to a just-acceptable level for imagined day and night situations. The sounds presented consisted of a set of tones across the low frequency range, "real" LFN extracted from field test recordings, and synthesized tones with varying degrees of fluctuation. The findings from these field and laboratory studies are summarized below.

A. Field measurements

1. Participant recruitment

LFN sufferers were identified with the help of Environmental Health Departments in areas that had ongoing complaint cases. Having circulated letters to local government authorities throughout the United Kingdom over 40 possible cases were identified and evaluated. Through telephone discussion with the environmental health practitioners in question, a detailed description of each case was obtained to aid selection of suitable sites. Cases where several complaints occurred in a cluster were selected in preference over those that were isolated complaints. Environmental health practitioners and sufferers alike were generally keen to participate. Cases were selected before any acoustical measurements were made, and no cases were discarded after recordings were performed.

2. Field measurement methodology

The main objective of the field measurements was to provide a database of field data for the development of a proposed criterion. Specifically this involved collecting data with which to test proposed criteria, and to provide audio recordings for use in the laboratory tests. Although the majority of environmental noise standards specify that sound measurements should be conducted outside, it is now generally agreed that LFN can only meaningfully be evaluated inside dwellings.[6] In this series of investigations, a single microphone was positioned at a point in the room where the complainant indicated that the sound was present. An unoccupied room was used for preference to avoid disturbance of the measurements by the household, and recordings were usually made between 2100 and 0900 h. Subjects performed a one-on-one interview with an experienced interviewer, detailed below in Sec. II B 1, and were asked to complete a log sheet giving comments on how they perceived the sound at particular times. The equipment was left to monitor unmanned for between 3 and 5 days. The microphone and measurement chain were calibrated down to 1 Hz against a traceable standard at the UKAS accredited Calibration Laboratory at the University of Salford immediately prior to the tests. Parameters recorded included 1/3 octave spectra and audio. Data were streamed directly to hard disk.

3. Analysis of the field measurements

Large amounts of data were collected and details of the analyses are presented in the project report.[2] Most of the problem and marginal cases were in the 40 and 50 Hz third octave spectrum bands. In all cases, the background noise levels in the residences were remarkably low. Such low levels of natural masking noise are thought to be a factor contributing to the disturbance of LFN.[1] Audio recordings were analyzed to detect tonal components, temporal structures, and modulations that cannot be adequately detected from 1/3 octave sound pressure levels alone and were played back at a higher level to help to distinguish between various noise sources. Combined with third octave and narrow band spectra, this provided the most successful identification of sources.

During field trials, there were no cases in which the LFN was reported to be present only during the day. This does not mean that the noise was absent during the day though, since most respondents said that while sound could be heard during the day, it was worst at night. Furthermore, in every case, the noise was reported to be present at night. This contrasts with consultancy experience where a random selection of general industrial noise complaints might be expected to include some complaints about industry that does not operate at night that causes disturbance in the daytime. While this observation does not contribute to the main method of the assessment of LFN complaints, a combination of very low background noise levels and intermittent interaction tones from domestic equipment such as refrigerators, with occasional transportation noise and room resonance modes, ought not to be overlooked when analyzing interview and measurement data.

4. Case where an environmental sound was positively identified

This case study took place in an apartment in a quiet urban area. Figure 1 shows a 1/3 octave band spectrum calculated over one of many periods identified by the complainant. Compared with the nighttime criterion curve from the procedure for the assessment of LFN complaints, it is seen that the 63 Hz 1/3 octave band predominates. Figure 2 shows a time history of the measurements in the 63 Hz 1/3 octave spectrum band, and it is evident that a source cycles on and off with periodicity of about 10 min on and 20 min off. Also shown in the time history are the 63 Hz daytime and nighttime criteria from the procedure. While the background level

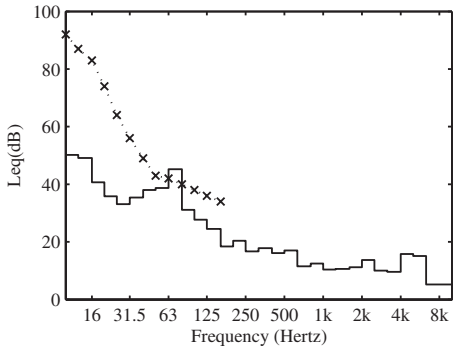


FIG. 1. Case where an environmental sound was positively identified. Solid line shows measured 1/3 octave band spectrum averaged over 9 min and 30 s starting 0700 h. Dashed (x) line is the nighttime criterion from the procedure for the assessment of LFN complaints.

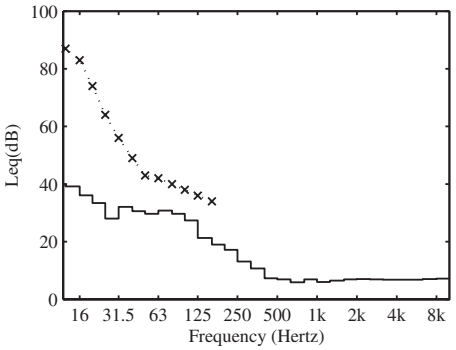


FIG. 3. Case where no environmental sound was identified. Solid line shows measured 1/3 octave band spectrum averaged over 9 min and 30 s starting 1930 h. Dashed (x) line is the nighttime criterion from the procedure for the assessment of LFN complaints.

during the nighttime is well below the criterion, the source levels clearly exceed the criterion. Given the correlation of the complainant's log with these recordings these results indicate that this source is likely to be the cause of the complaints.

5. Case in which no environmental sound was identified

One example of this category of case study took place in a house in a quiet urban area. Comparing the spectrum for one of the many periods identified by the complainant with the criterion curve in Fig. 3, it is seen that no particular 1/3 octave band dominates. The 63–100 Hz bands may just be audible, but the dominant source in this part of the spectrum was found to be road traffic. This was found to be quite common in the cases and control cases in this study. Figure 4 shows a time profile of the measurements in the 80 Hz 1/3 octave spectrum band. The profile of the sound levels during the night is again typical of road traffic. Occasional spikes on this plot are due to domestic movement or traffic events and are not associated with any steady low frequency sound. Also shown in the time history are the 80 Hz daytime and nighttime criteria from the procedure for the assessment of LFN complaints.

While the background level during the nighttime is well below the criterion, daytime levels are also seen to be remarkably low. More detailed frequency analyses were also performed, and several other times were evaluated. However, no relationship between noise levels and the complainant's log could be established. Given the exceptionally low levels as compared with the criteria and the lack of correlation between the complainant's log with these recordings, these results indicate that no environmental source was measured that is likely to be the cause of the complaints.

6. Categorization of case studies

The data from the field studies were combined with results from the laboratory tests to produce a criterion curve to assist environmental health practitioners in their assessment of LFN complaints. Details of the laboratory tests are presented below in Sec. II C, while the criterion curve is detailed in Sec. III A.

Examining the 11 cases where LFN was reported, three cases were identified where the criteria were exceeded and where there was also correlation between the residents' logged complaints and the LFN level. Two of these three cases were related, having been measured in the same apartment block. Five cases were identified where the criteria were not generally exceeded and where there was also a lack of correlation between comments and noise levels. Analysis of these eight cases using the procedure for the assessment of LFN complaints was straightforward. The remaining three cases were marginal in that the measured LFN was close to the criterion in level, and moreover, did not correlate with complainant comments. Investigation of these marginal cases

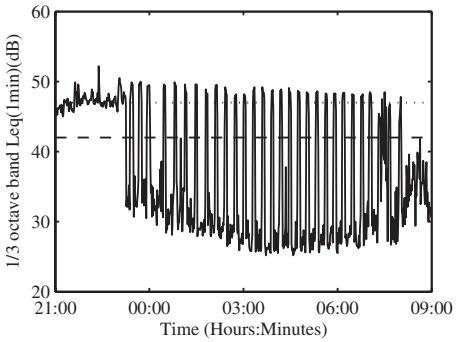


FIG. 2. Case where an environmental sound was positively identified. Time history showing 63 Hz 1/3 octave spectrum band (solid) with daytime (dotted) and nighttime (dashed) criteria.

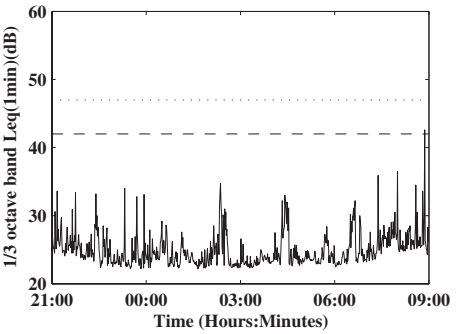


FIG. 4. Case where no environmental sound was identified. Time history showing 80 Hz 1/3 octave spectrum band (solid) with daytime (dotted) and nighttime (dashed) criteria.

TABLE I. Categorization of case studies.

Environmental sound category	No. of cases
Positively identified	3
Marginal	3
No environmental source found	5

was found to be most time-consuming. Categorization following analysis of the case studies is summarized in Table I, and examples are discussed below.

Examining the control cases, where there had been no complaints, it is interesting that four of the five exceeded the criterion curve, the fifth being an anechoic chamber with low background noise levels. This emphasizes that the criterion curve should not be treated as a strict indicator of acceptability or otherwise.

B. Social effects of LFN on sufferers

1. The qualitative methodology

In this section, the rationale for collecting details during the fieldwork about individual's residential and occupational histories is presented. Many complainants have ongoing problems that they associate with LFN, and which have a serious impact on their lives. However, human reaction to sound is known to be dependent not just on the sound itself, but on a complex array of other factors.[7] In addition to making physical recordings of the sounds within complainants' residences, it was therefore necessary to obtain a significant amount of personal data from the individuals themselves. This was important in order to obtain an overview of the background to the LFN complaint that might have a bearing on the responses.

Interviews were performed by an experienced interviewer in the complainants' home, during the daytime while the acoustical measurement equipment was being installed. Using a comprehensive one-to-one structured interview schedule, details were collected about each individual's residential and occupational histories, their general health, details of the noise they were exposed to, suspected sources of the noise, effects of the noise on themselves and their health, and any measures they may have taken to cope with or avoid the noise. Each participant of the field trials answered all questions without hesitation, and was forthcoming and open when answering questions relating to their general and mental health, and when providing detailed information about their noise problem.

2. Symptoms reported by LFN complainants

Reactions to the problem ranged from an annoyed interest to feeling suicidal. Symptoms were identified by asking complainants a number of personal questions about their current general health. First, they were asked to describe their general health in their own words. They were then asked to list any symptoms they suffered from, whether or not they attributed these to the noise problem, and to indicate for how long they had suffered each symptom. They were asked about any known hearing problems, when they had last had a hearing test, what the outcome of the test was, and whether they were satisfied with that outcome. Each complainant was specifically asked if they suffered from tinnitus, although their self-reports were not independently verified by a hearing specialist. Following this detailed health discussion, which allowed complainants to name their health problems in their own words, a final question was asked where a list of other symptoms was read out and the complainant was asked whether they suffered from any of them. The list of symptoms was based on that published by Leventhall *et al.*[1] Again, it was made clear that they should say whether they suffered from the symptom or whether or not they attributed it to exposure to LFN. The combination of open questions and the list of known symptoms meant that a full set of health issues was identified for each complainant. Some complainants practiced successful coping strategies at the time of the interview and so were asked to report health problems at the time when their suffering from the noise had been at its worst. Table II summarizes some of the more striking findings.

TABLE II. Numbers of LFN complainants reporting selected symptoms.

Health issue	No. of respondents	Percentage of respondents (%)
Sleep disturbance	11	92
Stress	10	83
Frustration	9	75
Difficulty falling asleep	8	67
Anxiety	8	67
Tiredness	7	58
Pressure or pain in ear or body	7	58
Headaches	7	58
Body vibration or pain	6	50
Frequent irritation	5	42
Insomnia	5	42
Depression	4	33
Migraine	3	25
Abdominal symptoms	3	25
Chronic fatigue	2	17
Suicidal	2	17
Tinnitus	1[a]	8

[a]Respondent attributed whistling in ear to sinusitis rather than tinnitus.

3. Discussion of findings from the semi-structured interview

The results indicate that all the complainants in the study had ongoing problems that they associated with LFN, and that had a serious impact on their quality of life. None of the complainants had a history of suffering from these problems at previous residences, and none had an employment or other discernable relationship with the company or organization suspected as the source of the LFN about which they complained. Furthermore, as far as can be judged by an experienced interviewer, the complaints were genuine, and there was no hint of ulterior motives.

4. Assessment of methodology

Combining measurements with a questionnaire gave a significant amount of personal data about the individuals and gave an overview of the background to the LFN complaint that might have a bearing on the responses. These sociological factors were incorporated into the procedure in the form of a questionnaire to be used by the investigating environmental health practitioners. The answers to the questions are intended to help local authorities distinguish cases where they should intervene from those where they can do nothing to help.

C. Laboratory tests

1. Objectives of the laboratory tests

The objective of the laboratory tests was to establish thresholds of acceptability for low frequency sounds, for day and night exposures. Previous work, including most national guidelines,[1] is based on the idea that the acceptability of a low frequency sound can be evaluated in relation to a frequency-dependent reference curve. Such a curve can be called the threshold of acceptability: Sounds with a higher intensity would be considered unacceptable, and those with a lower intensity acceptable.

A further objective of these tests was to investigate the effect of fluctuations on the disturbance caused by LFN. Specifically the questions to be addressed were as follows.

(i) Should fluctuating low frequency sounds be penalized compared with steady sounds?
(ii) If so, then by how much?
(iii) What measured parameter(s) should be used to determine when such a penalty should be applied?

The aim was to derive a method suitable for use by environmental health practitioners to quantify the effect of fluctuations. It is not possible to reproduce realistic field conditions in a laboratory test. In particular, the length of exposure does not give an adequate impression of what it is like to live with the sound. Therefore, it was not the objective of these laboratory tests to establish absolute levels for a reference curve.

2. Methodology for laboratory tests

The threshold of acceptability was defined as the level of a particular sound that the subject judged to be just acceptable for an assumed daytime or nighttime situation. Thresholds of acceptability were determined by the method of adjustment for a number of fluctuating and steady sounds. The subject was seated in a simulated living room into which pre-recorded low frequency sounds were to be played, and the following instructions were read to the subject.

"Imagine you are at home during the day. Press the button whenever you consider the sound is not acceptable to live with and keep it pressed. Whenever you consider the sound is acceptable to live with, release the button."

An operator then adjusted levels using similar techniques to those used in audiometry, reducing the level of the sound when the button was pressed until it was released. A coarse adjustment was made up and down to find an approximate threshold during the first few seconds followed by finer adjustments. Each sample lasted 90 s, which had been found during preliminary tests to be sufficient time to obtain a reliable threshold. It was found that after an initial training period the threshold levels were repeatedly set to within 1 dB. For the "nighttime" tests the main lights were switched off and the first sentence of the instruction was replaced with the following: "Imagine you are at home at night and trying to get to sleep."

TABLE III. Details of composition of synthesized tones.

Synthesized tone	Component sinusoids
Steady 1	40 Hz at 0 dB
Steady 2	60 Hz at 0 dB
Beating 1	40 Hz at 0 dB
	41.5 Hz at −8 dB
Beating 2	60 Hz at 0 dB
	61.5 Hz at −8 dB

The set of sounds presented to subjects comprised a combination of real and synthesized sounds that was developed and refined during a series of preliminary tests. Three sets of sounds were used:

(a) real sounds from field recordings,
(b) steady synthesized tones, and
(c) beating synthesized tones.

The advantage of real sounds is that they are known to have caused disturbance. The advantage of synthesized sounds is that they can be controlled so that only one aspect of the sound is varied at once. Specifically, this allowed control of the amount of fluctuation while keeping other characteristics of the sound constant. The real sounds were taken from the field measurements made in the dwellings of LFN sufferers. It was necessary to ensure that parameters such as tonality and frequency content were kept constant, and that only the fluctuation varied. Synthesized tones were constructed from 40 and 60 Hz sinusoids, using both steady tones of single frequencies, and beating tones formed from two sinusoids of similar frequencies as shown in Table III.

3. Choice of subjects

The choice of both the number and makeup of subjects was an important consideration. Regarding the profile of subjects, LFN sufferers tend to be middle aged or elderly, and the majority is women.[1] In addition, there is evidence that people known to be disturbed by LFN will judge sounds differently to a cross section of non-sufferers. Consequently, the profile summarized in Table IV was chosen.

4. Low frequency hearing thresholds

A conventional audiometric test was conducted on each subject over the frequency range 250 Hz–6 kHz to identify any hearing defects that could affect the results. In addition, low frequency audiometric tests were carried out in an anechoic chamber using pure tones played through a loudspeaker at the third octave band center frequencies between

TABLE IV. Makeup of subject groups for laboratory tests.

Group	Group profile	Average age	Male	Female	Total
0	Subjects known to be disturbed by low frequency sounds	62	0	3	3
1	Subjects with the age profile of typical LFN sufferers (55–70 years old) but without a history of disturbance by LFN	60	5	3	8
2	Subjects from a younger age group chosen at random	32	2	5	7
All		50	7	11	18

31.5 and 160 Hz. Each subject took part in three listening sessions and one training session, each lasting 20 min.

Figure 5 shows the hearing thresholds of all subjects averaged over each group. There was a spread of between 25 and 40 dB between the most and least sensitive subjects. The younger age group (group 2) has more sensitive hearing than the 55–70 year old group (group 1) by about 5 dB as might be expected. The shapes of the spectra follow the ISO reference threshold of hearing,[8] and the levels show good agreement given that the ISO curve applies to 18–25 year olds whereas the average age of the subjects was 60 and 32 years for groups 1 and 2, respectively. The least sensitive group in terms of hearing threshold is group 0 (sufferers).

5. Threshold of acceptability for real sounds

Thresholds of acceptability for the real sounds in the nighttime are shown in Fig. 6 for all subjects. These tracks were carefully produced from a selection of short recordings from the 5-day record from a case study in which the source was essentially the same, but the degree of fluctuation of the sound varied. There is a wide spread of results, which might be expected given the wide range of hearing thresholds. However, the lines are surprisingly parallel, indicating that all subjects responded in a similar way to the various sounds, but at a different overall level.

Figure 7 shows the same data as Fig. 6 but averaged by group. The authors see that group 0 (sufferers) is less sensitive in absolute terms than the other groups, by about 2–4 dB. There is no significant difference in the responses of the other two groups. Subjects were generally more tolerant of track 1, which displayed the smallest fluctuations by about 5 dB, and judged the other four sounds to be similar in terms of their acceptability.

It might be expected that acceptability thresholds would depend on hearing thresholds, and it is therefore interesting to examine the difference between these two thresholds for each group. These data are given in Fig. 8 for the nighttime scenario. On average respondents set the nighttime thresholds 2 dB lower than for the day, and the difference between day and night was almost identical for each sound. This result suggests that there was no qualitative difference in the sounds, and that no particular sound was relatively more disturbing just at night. Two important points can be derived from these data.

(i) Sufferers tend to set acceptable levels close to their threshold of hearing, both day and night.
(ii) The youngest group was most tolerant, and the older group less so, to these sounds.

In absolute terms, the sufferers in these tests were the least sensitive group to low frequency sounds. A major factor in this is that sufferers' thresholds of hearing were higher than those of other groups. The authors should avoid strong general conclusions because only three sufferers were tested, and there was variation between them. Nevertheless, this finding contradicts the view sometimes expressed that LFN problems are a result of exceptional sensitivity. In relative terms, sufferers tend to set the threshold of acceptability much closer to the threshold of hearing than other groups. Whether this is because they are naturally less tolerant or have become sensitized by exposure is not known.

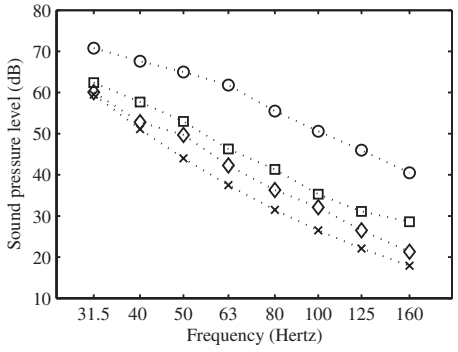


FIG. 5. Average low frequency hearing thresholds for each group. Group 0 (LFN sufferers) (○), group 1 (55–70 years old) (□), group 2 (younger age) (◇), and ISO 226 (0 dB threshold curve) (x).

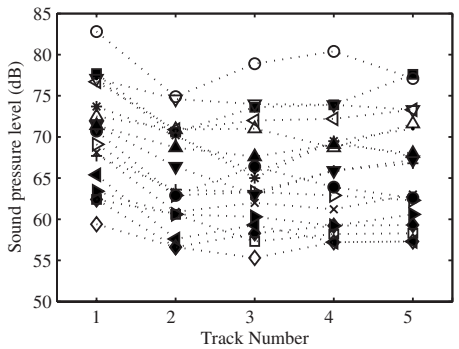


FIG. 6. Nighttime thresholds of acceptability to real sounds, all subjects.

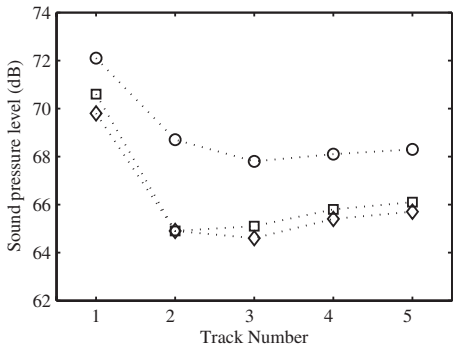


FIG. 7. Nighttime thresholds of acceptability to real sounds, averaged by group. Group 0 (LFN sufferers) (○), group 1 (55–70 years old) (□), and group 2 (younger age) (◇).

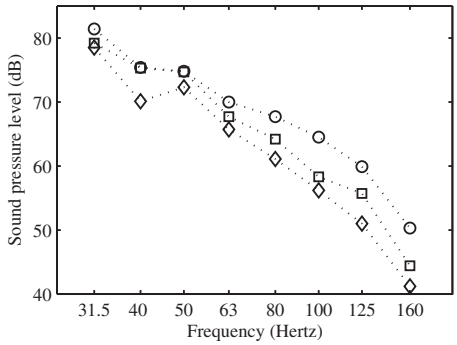


FIG. 9. Nighttime acceptability thresholds for tones by group. Group 0 (LFN sufferers) (○), group 1 (55–70 years old) (□), and group 2 (younger age) (◇).

6. Threshold of acceptability for steady tones

Figure 9 shows the thresholds of acceptability for steady tones in the nighttime scenario set by all subjects averaged over each group. There was a spread of about 30 dB between the most and least sensitive subject. This is not surprising given that the thresholds of hearing have a similar spread. In absolute terms, the LFN sufferers are the least sensitive group, followed by the older and then the younger group. As mentioned above, this contradicts the often-held view that LFN sufferers tend to be particularly sensitive.

Shown in Fig. 10 are the relative nighttime acceptability thresholds, i.e., the difference between the threshold of acceptability and of hearing for each individual, averaged by group. There was ~35 dB spread in the results. Some subjects set the threshold of acceptability only a few decibels above their hearing threshold, judging that a sound that was only just audible to be unacceptable. Others set the difference very much higher, so that the sound would be clearly audible before they judged it unacceptable.

Two points of interest can be made. First, there is a marked difference in the average response of LFN sufferers compared with the other two groups. LFN sufferers set the acceptable level about 10 dB higher than hearing threshold on average, whereas for non-sufferers, the difference was about 20 dB. Thus, the authors can say that relative to their hearing threshold the LFN sufferers are more sensitive than are non-sufferers, although as stated above in absolute terms they were less sensitive. However, the authors should again be cautious about drawing general conclusions based on three subjects.

The second point is that for the lower frequency bands, the threshold of acceptability reduces, i.e., gets closer to the threshold of hearing. This is significant since it suggests that the optimum shape of a reference curve does not follow the threshold of audibility over the whole of the low frequency range. Rather, it will tend to follow the hearing threshold for the lower bands but then move away from it above around 50 Hz.

7. Threshold of acceptability for beating tones

There are several clear trends. First, as before, group 0 (LFN sufferers) is the most sensitive group in relative terms, setting the acceptability threshold only 2–3 dB above audibility threshold for nighttime beating tones. Second, subjects were 3–5 dB more tolerant of steady tones than of the corresponding beating tone. This is consistent with previous published research,[9,10] and proposed revisions to American National Standard criteria for evaluating room noise with regard to quiet ventilation system design.[11,12] This is also consistent with the Danish standard[13] method of adding a 5 dB penalty for impulsive noise, as well as existing United Kingdom guidelines[14] where a 5 dB penalty is added for noise with noticeable features.

Third, daytime levels were set an average of 3–4 dB higher than the corresponding nighttime levels. This is a

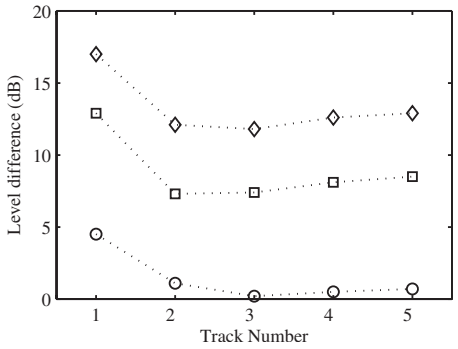


FIG. 8. Nighttime acceptability thresholds relative to hearing threshold for real sounds, averaged by group. Group 0 (LFN sufferers) (○), group 1 (55–70 years old) (□), and group 2 (younger age) (◇).

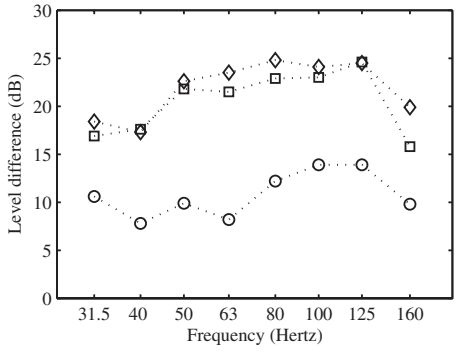


FIG. 10. Nighttime acceptability thresholds for tones relative to hearing thresholds by group. Group 0 (LFN sufferers) (○), group 1 (55–70 years old) (□), and group 2 (younger age) (◇).

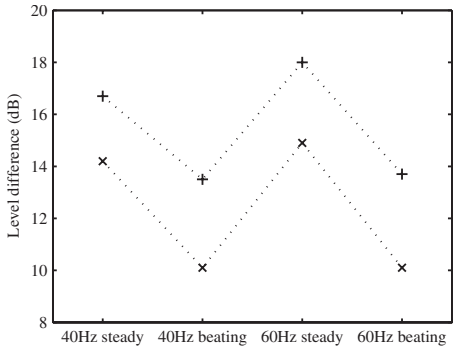


FIG. 11. Comparison of daytime (+) and nighttime (x) acceptability thresholds relative to hearing threshold for steady and beating tones. Average of all subjects.

slightly lower difference than the 5 dB daytime relaxation used in the German standard.[15] However, due to difficulty in reproducing realistic nighttime conditions, it is likely that this difference is underestimated in the laboratory tests.[16] Consequently, 5 dB is an appropriate relaxation to the limits for sounds only present during the day. Lastly, the effect of the beating on the response was essentially the same for day and night. This means that the procedure used to assess fluctuations can be applied equally to night and day. These last two points are illustrated most clearly in Fig. 11.

Two alternative methods are suggested here for the assessment of a sound for fluctuation. The first is based on the parameter known as prominence,[17] and is that a sound should only be considered fluctuating when the rate of change of the rms fast sound level in the third octave band of interest exceeds 10 dB/s. The second method uses the difference $L_{10}-L_{90}$ measured using a fast time constant, which has the additional advantage that it is generally available to environmental health practitioners. Shown in Fig. 12 are the relative nighttime acceptability levels plotted against the value of $L_{10}-L_{90}$ for each sound averaged for all subjects. In one of the preliminary tests, subjects were played a sequence of beating tones with varying degrees of fluctuation. It was found that the relative thresholds of acceptability were set at

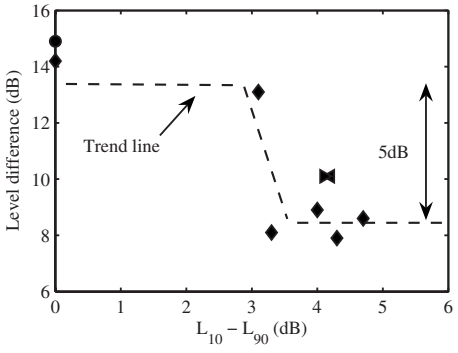


FIG. 12. Nighttime thresholds of acceptability relative to hearing threshold for real sounds (♦), steady tones (●), 40 Hz beating tones (▶), and 60 Hz beat tones (◀) averaged for all subjects. Showing variation with $L_{10}-L_{90}$. The beating tones' data points (▶) and (◀) are so close as to be indistinguishable. These listening test results are used to support the suggestion that a 5 dB penalty for fluctuations is appropriate when $L_{10}-L_{90}>5$ dB.

about the same level for the various beating tones, but that there was a clear difference of ~5 dB from those for the steady tones.

Arguably, Fig. 12 also displays this trend: The most fluctuating sounds, represented by points to the right, display a "penalty" of ~5 dB compared with steady sounds on the left. The overall trend can be simplified without much loss of accuracy. The simplified trend can then be described as follows.

(i) $L_{10}-L_{90}<3$ dB: no penalty.
(ii) $L_{10}-L_{90}>3$ dB: penalty of 5 dB.

Although based on a small data set, this relationship is in a pragmatic form that could be used by environmental health practitioners to decide whether to apply the 5 dB penalty. While useful, the difference $L_{10}-L_{90}$ is not a foolproof parameter since, for example, the same value of $L_{10}-L_{90}$ can be obtained for a slowly varying or for a rapidly varying sound, whereas experience suggests that they would be judged subjectively differently in terms of acceptability. An additional criterion was therefore introduced,[3] i.e., that the 5 dB penalty would only apply where the slope of the sound level (rms fast) curve exceeds 10 dB/s.

III. LFN ASSESSMENT PROCEDURE

The LFN assessment procedure is detailed in the procedure for the assessment of LFN complaints,[5] together with guidance notes and a pro-forma report with systematic instructions. Measurements for the procedure for the assessment of LFN complaints require detailed acoustical monitoring over a period of 3–5 days combined with a synchronized log completed by the complainant. There are then two aspects to the assessment procedure:

(1) comparison of the level of recorded sound with a third octave band criterion curve and
(2) evaluation of the correlation between the recorded sound and the complainant's log.

A. The criterion curve

The criterion curve is given in Table V and Fig. 13. If the noise occurs only during the day then 5 dB relaxation may be applied to all third octave bands. Note that the criterion curve sound levels given in Table V for 25 Hz and below can cause the vibration of windows, walls, and even floors in residential housing structures with the accompanying rattling of dishes and bric-a-brac. This induced vibration and the accompanying secondary noises will be noticed by residents, with annoyance the likely result. Some account of vibration-induced noise is made in the Japanese method for the assessment of LFN complaints.[18]

B. Evaluation of the recordings and complainant's log

The following provides a step-by-step guide to analysis.

(1) Consult the complainant's log to find times when the sound was considered most disturbing.

TABLE V. Proposed nighttime reference curve.

Hz	10	12.5	16	20	25	31.5	40	50	63	80	100	125	160
L_{eq} (dB)	92	87	83	74	64	56	49	43	42	40	38	36	34

(2) If possible, check the character of the sound at these times by audio playback.
(3) If the sound is predominantly due to traffic or movement within the building then reject this sample (the procedure excludes evaluation of traffic noise over which environmental health practitioners have no control in the United Kingdom).
(4) For the chosen time obtain the third octave band spectrum of $L_{eq,T}$ samples.
(5) Compare the $L_{eq,T}$ spectrum to the criterion curve to find any third octave bands for which the criterion curve is exceeded.
(6) For the third octave band, which exceeds the curve by the greatest margin, plot the time variation of the $L_{eq,T}$ for the 24 h period in which the event occurred.
(7) Compare the complainant's log with the time history to see whether there is correlation between the two.

As described in Sec. II C, the laboratories' tests, the threshold of the hearing varies significantly between subjects, while fluctuation can further affect the threshold of acceptability. Consequently, the criterion curve should not be considered as an absolute limit. Rather, the criteria curve is provided for guidance when investigating to identify an environmental sound potentially responsible for the complaint.

IV. FIELD TRIALS BY ENVIRONMENTAL HEALTH PRACTITIONERS

A. Objectives of the field trials

Generally, some caution is needed in applying laboratory test results to real situations since laboratory experiments cannot reproduce the possible effects of sensitization over time, or account for the physical modification and enhancement of the experienced sound field *in-situ*. Nevertheless, it is believed that findings from laboratory testing can be reliably applied to provide a clearer understanding of the disturbance experienced by LFN sufferers in their homes. Consequently, it was resolved to undertake genuine trials of the procedure. Cases were solicited by letter and electronic mail requests to 62 local authorities around the United Kingdom, and a series of six trials of live LFN complaints was conducted by volunteers from environmental health departments.

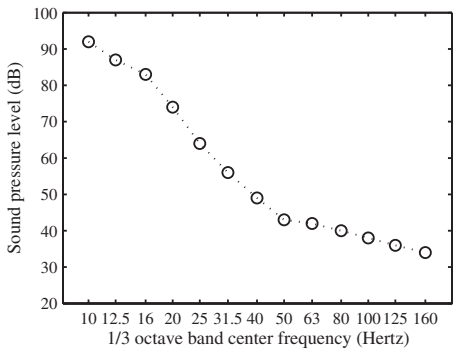


FIG. 13. Criterion curve from the procedure for the assessment of LFN.

B. Results of the field trials

In two out of the six cases an environmental noise was identified and its source located. In the remaining four cases, no environmental noise was found and the officers concluded that there was no remedial action they could take. In each field study, the sound measurements were supported by a semi-structured interview as detailed in the procedure, to determine whether sociological or other factors might influence the results. Combining measurements with a questionnaire provided a significant amount of personal data about the individuals and an overview of the background to each LFN complaint that might have a bearing on the responses to the perceived noise.

C. Findings from the debriefing session

1. General comments

A debriefing session was arranged to obtain feedback from the environmental health practitioners following application of the procedure to their "live" cases. The environmental health practitioners' experience in applying the procedure was generally very positive: The participating officers found the procedure easy to use and that working to a set procedure increased their confidence and the complainant's acceptance of the results. They also considered that the procedure achieved a good balance, giving a set method but allowing them the flexibility to form their own conclusions.

2. Experience of the semi-structured interview

With particular reference to the environmental health practitioners' evaluation of the subjective part of the procedure, it was felt that the interview provided a formal way of acquiring sensitive information that was relevant to the analysis. During the debriefing one practitioner confirmed the following: "[The complainant was] generally happy to be asked. It showed we were leaving no stone unturned." It also engendered trust and confidence in the environmental health practitioner on the part of the complainant who could see the rigor that was being applied to their case. The interviews gave the complainants the sense that they were being listened to and that everything possible was being done to help them. In particular, it was recognized; "Doing an interview that is formalized makes us able to tell them this is a way of gathering data and provides us with a pathway to give us confidence. It meant the complainant could see we'd done our best." Subsequently, some complainants in the field trials were satisfied that their case was closed even though no environmental source was found.

3. Assessment of application of the procedure in the field

The environmental health practitioners were generally able to draw firm conclusions and reach "closure" even if there was nothing they could do to help. As one practitioner stated, "The Procedure… raised our credibility and the complainant's acceptance of the findings." The environmental health practitioners who had found no low frequency environmental noise present commented that the lack of an alternative, more appropriate course of action for the complainants was a remaining difficulty. Currently there is no further formal advice that can be given to help complainants who are still suffering with their problem. There was a strong feeling that officers need somewhere to send people affected in this way. Consequently, there was a sentiment among the environmental health practitioners that an initiative to develop a further course of action would be endorsed, perhaps along the lines of "relief strategies."

V. DISCUSSION

In many cases, environmental health practitioners will find a noise source above audible thresholds that clearly correlates with the complainant's log, typically an industrial process of some kind, fans, pumps, or electrical equipment. However, in 8 of the 11 cases considered in the fieldwork, no environmental source could be found for the LFN complaint. In fact, a striking feature about many LFN sufferers' homes considered was the almost complete absence of any intrusive environmental noise. Further, in four out of the six field trials of the procedure performed by environmental health officers, no environmental noise consistent with the complaint could be found. Similar proportions of categories were found by Pedersen *et al.*[19] investigating whether it is real physical sound or low frequency tinnitus that causes the annoyance.

The proposed criterion curve is provided as guidance for environmental health officers in their evaluation of a LFN complaint, and not as an absolute limit. This means that tonal sounds at, or just below, the threshold of the hearing should be considered as environmental sources potentially responsible for the complaint. The course of action when no environmental noise consistent with the complaint can be found, and yet the complainant is clearly distressed, is unclear. On one hand, suggestion that a medical screening for tinnitus is in order is often rejected by the complainant. On the other hand, the environmental health officer could simply close the case and avoid further involvement as it may lead to frustration and false hopes. However, this strategy is unacceptable to many environmental health practitioners since it involves leaving a problem unsolved with the complainant still in distress.

An answer to this conundrum may lie in the existing clinical audiology and auditory neuroscience literature. A complement to the procedure would be the development of techniques by which the sufferer might acquire a degree of control over their adverse reactions. Applying neuropsychological understanding of human hearing and tinnitus to LFN complaints, Moorhouse and Baguley[20] proposed that environmental health officers who have applied the procedure and not identified an environmental sound that could account for the complaint could refer cases to strategically located audiology departments. Trials along similar lines have been conducted in Japan[18] and United Kingdom.[21] Such a network could be established by providing specialist audiologists with some additional background knowledge about LFN.

VI. CONCLUSIONS

Until recently, it has been extremely difficult for environmental health practitioners in the United Kingdom and elsewhere to deal with complaints about LFN. This was in part because no official guidance was available to support them. The UK Defra Procedure for the assessment of LFN complaints has addressed this point. Feedback from environmental health practitioners taking part in field trials of the procedure has been very positive, indicating that procedure was easy to follow and strengthened the authority's position with the complainant. Furthermore, complainants were said to be significantly reassured once they saw that a detailed procedure was being followed. The authors expect a reasonable proportion of cases to remain unresolved even with the application of the procedure, since a "no environmental source found" conclusion may not resolve the matter for many LFN sufferers. Nevertheless, this does not negate the value of a procedure that provides environmental health practitioners with a means of distinguishing cases where they should act from those where they can do nothing to help. It does, however, indicate the need for some alternatives for those LFN sufferers not satisfied with the outcome.

In absolute terms, the sufferers that participated in the laboratory tests were group least sensitive to low frequency sounds. A significant factor is that their thresholds of hearing were higher than other groups. This finding contradicts the view sometimes expressed that LFN problems are a result of exceptional sensitivity. Levels of acceptability were set typically 3–5 dB lower for sounds with strong fluctuations than for steady sounds. It is therefore appropriate to penalize fluctuating sounds compared with steady sounds, and that 5 dB is an appropriate level for such a fluctuation penalty. Although the laboratory tests yielded some interesting results, strong conclusions cannot be drawn due to the small sample size.

ACKNOWLEDGMENTS

The authors are grateful to Defra for funding the work leading to this paper. Thanks are due to Dr. Geoff Leventhall, and the authors would like to acknowledge the contribution of Dr. David Baguley. The authors would like to thank the environmental health practitioners that performed the field trials of the procedure. The authors would like to thank the referees for their helpful comments. Work funded by the Department for Environment, Food and Rural Affairs (Defra), United Kingdom.

[1]G. Leventhall, P. Pelmear, and S. Benton, "A review of published research on low frequency noise and its effects," Department for Environment, Food and Rural Affairs, London, 2003, http://www.defra.gov.uk/environment/noise/research/lowfrequency/pdf/lowfreqnoise.pdf (Last viewed 3/26/2009).

[2]H. Guest, "Inadequate standards currently applied by local authorities to determine statutory nuisance from LF and infrasound," Low Freq. Noise, Vib., Act. Control **22**, 1–7 (2003).
[3]A. T. Moorhouse, D. C. Waddington, and M. Adams, "Proposed criteria for the assessment of low frequency noise disturbance," Department for Environment, Food and Rural Affairs, London, 2004, http://www.defra.gov.uk/environment/noise/research/lowfrequency/pdf/nanr45-criteria.pdf (Last viewed 3/26/2009).
[4]A. T. Moorhouse, D. C. Waddington, and M. Adams, "Field trials of proposed procedure for the assessment of low frequency noise complaints," Department for Environment, Food and Rural Affairs, London, 2004, http://www.defra.gov.uk/environment/noise/research/lowfrequency/pdf/nanr45-fieldtrials.pdf (Last viewed 3/26/2009).
[5]A. T. Moorhouse, D. C. Waddington, and M. Adams, "Procedure for the assessment of low frequency noise disturbance," Department for Environment, Food and Rural Affairs, London, 2005, http://www.defra.gov.uk/environment/noise/research/lowfrequency/pdf/nanr45-procedure.pdf (Last viewed 3/26/2009).
[6]P. Schomer, "The importance of proper integration of and emphasis on the low-frequency sound energies for environmental noise assessment," Noise Control Eng. J. **52**, 26–39 (2004).
[7]B. Schulte-Fortkamp and A. Fiebig, "Soundscape analysis in a residential area: An evaluation of noise and people's mind," Acta Acust. Acust. **92**, 875–880 (2006).
[8]"Acoustics. Normal equal-loudness-level contours," BS ISO 226, British Standards, 2003.
[9]J. S. Bradley, "Annoyance caused by constant-amplitude and amplitude-modulated sounds containing rumble," Noise Control Eng. J. **42**, 203–208 (1994).
[10]T. Poulsen and F. R. Mortensen, "Laboratory evaluation of annoyance of low frequency noise," Working Report No. 1, Danish Environmental Protection Agency, 2002, http://www.mst.dk/udgiv/publications/2002/87-7944-955-7/pdf/87-7944-956-5.pdf (last viewed 7/17/2009).
[11]P. D. Schomer, "Proposed revisions to room noise criteria," Noise Control Eng. J. **48**, 85–96 (2000).
[12]P. D. Schomer and J. S. Bradley, "A test of proposed revisions to room noise criteria curves," Noise Control Eng. J. **48**, 124–129 (2000).
[13]J. Jakobsen, "Danish guidelines on environmental low frequency noise, infrasound and vibration," Low Freq. Noise, Vib., Act. Control **20**, 141–148 (2001).
[14]"Method for rating industrial noise affecting mixed residential and industrial areas," BS 4142, British Standards, 1997.
[15]"Measurement and rating of low frequency noise emission in the neighbourhood (in German)," Standard DIN 45680, Deutsches Institut für Normung e.V., Berlin, 1997.
[16]Y. Inukai, N. Nakamura, and H. Taya, "Unpleasantness and acceptable limits of low frequency sound," Low Freq. Noise, Vib., Act. Control **19**, 135–140 (2000).
[17]T. H. Pederson, "Objective method for measuring the prominence of impulsive sounds and for adjustment of LAeq," presented at the InterNoise, The Hague, The Netherlands (2001).
[18]T. Kitamura, M. Hasebe, and S. Yamada, "Psychological analysis of complainants on noise/low frequency noise and the relation between psychological response and brain structure," Low Freq. Noise, Vib., Act. Control **24**, 43–48 (2005).
[19]C. S. Pedersen, H. Møller, and K. P. Waye, "A detailed study of low-frequency noise complaints," Low Freq. Noise, Vib., Act. Control **27**, 1–33 (2008).
[20]A. Moorhouse and D. Baguley, "Sound advice: Solution to the often intractable problem of low frequency noise complaints," Environmental Health Practitioner **115**, 20–24 (2007).
[21]G. Leventhall, S. Benton, and D. Robertson, "Coping strategies for low frequency noise," Low Freq. Noise, Vib., Act. Control **27**, 35–52 (2008).

Leakage effect in Helmholtz resonators

Ahmet Selamet[a] and Hyunsu Kim
Department of Mechanical Engineering and The Center for Automotive Research, The Ohio State University, Columbus, Ohio 43212

Norman T. Huff
Owens Corning Automotive, Novi, Michigan 48377

(Received 11 November 2008; revised 26 June 2009; accepted 29 June 2009)

The effect of leakage in Helmholtz resonators has been investigated in this predominantly experimental study combined with a computational effort. A prototype has been built with varying levels of intentional leakage due to holes in the baffle and gaps between the baffle and the housing. The transmission loss is then measured with different combinations of holes and/or gaps. Such openings, even though their cross-sectional areas are small, are found to have a significant impact on transmission loss. The effect of holes versus gaps is also compared as a function of the leakage area. The present study illustrates the critical need to account for such leakages at the design stage for the proper tuning of these resonators.
 [DOI: 10.1121/1.3183416]

PACS number(s): 43.50.Gf [AH] Pages: 1142–1150

I. INTRODUCTION

Helmholtz resonators (HRs) are widely used in many applications, including engines, compressors, and ventilation systems as effective narrow band noise attenuators, particularly at low frequencies. As a result, an extensive literature has developed on this configuration dating back to Ingard (1953) who investigated the effect of neck geometry on the acoustic performance of HRs, as well as the interaction between a pair of circular necks. The studies that have followed Ingard (1953) on HRs are numerous, including Alster (1972) and Chanaud (1997), among others. This earlier literature has been elaborated upon by Selamet *et al.* (2005), which will not be repeated here.

During the past decade, Selamet and co-workers furthered the understanding of HRs by implementing multi-dimensional analytical and computational approaches (Selamet *et al.*, 1997 on circular concentric HRs; Selamet and Ji, 2000 on circular asymmetric HRs; Selamet and Lee, 2003 on HRs with extended necks; and Selamet *et al.*, 2005 on HRs with fibrous material). The computational approach used in these studies primarily employed a multi-dimensional boundary element method (BEM). Of particular interest here is the work of Selamet and Lee (2003) where the acoustic performance of a HR with perforations on the neck extended into the cavity was shown to vary as a strong function of the porosity. Such perforations were found to increase the resonance frequency, while reducing the peak transmission loss (TL) particularly at small openings. At high porosities, the acoustic impact of walls of neck extension had essentially diminished, with the behavior approaching that of a HR in the absence of neck extension.

Acoustic impedance of small orifices has been studied by a number of investigators following Sivian's original work in 1935 (Silvian, 1935). Bolt *et al.* (1949), Ingard and Labate (1950), Bies and Wilson (1957), Melling (1973), Goldman and Chung (1982), Sullivan and Crocker (1978), and more recently Dickey and Selamet (1998) determined the acoustic impedance of small holes, while observing an increase in resistance in the nonlinear region with oscillating flow velocity at the orifice. The measured resistance and reactance can be applied to BEM for improved predictions of acoustic properties of a system. Lee *et al.* (2006a) measured the acoustic impedance of perforated plates for 11 different hole geometries as an extension of Sullivan and Crocker (1978), followed by an implementation of this impedance into BEM for acoustic performance predictions.

As illustrated in the foregoing studies, the resonance of a reactive HR is typically dictated by its cavity and neck geometry, including their dimensions and relative orientation. However, when the cavity has an additional passage besides the neck to communicate, for example, with the main duct, the resonance characteristics of the HR can be altered significantly. Such passages or "leakages" may be present in practical designs due to drain holes implemented into baffles (confining the HR cavity) and/or gaps between the baffles and the outer housing due to manufacturing constraints/tolerances. Bemman *et al.* (2005), for example, reported louder exhaust tailpipe noise with increasing leakage in the HR at low engine speeds. However, they did not provide an explanation as to how the leakage affected the acoustic performance. The objectives of the present study are then to (1) investigate the leakage effect in HRs through systematic experiments with a prototype and (2) predict the resonance frequency by implementing a measured acoustic impedance of a small orifice into BEM.

Following this introduction, Sec. II develops the basics of the BEM model for a HR with leakage and describes the experimental approach to measure the acoustic impedance of a hole. Section III elaborates on the prototype HR, followed

[a]Author to whom correspondence should be addressed. Electronic mail: selamet.1@osu.edu

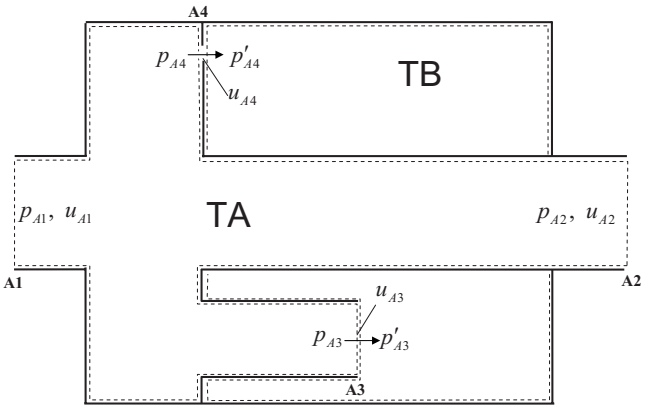


FIG. 1. BEM model of a HR with leakage.

by the discussion of experimental and computational results on the leakage effect in Sec. IV. The study is concluded with final remarks in Sec. V.

II. THEORY

A. BEM

A leakage model of a HR has been analyzed by BEM. Two domains used in the BEM model are designated in Fig. 1 by *TA* (highlighted) and *TB*. The domain *TA* includes the main pipe with inlet (*A*1) and outlet (*A*2). The domain *TA* is connected to *TB* (the HR volume) at the neck (*A*3) and the leakage (*A*4). At these connections, the absolute value of acoustic velocities (u_{A3},u_{A4}) is the same between two domains, with the directions being opposite. The acoustic pressure differentials at the connections may be expressed as

$$p_{A3} - p'_{A3} = Z_0 \zeta_{A3} u_{A3}, \quad p_{A4} - p'_{A4} = Z_0 \zeta_{A4} u_{A4}, \tag{1}$$

where $Z_0 = \rho_0 c_0$ is the characteristic impedance of air, u_{A4} is the velocity within the hole, and ζ_{A3} and ζ_{A4} are the nondimensional acoustic impedances of the connecting areas which will be described later. Then, the impedance matrix defined by Lee *et al.* (2006b) for domain *TA* may be written as

$$\begin{Bmatrix} p_{A1} \\ p_{A2} \\ p_{A3} \\ p_{A4} \end{Bmatrix} = Z_0 \begin{bmatrix} TA_{11} & TA_{12} & TA_{13} & TA_{14} \\ TA_{21} & TA_{22} & TA_{23} & TA_{24} \\ TA_{31} & TA_{32} & TA_{33} & TA_{34} \\ TA_{41} & TA_{42} & TA_{43} & TA_{44} \end{bmatrix} \begin{Bmatrix} u_{A1} \\ u_{A2} \\ u_{A3} \\ u_{A4} \end{Bmatrix}, \tag{2}$$

where TA_{ij} is the impedance matrix between i and j positions. The impedance matrix for domain *TB* can also be expressed as

$$\begin{Bmatrix} p'_{A3} \\ p'_{A4} \end{Bmatrix} = -Z_0 \begin{bmatrix} TB_{11} & TB_{12} \\ TB_{21} & TB_{22} \end{bmatrix} \begin{Bmatrix} u_{A3} \\ u_{A4} \end{Bmatrix}. \tag{3}$$

By combining Eqs. (2) and (3), the relationships between pressure and acoustic velocities at locations *A*1-*A*4 become

$$p_{A1} = Z_0(TA_{11}u_{A1} + TA_{12}u_{A2} + TA_{13}u_{A3} + TA_{14}u_{A4}), \tag{4}$$

$$p_{A2} = Z_0(TA_{21}u_{A1} + TA_{22}u_{A2} + TA_{23}u_{A3} + TA_{24}u_{A4}), \tag{5}$$

$$p_{A3} - p'_{A3} = Z_0\{TA_{31}u_{A1} + TA_{32}u_{A2} + (TA_{33} + TB_{11})u_{A3} + (TA_{34} + TB_{12})u_{A4}\}, \tag{6}$$

$$p_{A4} - p'_{A4} = Z_0\{TA_{41}u_{A2} + TA_{42}u_{A2} + (TA_{43} + TB_{21})u_{A3} + (TA_{44} + TB_{22})u_{44}\}. \tag{7}$$

Using the acoustic impedance of a hole defined by Eq. (1), Eqs. (6) and (7) can be rearranged in the matrix form as

$$\begin{bmatrix} TA_{31} & TA_{32} \\ TA_{41} & TA_{42} \end{bmatrix} \begin{Bmatrix} u_{A1} \\ u_{A2} \end{Bmatrix} = -\begin{bmatrix} (TA_{33} + TB_{11} - \zeta_{A3}) & (TA_{34} + TB_{12}) \\ (TA_{43} + TB_{21}) & (TA_{44} + TB_{22} - \zeta_{A4}) \end{bmatrix} \times \begin{Bmatrix} u_{A3} \\ u_{A4} \end{Bmatrix} \tag{8}$$

or simply as

$$\begin{Bmatrix} u_{A3} \\ u_{A4} \end{Bmatrix} = \begin{bmatrix} TY_{11} & TY_{12} \\ TY_{21} & TY_{22} \end{bmatrix} \begin{Bmatrix} u_{A1} \\ u_{A2} \end{Bmatrix}, \tag{9}$$

where

$$\begin{bmatrix} TY_{11} & TY_{12} \\ TY_{21} & TY_{22} \end{bmatrix} = -\begin{bmatrix} (TA_{33} + TB_{11} - \zeta_{A3}) & (TA_{34} + TB_{12}) \\ (TA_{43} + TB_{21}) & (TA_{44} + TB_{22} - \zeta_{A4}) \end{bmatrix}^{-1} \times \begin{bmatrix} TA_{31} & TA_{32} \\ TA_{41} & TA_{42} \end{bmatrix}. \tag{10}$$

Eliminating u_{A3} and u_{A4} from Eqs. (4) and (5), and rearranging them in the matrix form yield

$$\begin{Bmatrix} p_{A1} \\ p_{A2} \end{Bmatrix} = Z_0 \begin{bmatrix} TI_{11} & TI_{12} \\ TI_{21} & TI_{22} \end{bmatrix} \begin{Bmatrix} u_{A1} \\ u_{A2} \end{Bmatrix}, \tag{11}$$

where

$$TI_{11} = TA_{11} + TA_{13}TY_{11} + TA_{14}TY_{21}, \tag{12}$$

$$TI_{12} = TA_{12} + TA_{13}TY_{12} + TA_{14}TY_{22}, \tag{13}$$

$$TI_{21} = TA_{21} + TA_{23}TY_{11} + TA_{24}TY_{21}, \tag{14}$$

$$TI_{22} = TA_{22} + TA_{23}TY_{12} + TA_{24}TY_{22}. \tag{15}$$

Equation (11) may then be rewritten involving transfer matrix T_{ij} as

$$\begin{Bmatrix} p_{A1} \\ \rho_0 c_0 u_{A1} \end{Bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{Bmatrix} p_{A2} \\ \rho_0 c_0 u_{A2} \end{Bmatrix}. \tag{16}$$

Finally, the TL can be determined from

$$\mathrm{TL} = 20 \log_{10}\left(\tfrac{1}{2}|T_{11} + T_{12} + T_{21} + T_{22}|\right). \tag{17}$$

B. Acoustic impedance of a hole

The nondimensional acoustic impedance of a plate perforated with holes is expressed by Sullivan and Crocker (1978) as

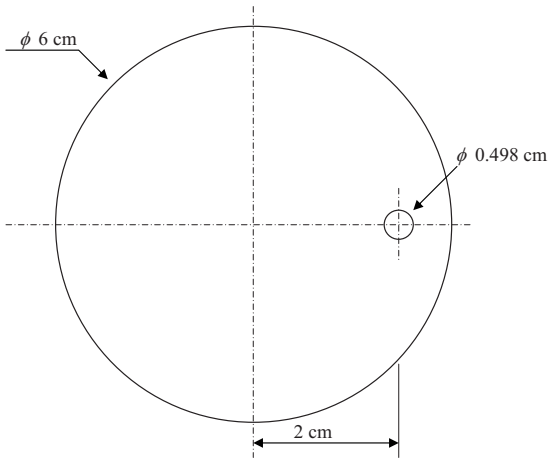


FIG. 2. Schematic of a plate with a single hole.

$$\zeta_p = \frac{Z_p}{Z_0} = \frac{R + ik_0(t_w + \alpha d_h)}{\phi}, \tag{18}$$

where k_0 is the wave number, t_w is the plate thickness, d_h is the hole diameter, and ϕ is the porosity of the plate. For $\phi = 4.2\%$, they obtained $R = 0.006$ for the resistance and $\alpha = 0.75$ for the end correction coefficient. Lee *et al.* (2006a) investigated 11 samples with various plate thicknesses, hole diameters, and porosity. They experimentally calculated R and α for four porosities (2.1%, 8.4%, 13.6%, and 25.2%), two wall thicknesses (0.08 and 0.16 cm), and two hole diameters (0.249 and 0.498 cm). The acoustic impedance of a hole (ζ_{A4}) needed in Eq. (1) may then be expressed by

$$\zeta_h = \phi\zeta_p. \tag{19}$$

One approximation to a single hole impedance may therefore be to use such R and α based on an equivalent porosity. An alternative approach would be to directly measure the acoustic impedance of the hole. Here, both approaches are adopted and contrasted.

A circular plate 6 cm in diameter and 0.14 cm thick, as shown in Fig. 2, has been fabricated to measure the acoustic impedance of a small hole 0.498 cm in diameter (corresponding to a porosity of 1%) and 2 cm away from the center. The hole is positioned intentionally closer to the perimeter to represent locations in actual hardware. An experimental setup for measuring the acoustic impedance of the hole is shown in Fig. 3. Using two microphone method, the pressure difference across and velocity through the hole can be calculated. In view of Fig. 3,

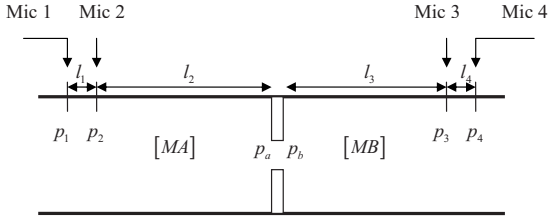


FIG. 3. Schematic of a two-microphone setup for acoustic impedance measurement of a hole.

$$\begin{Bmatrix} p_a \\ u_a \end{Bmatrix} = \begin{bmatrix} MA_{11} & MA_{12} \\ MA_{21} & MA_{22} \end{bmatrix}^{-1} \begin{Bmatrix} p_2 \\ u_2 \end{Bmatrix} = \begin{bmatrix} TX_{11} & TX_{12} \\ TX_{21} & TX_{22} \end{bmatrix} \begin{Bmatrix} p_2 \\ u_2 \end{Bmatrix}, \tag{20}$$

$$\begin{Bmatrix} p_b \\ u_b \end{Bmatrix} = \begin{bmatrix} MB_{11} & MB_{12} \\ MB_{21} & MB_{22} \end{bmatrix} \begin{Bmatrix} p_3 \\ u_3 \end{Bmatrix}, \tag{21}$$

where

$$[MA] = \begin{bmatrix} \cos kl_2 & j\rho_0 c_0 \sin kl_2 \\ j\dfrac{1}{\rho_0 c_0} \sin kl_2 & \cos kl_2 \end{bmatrix},$$

$$[MB] = \begin{bmatrix} \cos kl_3 & j\rho_0 c_0 \sin kl_3 \\ j\dfrac{1}{\rho_0 c_0} \sin kl_3 & \cos kl_3 \end{bmatrix}.$$

The acoustic impedance of plate, defined by

$$\zeta_p = \frac{p_a - p_b}{\frac{1}{2}\rho_0 c_0 (u_a + u_b)}, \tag{22}$$

may then be expressed by substituting, from Eqs. (20) and (21),

$$\begin{aligned} p_a &= TX_{11}p_2 + TX_{12}u_2, \\ u_a &= TX_{21}p_2 + TX_{22}u_2, \\ p_b &= MB_{11}p_3 + MB_{12}u_3, \\ u_b &= MB_{21}p_3 + MB_{22}u_3 \end{aligned} \tag{23}$$

as

$$\zeta_p = \frac{TX_{11}p_2 + TX_{12}u_2 - MB_{11}p_3 - MB_{12}u_3}{0.5\rho_0 c_0 (TX_{21}p_2 + TX_{22}u_2 + MB_{21}p_3 + MB_{22}u_3)}. \tag{24}$$

By further inserting

$$u_2 = \frac{p_1 - p_2 \cos kl_1}{j\rho_0 c_0 \sin kl_1}, \quad u_3 = \frac{p_3 \cos kl_4 - p_4}{j\rho_0 c_0 \sin kl_4}$$

into Eq. (24) and rearranging yields explicitly

$$\zeta_p = \frac{TX_{11}H_{23} + TX_{12}\dfrac{H_{13} - H_{23}\cos kl_1}{j\rho_0 c_0 \sin kl_1} - MB_{11} - MB_{12}\dfrac{\cos kl_4 - H_{43}}{j\rho_0 c_0 \sin kl_4}}{0.5\rho_0 c_0\left(TX_{21}H_{23} + TX_{22}\dfrac{H_{13} - H_{23}\cos kl_1}{j\rho_0 c_0 \sin kl_1} + MB_{21} + MB_{22}\dfrac{\cos kl_4 - H_{43}}{j\rho_0 c_0 \sin kl_4}\right)}, \tag{25}$$

where $H_{ij}=p_j^* p_i / p_j^* p_j$, with $p_j^* p_i$ being the cross-spectrum between microphones i and j, and $p_j^* p_j$ the auto-spectrum at microphone j. The resistance R and the end correction coefficient α of acoustic impedance obtained by these experiments for a specific hole and given by Lee *et al.* (2006a) for perforated plates are implemented into Eq. (8) for the BEM predictions of the leakage effect. The comparisons will be illustrated later.

The inherent multi-dimensional acoustic field in the immediate vicinity of neck/hole/gap and cavity or main duct has dictated in the present study the use of BEM capable of capturing such spatial variations. Yet, to develop an inherent understanding into the dramatic effect of "parallel" acoustic paths (here through holes and gaps) between the cavity of HRs and the main duct, a simplified lumped analysis has been provided in Appendix. The results for peak TL frequencies from such one-dimensional (1D) approach have then been compared with experiments (as well as BEM for holes) as elaborated within the Appendix for the hole and gap geometries described in the following section.

III. PROTOTYPE

A prototype HR, shown in Fig. 4, has been fabricated, consisting of a main chamber, a cavity volume, a neck, and a baffle that splits the entire chamber into a HR cavity and an expansion chamber. The outer housing is removable so that the number of holes and gaps in the baffle can be controlled by blocking those. Figure 5 provides the details of the baffle designed with various holes and gaps to study the relationship between the open area and TL. In addition to the main duct and HR neck, there are 13 holes (identified by H1–H13) of 0.5 cm diameter and four gaps (identified by G1–G4) of 0.07 cm width. The length of the inside gaps (G1 and G3) are 7.55 cm, while the outside gaps (G2 and G4) are 8.44 cm. The holes are located 15° apart at a radius of 7.22 cm. The inner gaps are located at the same radius as the hole centers. The opening area is 0.196 cm^2 for a hole, 0.529 cm^2 for an inner gap, and 0.591 cm^2 for an outer gap; therefore, gaps have 2.7 (inner) to 3.0 (outer) times larger open area than a single hole. Figure 6 provides the dimensions of the prototype, including the main duct diameter of 4.9 cm; neck diameter and length of 3.6 cm and 8 cm, respectively. The baffle thickness is 0.14 cm.

IV. RESULTS

The TL of the prototype is measured on an impedance tube setup. Figure 7 compares the TL by opening holes gradually from H1 to H13 while keeping all the gaps closed. Holes are opened beginning with H1 until all 13 holes are opened. As the open area is increased, the resonance frequency increases, whereas the magnitude of the TL first decreases sharply followed by an increase. With the given dimensions of the prototype, the resonance frequency of HR is estimated to be about 100 Hz from the classical lumped model. The measured frequency and the magnitude of peak TL are summarized in Table I as a function of hole opening area. In the absence of any leakage, these quantities are measured to be 102 Hz and 29 dB, respectively. With one hole open (H1), the resonance frequency increases by 11% (to 113 Hz), and the magnitude decreases by 43% (to 16.6 dB). As additional holes open, the peak frequency increases (up to 229 Hz with all 13 holes open), while the magnitude decreases first followed by an increase up to about 30 dB.

Figure 8 shows the comparison of TLs with gap openings, while maintaining the holes closed. In general, the TL with gap opening exhibits a phenomenon similar to that of

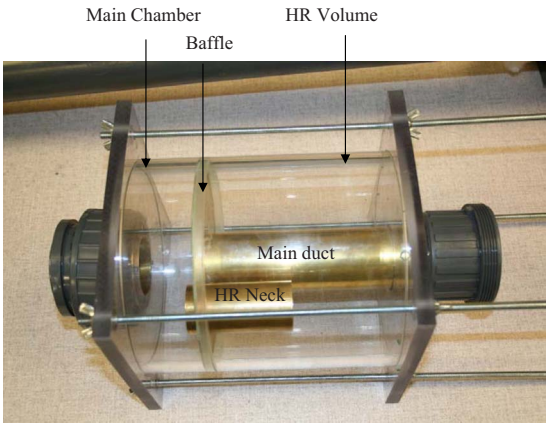


FIG. 4. (Color online) Picture of the prototype.

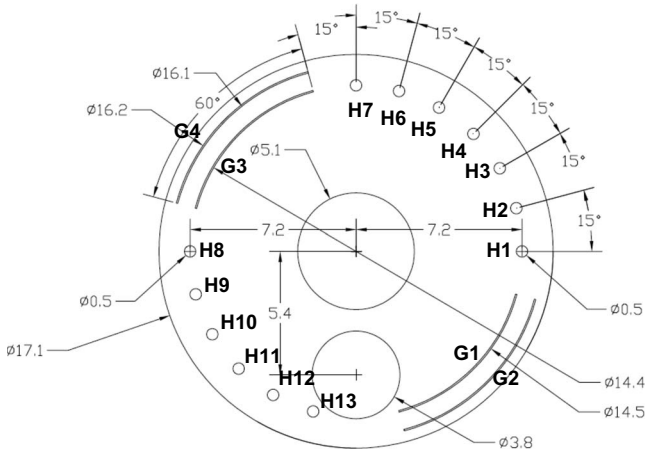


FIG. 5. Schematic of the baffle (unit in cm).

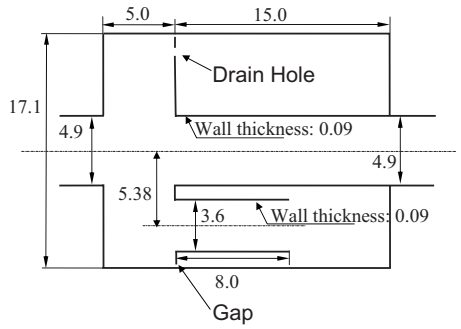


FIG. 6. Schematic of the prototype including HR with leakages (unit in cm).

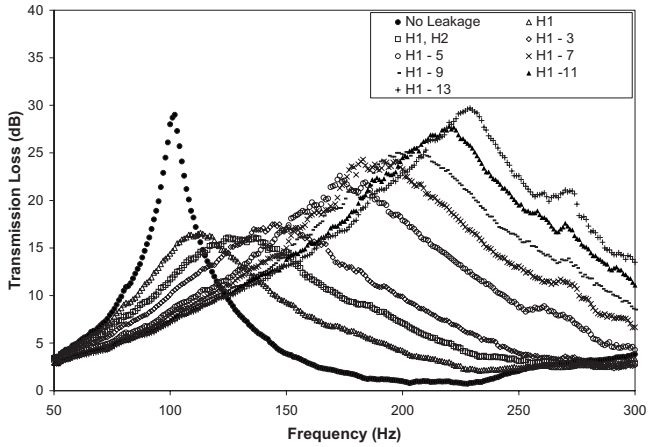


FIG. 7. Experimental results for TL with the prototype (holes open).

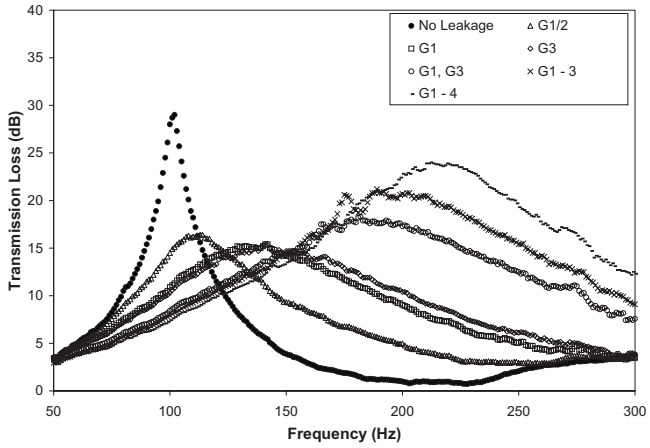


FIG. 8. Experimental results for TL with the prototype (gaps open).

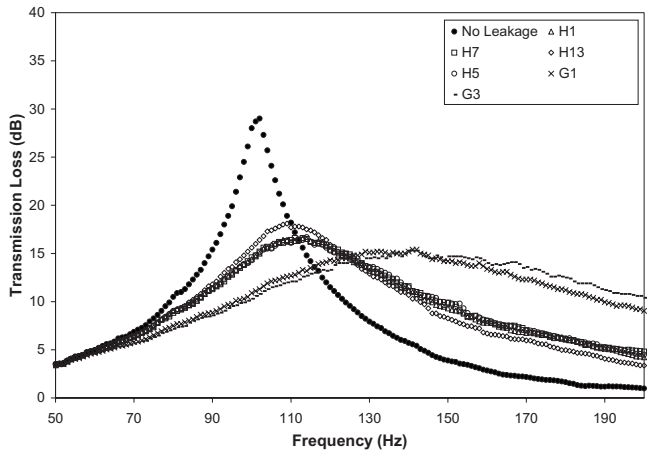


FIG. 9. Experimental results for TL with the prototype (varying hole and gap locations).

TABLE I. The hole openings from H1 to H13 (all gaps are closed).

Hole open	Open area (cm^2)	Peak TL freq. (Hz)	Peak TL mag. (dB)
None	0	102	29
H1	0.196	113	16.6
H1, H2	0.393	131	15.8
H1–3	0.589	147	17.5
H1–4	0.785	167	19.3
H1–5	0.982	177	21.3
H1–6	1.178	181	23.1
H1–7	1.374	189	23.3
H1–8	1.571	195	25
H1–9	1.767	204	25
H1–10	1.963	211	26.5
H1–11	2.160	220	27.8
H1–12	2.356	225	28.4
H1–13	2.553	229	29.7

holes. A half open gap is designated by G1/2, with its peak TL being characterized by 111 Hz and 16.2 dB in frequency and magnitude. Table II summarizes the peak frequency and magnitude as a function of gap opening area. With all gaps open, the peak TL frequency moves up to 217 Hz.

Figure 9 compares the TLs with either a single hole or a gap open with different distances from the center of HR neck; the closest hole to the HR neck is H13 with 2.5 cm distance, and the farthest hole is H7 with 12.5 cm; the closest gap is G1 and one of the farthest gaps is G3. With any open hole, the frequency and magnitude of peak TL are measured, respectively, around 113 Hz and 16.6 dB regardless of the distance except H13 which yields 17.7 dB. With the same amount of gap opening, the attenuation behavior remains similar even though the location of each gap is different. The single gap opening exhibits about 138–140 Hz for the resonance frequency with a corresponding 14.5–15 dB in TL. Table III summarizes the frequency and magnitude of peak TL as a function of the opening location.

Figure 10 compares TLs for the extreme cases of fully open holes and/or gaps, as well as a no-baffle case. As anticipated from earlier results, the peak frequency further increases when hole and gap openings are combined. The frequency of peak TL with no baffle is measured to be 497 Hz with a magnitude of 46.3 dB. In the absence of the baffle, the prototype is no longer a HR as it becomes essentially a quarter wave resonator. For example, using the overall length of 15 cm as a crude estimate (the effective length will be longer because of the 5 cm opening; recall Fig. 6), the quarter wave

TABLE II. The gap openings from G1 to G4 (all holes are closed).

Gap open	Open area (cm^2)	Peak TL freq. (Hz)	Peak TL mag. (dB)
None	0	102	29
G1/2	0.264	111	16.2
G1	0.529	138	15
G1, G3	1.058	179	17.6
G1–3	1.649	196	20.4
G1–4	2.240	217	23.6

TABLE III. Effect of location of hole and gap openings.

Hole open	Gap open	Open area (cm^2)	Peak TL freq. (Hz)	Peak TL mag. (dB)
None	None	0	102	29
H1	None	0.196	113	16.6
H7	None	0.196	112	16.4
H13	None	0.196	110	17.7
H5	None	0.196	112	16.3
None	G1	0.529	138	15
None	G3	0.529	140	14.5

resonance frequency becomes 571 Hz with a speed of sound $c_0=343.2$ m/s (at a room temperature of 20 °C). Hence, if the leakage is substantial, the increasing frequency of peak TL approaches that of the quarter wave resonator. For the no-baffle case, BEM predictions are also superimposed in this figure, showing a reasonable agreement with the experimental results.

The behavior of the frequency and magnitude of peak TL as a function of opening area by holes or gaps is compared in Figs. 11 and 12, respectively. Figure 11 suggests that the frequency of peak TL shifts to higher values with increasing leakage, almost independent of whether the openings are due to holes or gaps. The magnitude of peak TL in Fig. 12, however, exhibits some differences between the two types of openings. With a small amount of leakage, the peak attenuation is reduced similarly for both holes and gaps. With further increase in leakage, however, the holes reveal 3–5 dB higher attenuation than gaps.

Figure 13 shows the experimental results for nondimensional acoustic impedance (resistance and reactance) of a single hole considered in this study as a function of frequency. Sound pressure level has been controlled to maintain the input level the same at the orifice between acoustic impedance measurements and HR-TL experiments with leakage. The resistance and reactance of the hole are then curve-fitted by $\mathrm{Re}(\zeta_h)=R=-0.000\,003f+0.004\,782$ (with 1% porosity) and $\mathrm{Im}(\zeta_h)=0.000\,061f$, respectively. The latter leads, in view of Eq. (18), to an end correction coefficient of $\alpha=0.3859$ for the hole.

Figure 14 compares the TL between experiments and BEM predictions for a single hole open (H1), as well as the baseline with no leakage. The resonance of the latter is identical at 102 Hz between the experiment and BEM, whereas a discrepancy in peak TL magnitude is observed, presumably due to the acoustic impedance of the neck which BEM has treated as a simple opening (pressure difference is zero at the neck-cavity interface, $\zeta_{A3}=0$). When the hole is introduced with experimentally obtained $R=-0.000\,003f+0.004\,782$ and $\alpha=0.3859$, BEM exhibits an increase in frequency and a sharp drop in the magnitude of peak TL, in agreement with the experimental results. While the trend of increasing frequency and decreasing magnitude of TL peak is well captured in these predictions, the attenuation levels remain about 5 dB higher than the experiments for the reasons associated with the baseline HR. The acoustic impedance of perforated holes as suggested by Lee *et al.* (2006a), $R=0.004\,395, 0.005\,013$ and $\alpha=0.2471, 0.4473$ corresponding

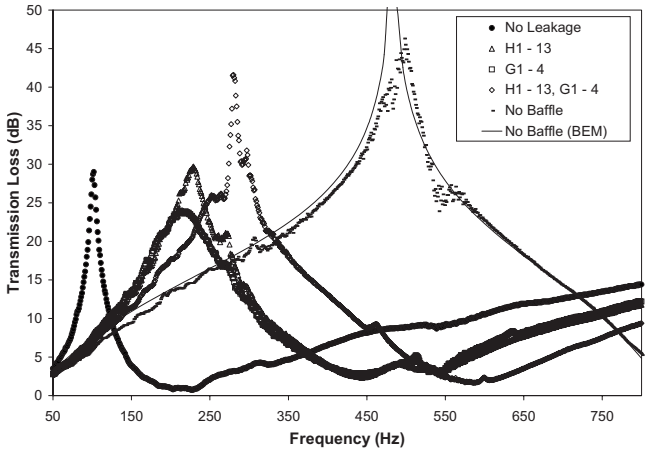


FIG. 10. Experimental results for TL with the prototype (holes and/or gaps fully open).

to $\phi=13.6\%$ and 25.2%, respectively, are also used with BEM to assess its applicability to a single hole considered here. It may be observed that R and α determined experimentally in the present study for a single hole provides a better match for the resonance frequency and magnitude near peak TL. Since TLs with acoustic impedances of (Lee *et al.*, 2006a) show similar behavior at both porosities, comparisons hereafter will use $\phi=25.2\%$ only.

Two-hole (H1 and H8) open case is investigated next through BEM predictions as depicted in Fig. 15, along with single hole open results as superimposed from Fig. 14. Similar to the single hole open case, the experimentally obtained acoustic impedance shows a more reasonable match. Finally, the results are contrasted with the TL predictions employing zero acoustic impedance ($\zeta_{A4}=0$) for the two holes. While this approach shows a frequency shift qualitatively in the right direction, neither the exact location nor the magnitude near peak TL agrees with the experimental results, rendering such a simplification inappropriate for the geometry considered here.

V. CONCLUDING REMARKS

As pointed out in the Introduction, in the commercial manufacture of mufflers, there are often holes in the baffles

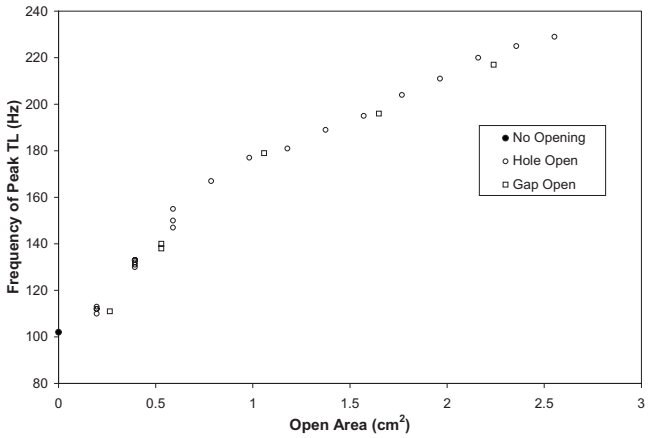


FIG. 11. Relationship between peak TL frequency and open area on the baffle; solid circle represents no leakage case.

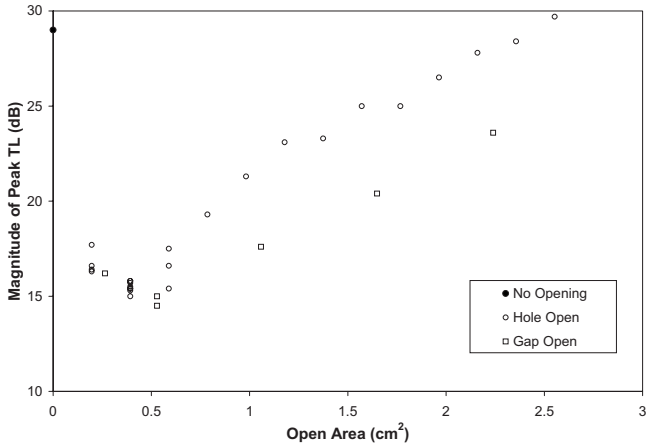


FIG. 12. Relationship between peak TL magnitude and open area on the baffle; solid circle represents no leakage case.

and gaps between the baffles and shell. When a baffle constitutes one of the chamber boundaries of a HR, the acoustic performance of the HR can be significantly impacted. A prototype HR has been built to study these phenomena with holes and gaps being intentionally implemented on the baffle to produce leakage. The TL is measured by gradually opening these holes and/or gaps on the baffle. The resonance frequency is observed to shift to higher values with increasing leakage. Peak TL is found first to drop drastically with small leakages, followed by an increase with further opening. At large openings or in the ultimate limit of no baffle in the prototype silencer, the TL behavior approaches that of a quarter wave resonator. The shift in resonance frequency with opening area appears to be independent of the type of leakage (hole or gap). The magnitude of peak TL reveals a similar behavior only at the small openings, whereas, at larger openings, holes yield higher TL than gaps presumably due to different acoustic characteristics of open areas.

TL predictions have been performed to capture the effect of leakage by BEM which implements a measured acoustic impedance of the hole. The results are also compared with those obtained by using an acoustic impedance available in the perforated plate literature. While the latter captures the trends reasonably well, the former acoustic impedance determined here for a hole experimentally leads to a more accurate prediction of the frequency dependence of TL. Some discrepancy remains between the predictions and experiments in terms of magnitudes near peak TL, due possibly to non-negligible acoustic impedance of the neck interfaces, an aspect that requires further study in future.

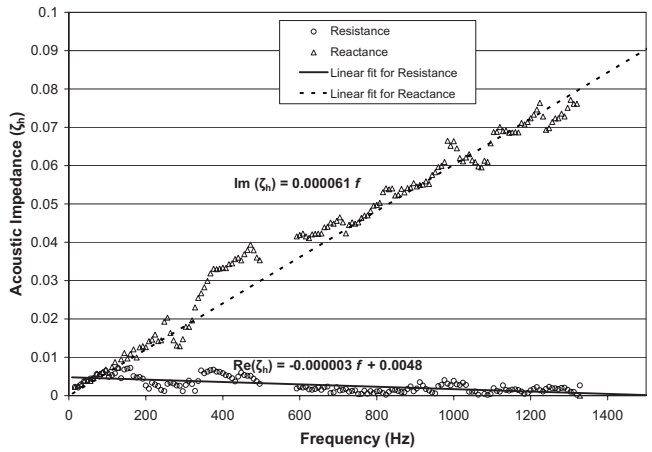


FIG. 13. Measured acoustic impedance (resistance and reactance) of a hole.

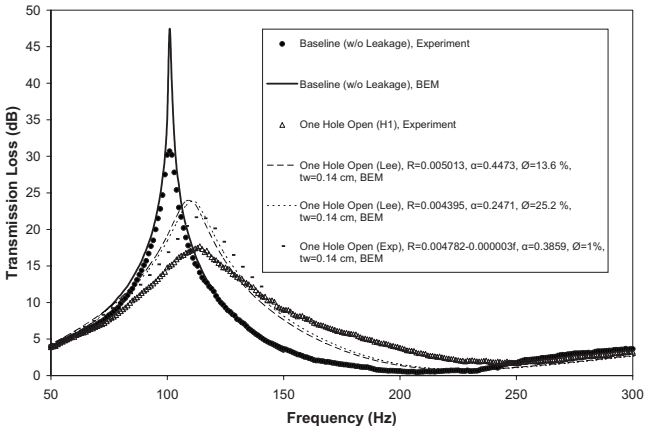


FIG. 14. Comparison between experimental results and BEM predictions for leakage model (one hole open).

APPENDIX

A leakage in HR is modeled, as shown in Fig. 16, through an additional short neck. HR cavity volume and main duct are thus connected through both the HR neck and a small opening that provides an additional path of interaction besides the neck.

The acoustic pressure and velocity at each location i of this configuration may be expressed by

$$p_i = A_i e^{j(\omega t - kx)} + B_i e^{j(\omega t + kx)}, \tag{A1}$$

$$u_i = \frac{1}{\rho_0 c_0}(A_i e^{j(\omega t - kx)} - B_i e^{j(\omega t + kx)}). \tag{A2}$$

Boundary conditions at location I defined by pressure and volumetric velocity equalities may be written as

$$p_1 = p_2 = p_5 = p_6, \tag{A3}$$

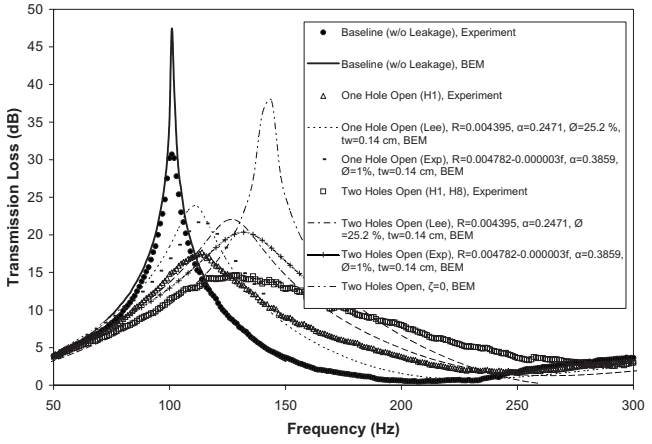


FIG. 15. Comparison between experimental results and BEM predictions for leakage model (two holes open).

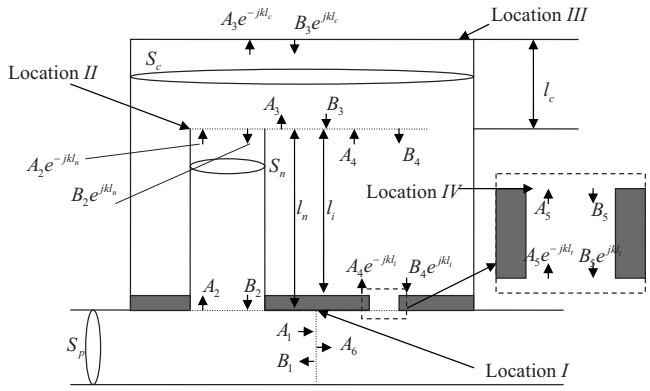


FIG. 16. The schematic of HR with leakage for 1D analysis.

$$S_p u_1 = S_n u_2 + S_t u_5 + S_p u_6, \tag{A4}$$

where S_p and S_n are the cross-sectional areas of main duct and neck, respectively. Similarly, other boundary conditions yield

$$p_2 = p_3 = p_4, \tag{A5}$$

$$S_n u_2 = S_c u_3 + m_1 u_4, \tag{A6}$$

$$u_3 = 0 \quad \text{at location III (wall)}, \tag{A7}$$

$$p_4 = p_5, \tag{A8}$$

$$m_1 u_4 = S_t u_5, \tag{A9}$$

where S_c is the cross-sectional area of cavity, $m_1 = S_c - S_n$, and S_t is the cross-sectional area of the hole.

The algebraic manipulation of the foregoing ten equations [within Eqs. (A3)–(A9)] leads to the following relationships for 11 unknowns (A_i, $i=1-6$ and B_i, $i=1-5$) expressed in the matrix format as

$$\begin{Bmatrix} A_1 \\ B_1 \\ A_2 \\ B_2 \\ A_3 \\ B_3 \\ A_4 \\ B_4 \\ A_5 \\ B_5 \end{Bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & \frac{S_n}{S_p} & -\frac{S_n}{S_p} & 0 & 0 & 0 & 0 & \frac{S_t}{S_p}e^{-jkl_t} & -\frac{S_t}{S_p}e^{jkl_t} \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & -e^{-jkl_t} & -e^{jkl_t} \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -e^{2jkl_c} & 0 & 0 & 0 & 0 \\ 0 & 0 & -e^{-jkl_n} & e^{jkl_n} & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 0 & \frac{S_n}{m_1}e^{-jkl_n} & -\frac{S_n}{m_1}e^{jkl_n} & -\frac{S_c}{m_1} & \frac{S_c}{m_1} & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -e^{-jkl_i} & -e^{jkl_i} & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{m_1}{S_t}e^{-jkl_i} & -\frac{m_1}{S_t}e^{jkl_i} & -1 & 1 \end{bmatrix}^{-1} \begin{Bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} A_6, \tag{A10}$$

or

$$\begin{Bmatrix} A_1 \\ B_1 \\ A_2 \\ B_2 \\ A_3 \\ B_3 \\ A_4 \\ B_4 \\ A_5 \\ B_5 \end{Bmatrix} = [TM] \begin{Bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix} A_6. \tag{A11}$$

The relationship between A_1 and A_6 may then be rearranged as

$$\frac{A_6}{A_1} = \frac{1}{TM_{11} - TM_{12} + TM_{14}}, \tag{A12}$$

where TM_{ij} are the matrix components at ith row and jth column. Finally, the TL of the HR with the leakage is determined from

$$\mathrm{TL} = 20 \log_{10} \left| \frac{A_6}{A_1} \right| = 20 \log_{10} \left| \frac{1}{TM_{11} - TM_{12} - TM_{14}} \right|. \tag{A13}$$

The peak TL frequencies with leakage from both experimental results and the foregoing simplified 1D analysis are shown in Tables IV (for holes) and V (for gaps). The following approximate end correction coefficients (α) [see, for example, Selamet and Ji (2001)] are implemented into 1D analysis for the neck and leakage opening area: $\alpha = 0.8217$

TABLE IV. The peak TL frequencies with hole opening.

		Experimental results	1D analysis	3D BEM
Hole open	Open area (cm^2)	Peak TL freq. (Hz)	Peak TL Freq. (Hz)	Peak TL freq. (Hz)
None	0	102	101	102
H1	0.196	113	121	112
H1,H2	0.393	131	137	132
H1–3	0.589	147	151	148
H1–5	0.982	177	177	176

for neck to main duct interface and both sides of leakage opening area, and $\alpha=0.6133$ for neck to HR cavity interface. Without leakage, the peak TL frequency is 101 Hz from 1D analysis which matches well the experimental result. With the introduction of leakage, the peak TL frequencies increase markedly even though the amount of opening area is small. For example, the approximate lumped analysis here suggests that a single hole (H1) with an open area of 0.196 cm^2 (a 1.93% increase over the original neck area of $\pi(3.6)^2/4=10.179$ cm^2, $0.196/10.179=0.0193$) raises the resonance frequency by 20% instead of

$$\left(\sqrt{\frac{10.179+0.196}{10.179}}-1\right)100 \cong \frac{1.93}{2} \cong 1\ \%\ ,$$

which would have been predicted by simply adding two areas in the expression for resonance frequency $f=(c/2\pi)\sqrt{S/VL}$ from lumped analysis, where c is the speed of sound, S is the neck cross-sectional area, V is the volume of the HR cavity, and L is the neck length. The reasonable comparison of trends in experimental vs 1D analysis results presented here demonstrates the dramatic impact of the additional small passage, which could not be predicted by simply enlarging the area of the original neck. However, it should also be noted that the 1D analysis performed here is not precise (particularly in representing the inertia of short necks) since it does not capture the multi-dimensional behavior of acoustics (other than through approximate end corrections). The clear agreement of BEM results also included in Table IV with experiments, on the other hand, illustrate the need for such a multi-dimensional approach for accurate predictions.

TABLE V. The peak TL frequencies with gap opening.

		Experimental results	1D analysis
Gap open	Open area (cm^2)	Peak TL freq. (Hz)	Peak TL Freq. (Hz)
None	0	102	101
G1	0.529	138	133
G1,G3	1.058	179	159
G1–3	1.649	196	179
G1–4	2.240	217	199

Alster, M. (**1972**). "Improved calculation of resonant frequencies of Helmholtz resonator," J. Sound Vib. **24**, 63–85.

Bies, D. A., and Wilson, O. B. (**1957**). "Acoustic impedance of Helmholtz resonator at very high amplitude," J. Acoust. Soc. Am. **29**, 711–714.

Bolt, R. H., Labate, S., and Ingard, U. (**1949**). "The acoustic reactance of small circular orifices," J. Acoust. Soc. Am. **21**, 94–97.

Bemman, Y. J., Frei, T., Jones, C., and Keck, Mathias. (**2005**). "Passive exhaust system with cylinder deactivation," SAE Paper No. 2005-01-2351.

Chanaud, R. C. (**1997**). "Effects of geometry on the resonance frequency of Helmholtz resonators, Part II," J. Sound Vib. **204**, 829–834.

Dickey, N. S., and Selamet, A. (**1998**). "Acoustic nonlinearity of a circular orifice: An experimental study of the instantaneous pressure/flow relationship," Noise Control Eng. J. **46**, 97–107.

Goldman, A., and Chung, C. H. (**1982**). "Impedance of an orifice under a turbulent boundary layer with pressure gradient," J. Acoust. Soc. Am. **71**, 573–579.

Ingard, U. (**1953**). "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**, 1037–1061.

Ingard, U., and Labate, S. (**1950**). "Acoustic circulation effects and the nonlinear impedance of orifices," J. Acoust. Soc. Am. **22**, 211–218.

Lee, I.-J., Selamet, A., and Huff, N. T. (**2006a**). "Acoustic impedance of perforations in contact with fibrous material," J. Acoust. Soc. Am. **119**, 2785–2797.

Lee, I.-J., Selamet, A., and Huff, N. T. (**2006b**). "Impact of perforation impedance on the transmission loss of reactive and dissipative silencers," J. Acoust. Soc. Am. **120**, 3706–3713.

Melling, T. H. (**1973**). "The acoustic impedance of perforates at medium and high sound pressure levels," J. Sound Vib. **29**, 1–65.

Selamet, A., and Ji, Z. L. (**2000**). "Circular asymmetric Helmholtz resonators," J. Acoust. Soc. Am. **107**, 2360–2369.

Selamet, A., and Ji, Z. L. (**2001**). "Wave reflection from duct terminations," J. Acoust. Soc. Am. **109**, 1304–1311.

Selamet, A., and Lee, I. J. (**2003**). "Helmholtz resonator with extended neck," J. Acoust. Soc. Am. **113**, 1975–1985.

Selamet, A., Radavich, P. M., Dickey, N. S., and Novak, J. M. (**1997**). "Circular concentric Helmholtz resonator," J. Acoust. Soc. Am. **101**, 41–51.

Selamet, A., Xu, M. B., Lee, I.-J., and Huff, N. T. (**2005**). "Helmholtz resonator lined with absorbing material," J. Acoust. Soc. Am. **117**, 725–733.

Sivian, L. J. (**1935**). "Acoustic impedance of small orifices," J. Acoust. Soc. Am. **7**, 94–101.

Sullivan, J. W., and Crocker, M. J. (**1978**). "Analysis of concentric-tube resonators having unpartitioned cavities," J. Acoust. Soc. Am. **64**, 207–215.

Bicylindrical model of Herschel–Quincke tube-duct system: Theory and comparison with experiment and finite element method

B. Poirier, J. M. Ville,[a] and C. Maury
Laboratoire Roberval, UMR UTC-CNRS No. 6253, Université de Technologie de Compiègne, BP 20529 F60205, Compiègne Cedex, France

D. Kateb
Laboratoire de Mathématiques Appliquées de Compiègne, Université de Technologie de Compiègne, BP 20529 F60205, Compiègne Cedex, France

(Received 22 April 2008; revised 3 June 2009; accepted 5 June 2009)

An analytical three dimensional bicylindrical model is developed in order to take into account the effects of the saddle-shaped area for the interface of a *n*-Herschel–Quincke tube system with the main duct. Results for the scattering matrix of this system deduced from this model are compared, in the plane wave frequency domain, versus experimental and numerical data and a one dimensional model with and without tube length correction. The results are performed with a two-Herschel–Quincke tube configuration having the same diameter as the main duct. In spite of strong assumptions on the acoustic continuity conditions at the interfaces, this model is shown to improve the nonperiodic amplitude variations and the frequency localization of the minima of the transmission and reflection coefficients with respect to one dimensional model with length correction and a three dimensional model.
 [DOI: 10.1121/1.3159370]

PACS number(s): 43.50.Gf, 43.20.Mv, 43.50.Lj [JWP] Pages: 1151–1162

I. INTRODUCTION

In many domains of application such as aircraft engine, building ventilation, and internal combustion engine (IC-engine) exhaust and intake automotive systems, mufflers are used to reduce noise radiated outside by sources located in ducts. Research is still conducted to improve passive or active technologies and also to find out other concepts such as the Herschel–Quincke (HQ) tube system. Indeed, this HQ tube system concept,[1] which consists of inter-connecting several cylindrical pipes, has already been shown to have potential to reduce the broadband and the blade pass frequency noise radiated by a full scale production aircraft engine.[2] Moreover Griffin *et al.*[3] made a theoretical study of an actively controlled membrane introduced in the longer of the two ducts that allowed the characteristics of the system to adapt to changes in the incoming disturbance. In automotive, Hwang *et al.*[4] implemented this concept in exhaust systems using two HQ tubes in series and actively changing the length of the two interference paths in order to adapt the attenuation to follow the harmonics of the engine with variations of engine speed. Trochon,[5] and later McLean,[6] reported on the use of quarter-wave resonators at the nodal points of the HQ tube arrangement, thereby achieving a broader and smoother attenuation curve for turbo- and intake noise in internal combustion engines. If, for the aeronautical and building applications, the tube to duct diameter ratio is low, in car industry, applications deal with problems where the HQ tube diameters are of the same dimensions as the main duct. To predict or optimize the efficiency of such HQ tube system, a three dimensional (3D) analytical modeling based on the theory of derivative tubes is needed. But as pointed out by previous works on the branched tube, which is a closed derivation, two main difficulties occur.

(1) Even for low frequencies when the sound field in the main pipe and the side-branch is one dimensional (1D), near the junction it becomes 3D.
(2) When both the main duct and the branched tube are cylindrical, the surface at the junction obviously is saddle-shaped.

Indeed, Green's function integral technique analysis based on a two dimensional (2D) 45° and 90° bifurcation ducts[7] has shown deviations of the resonance frequency locations compared with those predicted by the 1D model of a straight duct. In particular, the region around the branch is shown to be highly 2D even at low frequency. A work[8] based on a 3D mode coupling theory applied to a squared cross-sectional duct and branch pointed out that higher order modes can propagate in the branch even if a planar pressure distribution is assumed in the main duct. Then, a modal decomposition method[9] was developed to obtain an exact result both in a matrix form and a variational formulation. It describes the insertion of a branched tube on a main waveguide in which only the planar mode propagates. In case of cylindrical intersection, new formulas were compared with results obtained with the boundary element method. The modal decomposition method has shown to be not well adapted because the matching of the modes cannot be applied over a

[a]Author to whom correspondence should be addressed. Electronic mail: jean-michel.ville@utc.fr

saddle-shaped junction, except in the limit case where the area of the branched tube is small compared to the area of the duct. Moreover, discretization methods lead to large uncertainties for small-diameter branched tubes, because of large discontinuities. Therefore, in order to take into account the effect of the evanescent modes at the interface, an adequate acoustic length correction to the tube length based on a 3D boundary element approach has been applied in 1D modeling.[10] A good agreement was found with the correction calculated for the limit case where the branch radius is much smaller than the main duct radius.

Assuming pressure equality and conservation of the volumetric flow, a 1D model[11] was developed to calculate the transmission loss of a system with one HQ tube system. The main motivation of the model was to improve early works done in 1928.[12] A good agreement was found with the experiment and a 1D finite-difference scheme valid at very low adimensional frequencies $ka<0.7$ (with a radius of the main duct), but introducing an "effective" tube length, which is supposed to take into account the curved shape of the intersection between the cylindrical duct and the tube. In order to be applied to real aircraft engine conditions, a 3D model for HQ tubes' derivations was recently developed.[13,14] This 3D modeling technique considers the tubes-inlet interfaces as finite piston sources that couple the acoustic field inside a hard-walled duct with the acoustic field within the HQ tubes for higher order modes' propagation conditions. This model represents an improvement compared with 1D model[11] since it emphasizes the scattering mode effects inside the duct due to the presence of HQ tubes.

The aim of the paper is to improve the 3D modeling[14] for the propagation in a main duct connected with HQ tubes. Indeed, the authors take into account the real duct-tube interface by using a bicylindrical technique. Results with Green's formalism for wave propagation modeling that accounts for the exact saddle shape of the inter-connecting surface are compared with numerical and experimental results for frequencies below the first duct cut-on frequency. The scattering matrix formalism is used to characterize the acoustical properties of the system made of two HQ tubes with the same diameter as the main duct system.

The paper is organized as follows. The authors begin in Sec. II to describe the theoretical basis of the propagation model. After, in Sec. III, the experimental set-up and the finite element method (FEM) are used to validate the model. In Sec. IV the results obtained from both the analytical approaches and the FEM are compared with the experiment. Also the validity of the bicylindrical technique is discussed. The study is concluded with final remarks in Sec. V.

II. THEORETICAL BASIS

A. Basic integral equation

Consider an infinite cylindrical rigid-walled duct of radius a connected with N_{HQ} cylindrical HQ tubes of radius d located circumferentially around the duct (Fig. 1 with $N_{HQ}=1$). The pressure $p(w)$ at any point with coordinates $w=(r,\theta,z)$ in the duct is written as the sum of the incident pressure $p^{inc}(w)$ assumed to propagate toward the positive

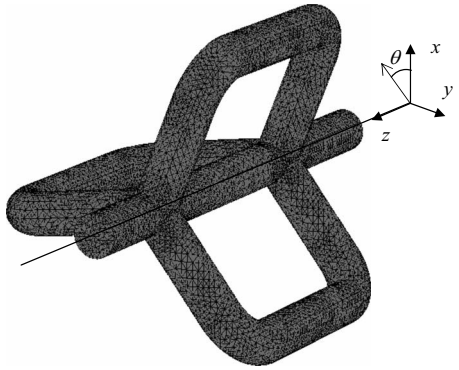


FIG. 1. The main duct with 3-HQ tubes.

z-direction in the straight duct without the HQ arrays and an integral over S, a closed surface within the infinite main duct, according to the Helmholtz–Huyghens equation

$$p(w)=p^{inc}(w)+\int\int G^{\pm}(w,w')\partial_n p(w') - p(w')\partial_n G^{\pm}(w,w')dS, \quad (1)$$

where w' belongs to the surface S, ∂_n is the outward normal derivative to the surface, and G is Green's function for the main duct. The following assumptions simplify this equation by restricting the integration over the surface at the interfaces between the main duct and the branched tubes. The integrals over both ends of the infinite main waveguide, which represent the pressure generated on these surfaces, vanish when Sommerfeld's condition is applied. Moreover, the authors choose Green's function for the duct to satisfy Neumann's conditions: Its normal derivative vanishes on the surface of the waveguide. The normal derivative of the pressure vanishes also on the hard walls of main duct. Therefore the integral in Eq. (1) is limited to the surface $S_{int}=\Sigma_{u=1}^{2N_{HQ}}S_u$ made up of all the interfaces between the duct and the tubes, S_u being the elementary area of the main duct, which communicates with an input or output of a HQ tube (Fig. 2). Neglecting the dissipation and according to Euler's equation

$$p(w)=p^{inc}(w)-j\omega\rho\sum_{u=1}^{2N_{HQ}}\int_{S_u}G^{\pm}(w,w')v_u(w')dS, \quad (2)$$

where v_u is the outward particle acoustic velocity normal to S_u, ρ is the fluid density, and ω is the pulsation. If the un-

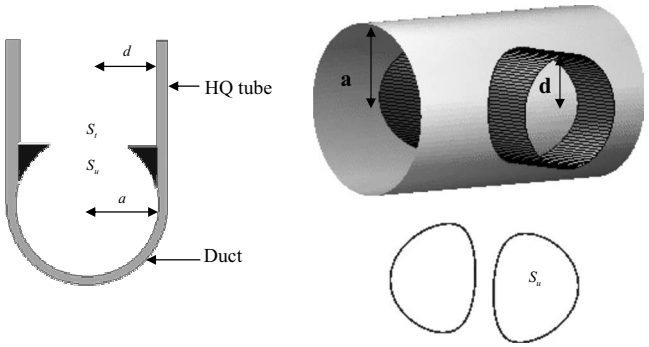


FIG. 2. Bicylindrical geometry: S_t the HQ tube area perpendicular to the duct and S_u the bicylindrical intersection of the duct with the HQ tube.

knowns of the problem $\bar{v}_u$, $u=1,\ldots,2N_{\mathrm{HQ}}$, are assumed to be constant on all surfaces S_u, then Eq. (2) becomes

$$p(w)=p^{\mathrm{inc}}(w)-j\omega\rho\sum_{u=1}^{2N_{\mathrm{HQ}}}\bar{v}_u\int_{S_u}G^{\pm}(w,w')dS. \tag{3}$$

Equation (3) shows that to determine the pressure in the main duct with the N_{HQ} tubes, the incident pressure $p^{\mathrm{inc}}(w)$, the integral $\int_{S_u}G^{\pm}(w,w')dS$, and $\bar{v}_u$ have to be computed.

Green's function of the main duct, which is the pressure field generated at $w=(r,\theta,z)$ by a point source located at $w'=(a,\theta',z')$, is solution of the following differential equation:

$$\frac{\partial^2 G}{\partial r^2}+\frac{1}{r}\frac{\partial G}{\partial r}+\frac{1}{r^2}\frac{\partial^2 G}{\partial \theta^2}+\frac{\partial^2 G}{\partial z^2}+k^2G=\frac{1}{r}\delta(r-a)\delta(\theta-\theta')\delta(z-z'), \tag{4}$$

where δ is Dirac's delta function, $k=\omega/c$ is the wave number, and c is the speed of sound. Green's function is represented on the basis of the rigid-walled cylindrical duct eigenvalues $\Phi_{mn}(r,\theta)=J_m(\chi_{mn}r/a)e^{-im\theta}$ where J_m is the Bessel function of the first kind and order m and χ_{mn} the nth derivative of J_{mn} (Ref. 15)

$$G^{\pm}(w|w')=\frac{i}{\pi a^2}\sum_{m=-M_g}^{M_g}\sum_{n=0}^{N_g}\frac{\Phi_{mn}(r,\theta)\Phi_{mn}(a,\theta')}{N_{mn}(k_{mn}^{+}-k_{mn}^{-})}\times e^{-ik_{mn}^{\pm}(z-z')}, \tag{5}$$

where $k_{mn}^{\pm}=\pm\sqrt{k^2-\chi_{mn}^2/a^2}$ are the axial wave numbers of the mode (m,n), which propagates, respectively, toward the positive $(z>z')$ and negative $(z<z')$ z-directions, and M_g and N_g being the number of azimuthal and radial modes. The normalization factor N_{mn} is defined by

$$N_{mn}=\begin{cases}J_m^2(\chi_{mn}) & \text{if } m=0\\ \frac{1}{2}\left[1-\frac{m^2}{(\chi_{mn})^2}\right] & \text{if } m\neq 0.\end{cases} \tag{6}$$

The term $-j\omega\rho\bar{v}_u\int_{S_u}G^{\pm}(w,w')dS$ represents the pressure radiated at $w=(r,\theta,z)$ by an input or output HQ tube area located at $w_u=(a,\theta_u,z_u)$. In a 3D model,[14] v_u is supposed to be constant over the surface area S_u assumed to be equal to S_t, the HQ tube cross-sectional area. In the present study, the exact value of the interface area S_u between two cylinders will be used to calculate the integrals. Both approaches are similar only when the branched tube radius d is much smaller than a the duct radius. Indeed, the saddle-shaped area S_u is evaluated by a bicylindrical technique and is computed as an elliptic integral of the second kind (see Appendix A)

$$S_u=4a\int_0^d\sqrt{\frac{d^2-z^2}{a^2-z^2}}dz=4ad\int_0^{d/a}\sqrt{\frac{1-(a/d)^2t^2}{1-t^2}}dt,$$

$$t=z/a. \tag{7}$$

The relative difference δS defined by $\delta S=100\times|(S_u-S_t)/S_u|$ calculated for the radius of the main duct $a=25$ mm and plotted in Fig. 3 versus d the radius of the tube

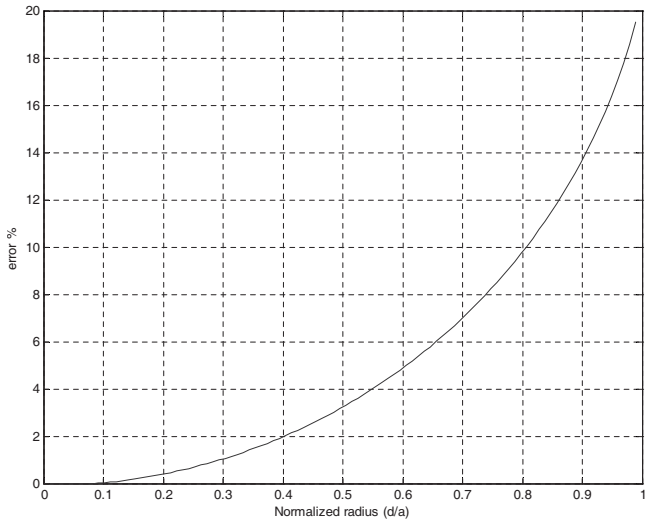


FIG. 3. Relative difference δS in area between the duct and the tube at the interface versus the normalized radius d/a.

shows that the real geometry of the interface cannot be approximated when d is not small compared to a. Indeed, for example, if $d=20$ mm, the error becomes higher than 10%. Previous work[9] introduced the specificity of this saddle-shaped intersection and pointed out also the difficulty to write the acoustic continuity conditions on S_u, which have to be theoretically the same form in the main duct and in the tube.

B. Modeling technique

From Eq. (3), the pressure in the main duct with the N_{HQ} tubes is the addition of the incident pressure, which propagates toward the positive z and the pressure, which is produced in both directions by the $2N_{\mathrm{HQ}}$ sources. To solve the problem, especially to determine $\bar{v}_u$, coupling conditions on the pressures and normal acoustic particle velocities between both ends of each HQ tube are written.

1. The incident pressure

The incident pressure $p^{\mathrm{inc}}(w)$ in the infinite hard wall cylindrical duct can be expressed as the sum of a set of modes of circumferential order m and radial order n given by

$$p^{\mathrm{inc}}(w)=\sum_{m=-M_g}^{M_g}\sum_{n=0}^{N_g}\Gamma_{mn}\Phi_{mn}(r,\theta)e^{-ik_{mn}^{+}z}, \tag{8}$$

where Γ_{mn} is the complex amplitude of mode (m,n). The sound field is assumed to consist of plane wave only and $(m,n)=(0,0)$.

2. Determination of $\bar{v}_u$

a. The relationship between $\bar{v}_u$ and $\bar{p}_u$ the average pressure on S_u The acoustic particle velocity $\bar{v}_u$ considered as an average value on S_u is related to the average value of the pressure $\bar{p}_u$ on S_u through the relationship that can be expressed in matrix form as follows:

$$\{\bar{p}_u\}_{2N_{HQ}} = [Z_{rs}]_{2N_{HQ}\times 2N_{HQ}}\{\bar{v}_u\}_{2N_{HQ}} + \{\bar{p}_u^{inc}\}_{2N_{HQ}},$$
$$u = 1{:}2N_{HQ}. \quad (9)$$

Indeed, $\bar{p}_u$ is the addition of $\bar{p}_u^{inc}$ the average incident pressure over S_u due to the incident pressure and the average pressure produced by the HQ sources and defined by $[Z_{rs}]_{2N_{HQ}\times 2N_{HQ}}\{\bar{v}_u\}_{2N_{HQ}}$. Z_{rs} is an impedance matrix that relates the average pressure on a "receiving" source due to an "emission" source with unit velocity.

Determination of $\bar{p}_u^{inc}$

The average pressure over S_u, the area source located at $w_u=(a,\theta_u,z_u)$ due to the incident pressure, is given by

$$\bar{p}_u^{inc}(w_u) = \frac{1}{S_u}\int_{S_u} p^{inc}(a,\theta,z)dS. \quad (10)$$

If N denotes the number of modes, then Eq. (10) can also be written under a matrix formulation

$$\bar{p}_u^{inc}(w_u) = \boldsymbol{\beta}_u^+\boldsymbol{\Gamma}. \quad (11)$$

where the modal incident vector $\boldsymbol{\Gamma}$ is defined as $\boldsymbol{\Gamma}=\{\Gamma_{00}\cdots\Gamma_N\}^T$ and $\boldsymbol{\beta}_u^+=\{\beta_{u,1}^+\cdots\beta_{u,N}^+\}$. The computation of $\beta_{u,mn}^+=J_{mn}(\chi_{mn})1/S_u\int_{S_u}e^{-im\theta}e^{-ik_{mn}^+z}dS$ in the bicylindrical form is detailed in Appendixes B 1 and C 1. The incident average pressure vector over all tube-duct interfaces can also be written as

$$\{\Gamma_u\}_{2N_{HQ}} = \{\bar{p}^{inc}(w_1),\bar{p}^{inc}(w_2),\ldots,\bar{p}^{inc}(w_{2N_{HQ}})\}^T$$
$$= \{\boldsymbol{\beta}_1^+\boldsymbol{\Gamma},\boldsymbol{\beta}_2^+\boldsymbol{\Gamma},\ldots,\beta_{2N_{HQ}}^+\boldsymbol{\Gamma})\}^T. \quad (12)$$

Determination of Z_{rs}

$Z_{u,rs}$ is defined as the ratio between $\bar{p}_{ur}$, the average pressure radiated by the source area S_{us} located at w_{us} on S_{ur} the area of the receiving surface, and $\bar{v}_{us}$ the average particle velocity on S_{us}. Therefore it can be expressed as

$$Z_{u,rs} = \frac{1}{v_{us}S_{ur}}\int_{S_{ur}} p(w,w_{us})dS, \quad (13)$$

where, as shown in Eq. (3), the pressure in the main duct radiated by a source area S_{us} located at $w_{us}=(a,\theta_s,z_s)$ is given in terms of the infinite rigid-walled duct Green's function

$$p(w,w_{us}) = -i\omega\rho\bar{v}_{us}\int_{S_{us}} G^{\pm}(w,w')dS. \quad (14)$$

As S_{us} is a saddle-shaped area, Eq. (14) is written in the bicylindrical form leading to the following integral as detailed in Appendix B 2:

$$p(w,w_{us}) = -i\omega\rho v_{us}\int_{y'=-d}^{y'=+d}\int_{z'=z_s-\sqrt{d^2-y'^2}}^{z'=z_s+\sqrt{d^2-y'^2}} \times G^{\pm}(w,w')C(y')dz'dy', \quad (15)$$

where $C(y)=a/\sqrt{a^2-y^2}$. The calculation of Z_{rs} from Eq. (13) depends on the axial location z_r of the receiving area relative to the axial source area position z_s. For the $u=1{:}2N_{HQ}$ sources, three cases are considered with $L_{mn}=J_m(\chi_{mn}r/a)J_m(\chi_{mn})/N_{mn}(k_{mn}^+-k_{mn}^-)$:

- For $z_r>z_s$,

$$p(w|w_{us}) = \frac{\bar{v}_{us}k\rho c}{\pi a^2}\sum_m\sum_n L_{mn}\int_{y'=-d}^{y'=+d}\int_{z'=z_s-\sqrt{d^2-y'^2}}^{z'=z_s+\sqrt{d^2-y'^2}} \times C(y')e^{-ik_{mn}^+(z-z')}e^{-im(\theta-\theta')}dy'dz', \quad (16)$$

where $\theta=\arcsin(y/a)$ and $\theta'=\arcsin(y'/a)$, and then from Eq. (13) where w' is the coordinates of a point that belongs to S_{ur},

$$Z_{u,rs} = \frac{1}{\bar{v}_{us}S_{ur}}\int_{y'=-d}^{y'=+d}\int_{z'=z_r-\sqrt{d^2-y'^2}}^{z'=z_r+\sqrt{d^2-y'^2}} \times C(y')p(w'|w_{us})dz'dy'. \quad (17)$$

- For $z_r<z_s$,

$$p(w,w_{us}) = \frac{\bar{v}_{us}k_0\rho c}{\pi a^2}\sum_m\sum_n L_{mn}\int_{y'=-d}^{y'=+d}\int_{z'=z_s-\sqrt{d^2-y'^2}}^{z'=z_s+\sqrt{d^2-y'^2}} \times C(y')e^{-ik_{mn}^-(z-z')}e^{-im(\theta-\theta')}dy'dz', \quad (18)$$

and then from Eq. (13) where w' is the coordinates of a point that belongs to S_{ur},

$$Z_{u,rs} = \frac{1}{\bar{v}_{us}S_{ur}}\int_{y'=-d}^{y'=+d}\int_{z'=z_r-\sqrt{d^2-y'^2}}^{z'=z_r+\sqrt{d^2-y'^2}} \times C(y')p(w'|w_{us})dz'dy', \quad (19)$$

- For $z_r=z_s$,

$$p(w|w_{us}) = \frac{\bar{v}_{us}k\rho c}{\pi a^2}\sum_m\sum_n L_{mn} \times\left[\int_{y'=-d}^{y'=+d}\int_{z'=z_s-\sqrt{d^2-y'^2}}^{z'=z_s} \times C(y')e^{-ik_{mn}^+(z-z')}e^{-im(\theta-\theta')}dy'dz' + \int_{y'=-d}^{y'=+d}\int_{z'=z_s}^{z'=z_s+\sqrt{d^2-y'^2}} C(y')e^{-ik_{mn}^-(z-z')} \times e^{-im(\theta-\theta')}dy'dz'\right], \quad (20)$$

and then from Eq. (13) where w' is the coordinates of a point that belongs to S_{ur},

$$Z_{u,rs} = \frac{1}{\bar{v}_{us}S_{ur}}\int_{y'=-d}^{y'=+d}\int_{z'=z_r-\sqrt{d^2-y'^2}}^{z'=z_r} C(y')p(w'|w_{us})dz'dy' + \frac{1}{\bar{v}_{us}S_{ur}}\int_{y'=-d}^{y'=+d}\int_{z'=z_r}^{z'=z_r+\sqrt{d^2-y'^2}} C(y')p(w'|w_{us})dz'dy'. \quad (21)$$

3. The tube modeling

The HQ tubes are connected to the main duct through two interfaces, an input and an output with the same area S_t, which is plane then different of S_u (Fig. 2). The sound field

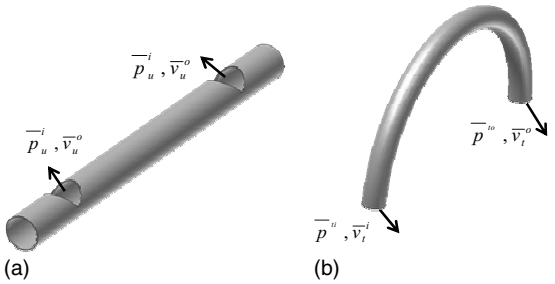


FIG. 4. Acoustic boundary conditions at the tube and duct interfaces: (a) for the duct and (b) for the tube.

inside the tubes is assumed to consist of plane wave only, a valid assumption well below the first tube cut-on frequency. Then the pressure and normal acoustic particle velocity are constant on S_t and represented by averaged values $(\bar{p}_t^i, \bar{v}_t^i)$ and $(\bar{p}_t^o, \bar{v}_t^o)$ (Fig. 4), respectively, at the input interface and at the output interface. In practice, the HQ tubes are hemitoroidal. However, for modeling purposes, they are considered as straight tubes with uniform cross-section. Let L be the length of the tube. The well known transfer matrix[17] links $(\bar{p}_t^o, \bar{v}_t^o)$ and $(\bar{p}_t^i, \bar{v}_t^i)$.

$$\begin{Bmatrix} \bar{p}_t^i \\ \rho c \bar{v}_t^i \end{Bmatrix} = \begin{bmatrix} \cos(kL) & i\sin(kL) \\ i\sin(kL) & \cos(kL) \end{bmatrix} \begin{Bmatrix} \bar{p}_t^o \\ \rho c \bar{v}_t^o \end{Bmatrix},$$

$$t = 1 \cdots 2N_{\mathrm{HQ}}, \tag{22}$$

from which can be deduced the mobility matrix $Z_{t,\mathrm{HQ}}$ linking $(\bar{p}_t^i, \bar{p}_t^o)$ and $(\bar{v}_t^i, \bar{v}_t^o)$. A consistent convention for the positive direction for the normal acoustic particle velocity must be kept. To this end, the positive particle velocity in the entrance end of the tube, which is opposite to the positive source velocity, is reversed by changing the sign of the first column of the matrix $Z_{t,\mathrm{HQ}}$. Furthermore, the following relationship between pressure and normal acoustic particle velocity of all input and output of the HQ tubes is then deduced in the matrix form:

$$\{\bar{p}_t\}_{2N_{\mathrm{HQ}}} = [Z_{\mathrm{HQ}}]_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} \{\bar{v}_t\}_{2N_{\mathrm{HQ}}}. \tag{23}$$

4. Coupling conditions at the tube-duct interfaces

Let us consider the case of a bicylindrical intersection between the tube and the duct (Fig. 2). Exact calculations of the integrals, which describe the influence of the tubes, are not possible since the area of the intersection between the duct and the tube is not plane but saddle-shaped: S_u, the surface of the main duct, does not coincide with S_t, the surface of the side-branch. Neglecting the effect of the volume in black in Fig. 2, the coupled tube-duct system is obtained by matching the average pressure (velocity) on S_u to the pressure (velocity) on S_t, i.e., $\bar{p}_u = \bar{p}_t$ and $\bar{v}_u = \bar{v}_t$. Then Eq. (23) relates the velocity vector at all the interfaces to the pressure vector. When substituted in Eq. (9), the following relationship, which gives the velocity at the tube-duct interfaces as a function of the impedance matrix of the coupled system and the incident pressures calculated at all the interfaces, is deduced:

$$\{v_u\}_{2N_{\mathrm{HQ}}} = [Z_{\mathrm{HQ}} - Z_{rs}]^{-1}_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} \{\Gamma_u\}_{2N_{\mathrm{HQ}}}. \tag{24}$$

C. Analytical expression of HQ scattering matrix

From Eq. (24) the expression of the velocity at all the tube-duct interfaces is given by

$$\{v_u\}_{2N_{\mathrm{HQ}}} = [Y]_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} \{\Gamma_u\}_{2N_{\mathrm{HQ}}}, \tag{25}$$

where $[Y]_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} = [Z_{\mathrm{HQ}} - Z_{rs}]^{-1}_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}}$. From Eq. (12), it comes

$$\{v_u\}_{2N_{\mathrm{HQ}}} = [Y]\{\boldsymbol{\beta}_1^+; \boldsymbol{\beta}_2^+; \ldots; \boldsymbol{\beta}_{2N_{\mathrm{HQ}}}^+\}\boldsymbol{\Gamma}. \tag{26}$$

The modal pressure radiated in positive and negative z-directions by the source u with acoustic velocity $\bar{v}_u$ is given by

$$\{p_{u,mn}^{\pm}\}_{2N_{\mathrm{HQ}}} = \{\mathrm{diag}(\alpha_{u,mn}^{\pm})\}_{2N_{\mathrm{HQ}}} \{\bar{v}_u\}_{2N_{\mathrm{HQ}}}, \tag{27}$$

$$\{p_{u,mn}^{\pm}\}_{2N_{\mathrm{HQ}}} = \{\mathrm{diag}(\alpha_{u,mn}^{\pm})\}_{2N_{\mathrm{HQ}}} [Y]_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} \times [\beta^+]_{2N_{\mathrm{HQ}}\times N} \boldsymbol{\Gamma}, \tag{28}$$

where $\alpha_{u,mn}^{\pm} = (k\rho c/\pi a^2) L_{mn} e^{-ik_{mn}^{\pm} z} \int_{y'=-d}^{y'=+d} \int_{z'=z_u-\sqrt{d^2-y'^2}}^{z'=z_u+\sqrt{d^2-y'^2}} \times (a/\sqrt{a^2-y'^2}) e^{ik_{mn}^{\pm} z'} e^{-im(\theta-\theta')} dz'\,dy'$ and $[\beta^+] = \{\boldsymbol{\beta}_1^+; \boldsymbol{\beta}_2^+; \ldots; \boldsymbol{\beta}_{2N_{\mathrm{HQ}}}^+\}$. The authors sum then on all the sources radiated in the positive z-direction to have the total transmitted modal pressure

$$p_{\mathrm{tot},mn}^+ = \Gamma_{mn} + \sum_{u=1}^{2N_{\mathrm{HQ}}} \alpha_{u,mn}^+ [Y][\beta^+]\boldsymbol{\Gamma}. \tag{29}$$

Hence, the vector of the total modal transmitted amplitudes is expressed by means of the vector of the modal incident pressures

$$\{p_{\mathrm{tot}}^+\}_{N\times 1} = [S^{12}]_{N\times N} \boldsymbol{\Gamma}_{N\times 1} \tag{30}$$

with

$$[S^{21}]_{N\times N} = \begin{bmatrix} \alpha_{1,00}^+ & \cdots & \alpha_{2N_{\mathrm{HQ}},00}^+ \\ \alpha_{1,10}^{+,} & \cdots & \alpha_{2N_{\mathrm{HQ}},10}^+ \\ & \vdots & \\ \alpha_{1,N}^+ & \cdots & \alpha_{2N_{\mathrm{HQ}},N}^+ \end{bmatrix} \times [Y]_{2N_{\mathrm{HQ}}\times 2N_{\mathrm{HQ}}} [\beta^+]_{2N_{\mathrm{HQ}}\times N} + [I]_{N\times N},$$

where $[S^{21}]_{N\times N}$ is the transmission matrix related to the wave coming into the test element from the left side.

To get the total reflected modal pressure, the authors sum then on all the sources radiated in negative z-direction

$$p_{\mathrm{tot},mn}^- = \sum_{u=1}^{2N_{\mathrm{HQ}}} \alpha_{u,mn}^- [Y][\beta^+]\boldsymbol{\Gamma}. \tag{31}$$

The vector of the total modal reflected amplitudes can be expressed as a function of the vector of the modal incident pressures

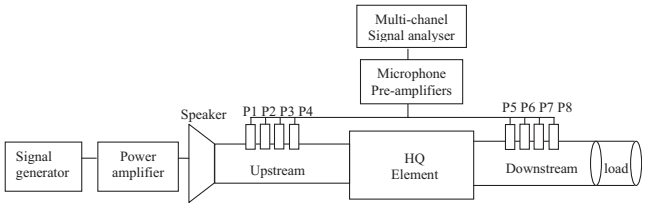


FIG. 5. Experimental set-up. Source on the left side configuration.

$$\{p_{\text{tot}}^{-}\}_{N\times 1} = [S^{11}]_{N\times N}\mathbf{\Gamma}_{N\times 1}, \tag{32}$$

with

$$[S^{11}]_{N\times N} = \begin{bmatrix} \alpha_{1,00}^{-} & \cdots & \alpha_{2N_{\text{HQ}},00}^{-} \\ \alpha_{1,10}^{-} & \cdots & \alpha_{2N_{\text{HQ}},10}^{-} \\ & \vdots & \\ \alpha_{1,N}^{-} & \cdots & \alpha_{2N_{\text{HQ}},N}^{-} \end{bmatrix} \times [Y]_{2N_{\text{HQ}}\times 2N_{\text{HQ}}}[\beta^{+}]_{2N_{\text{HQ}}\times N},$$

where $[S^{11}]_{N\times N}$ is the reflection matrix of the wave coming in the element from the left side.

The symmetry of the HQ element imposing the relations $[S^{1,1}]_{N\times N}=[S^{2,2}]_{N\times N}$ and $[S^{2,1}]_{N\times N}=[S^{1,2}]_{N\times N}$, the expression of the matrix $[S]_{2N\times 2N}$ is known as

$$[S]_{2N\times 2N} = \begin{bmatrix} [S^{1,1}]_{N\times N} & [S^{1,2}]_{N\times N} \\ [S^{2,1}]_{N\times N} & [S^{2,2}]_{N\times N} \end{bmatrix}. \tag{33}$$

The physical meaning of each elementary matrix is as follows:[18] $[S_{mn,pq}^{1,1}]_{N\times N}$ describes the reflection of the wave coming into the element from the left side, $[S_{mn,pq}^{2,1}]_{N\times N}$ describes the transmission of the wave coming into the element from the left side, $[S_{mn,pq}^{2,2}]_{N\times N}$ describes the reflection of the wave coming into the element from the right side, and $[S_{mn,pq}^{1,2}]_{N\times N}$ describes the transmission of the wave coming into the element from the right side.

III. THE MODEL VALIDATION

The validation of the modeling described before is conducted on a duct configuration with radius $a=25$ mm and a HQ element made of two branched tubes with radius $d=a=25$ mm. The length of the centerline of each tube is $L=56$ cm, the distance between the input and output area centers is $L'=30$ cm, and an angular spacing of $\theta_0=180°$ is chosen as shown in Fig. 6. In this case, as pointed out in Fig. 3, the effect of the saddle-shaped area is maximum. Each tube is branched to the duct through a straight duct section long enough to avoid the effects produced by a curved tube. If all the modeling is valid for a 3D configuration, the validation procedure is applied up to a frequency domain $ka<1.84$ in order to make the data analysis clearer. The incident pressure will then be assumed to be a plane wave. The validation procedure is based on the comparison of the scattering matrix coefficients of the HQ system deduced from the model with experimental and numerical results.

A. Experimental set-up and data processing

The experimental set-up is shown in Fig. 5. At one end

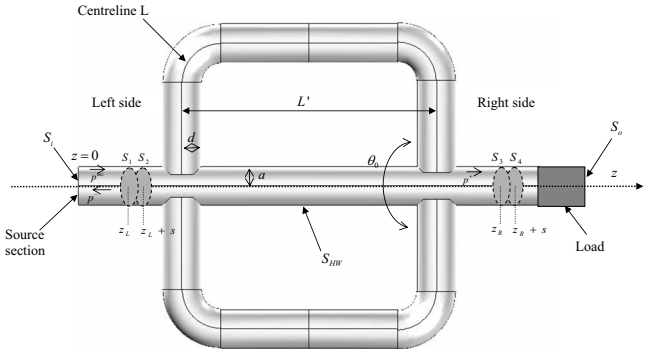


FIG. 6. The 2-HQ configuration tested.

of the duct, a loudspeaker is driven by a signal generator module and produces white noise and at the other end noise radiates outside through an unflanged inlet, which constitutes the load. Pressure measurements deduced from the transfer function between an electret microphone, which is moved, and the signal coming out from the power amplifier are acquired at eight positions, four on either side the HQ element. Within each group of microphones, the measurement points are equally spaced by $s=14$ mm and the distance between the measurement point closed to the HQ system has been chosen to avoid the effect of the evanescent modes produced at the discontinuity. This overdetermination technique is chosen to improve the separation step between the incident and

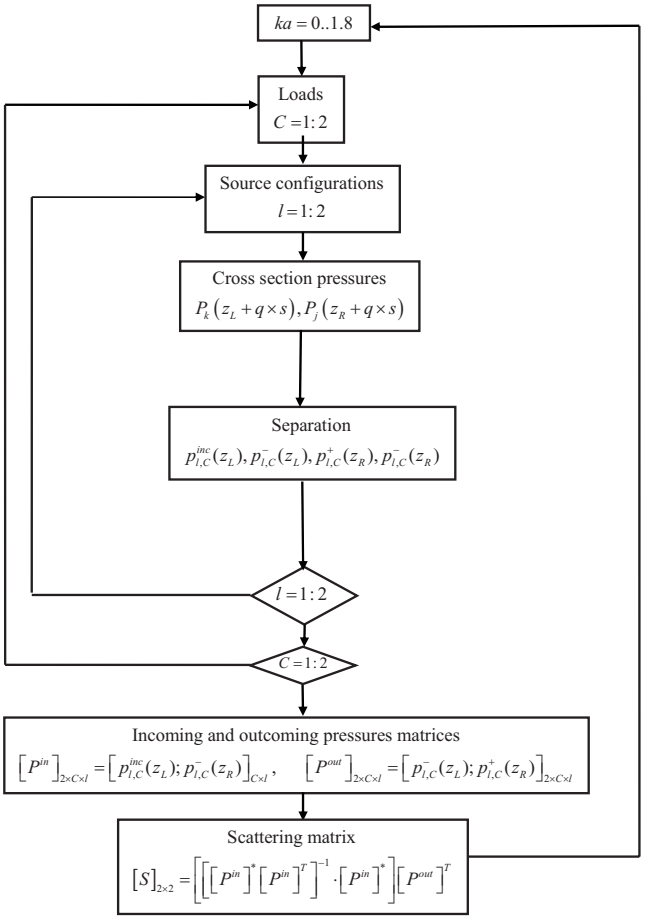


FIG. 7. Flow-chart for experimental scattering matrix calculation; $q \in [0,1,2,3] j \in [4,5,6,7]$.

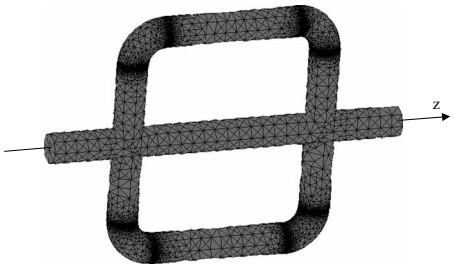


FIG. 8. The numerical meshing of the 2-HQ configuration.

the reflected wave. The measurement of $[S]$ is based on a technique that is a combination between the two-source technique and the two-load one[18] leading to an overdeterminated system. Indeed, the loudspeaker is located on the left side, then on the right side of the HQ element, and for each source location two different load conditions are created by changing only the length of the duct. The scattering matrix $[S]$ of the element located between z_L and z_R (Fig. 6) is then deduced following the steps described in the flow-chart (Fig. 7).

B. Numerical approach

A comparison has been made with the results from the 3D-FEM analysis of the same HQ configuration performed with FEMLAB®, a commercial FEM package.

The acoustic pressure in the duct presented in Fig. 6 is solution of the following system:

$$\Delta p + k^2 p = 0(\Omega),$$

$$\frac{\partial p}{\partial n_W} = 0(S_{\mathrm{HW}}), \tag{34}$$

where ∂n_W is the outward derivative normal to the surface S_{HW}. Helmholtz's equation is integrated over a finite internal volume Ω limited by the inlet section S_i and the outlet section S_o and by the hard walls of the duct and the two HQ tubes. At the source section S_i, a constant total pressure is imposed to produce an incident plane wave and on the right side an anechoic termination condition $Z=\rho c$ is imposed. The FEM is implemented using isoparametric tetrahedral elements. To ensure enough accuracy and to approximate as close as possible the duct-tube interface area a fine mesh with 38296 nodes was chosen (Fig. 8). The origin of the z axis is taken at the reference phase in the source z axis position (Fig. 6). A 625-point discretization of the acoustic pressure field computed in each of the two pairs of cross-sectional areas is achieved to control the plane wave assumption and the anechoic condition on the right side. The procedure for computing the scattering matrix is the same described by the flow-chart (Fig. 7) where only one source configuration and two calculation cross-sections are chosen.

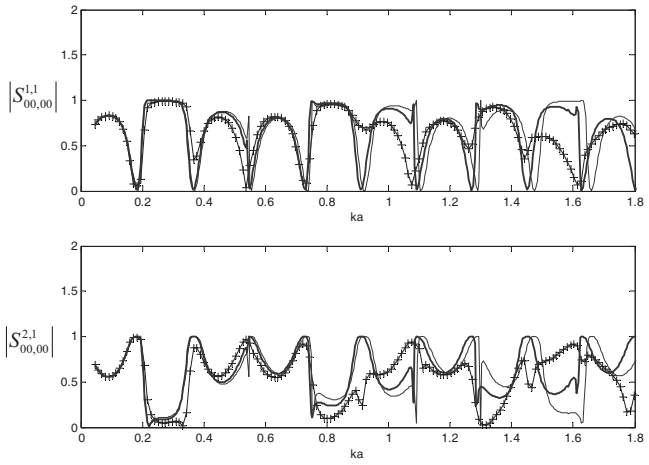


FIG. 9. Scattering coefficient of reflection $S^{1,1}_{00,00}$ and transmission $S^{2,1}_{00,00}$ of the mode (0,0) for the two tubes configuration; 1D (Ref. 11) (—), 1D with length correction (Ref. 10) (—), and FEM (--+--).

IV. RESULTS

The modulus of the transmission coefficient from the left side to the right side $|S^{2,1}_{00,00}|$ and the reflection coefficient from source side $|S^{1,1}_{00,00}|$ deduced from the 1D model with[10] and without[11] length correction are plotted versus ka in Fig. 9. The analytical model without length correction shows a good agreement with results deduced from FEM at low frequencies below $ka=0.8$ on the amplitude of both coefficients but a small shift (Table I) on the minimum reflection frequencies increases with ka. When $ka>0.8$ this model disagrees as expected with the FEM curve. Indeed, previous works[10,11] have shown that this shift was associated with the

TABLE I. Relative difference of HQ frequencies (in percent) referenced to FEM, 1D model, 1D model with length correction, 3D model, 3D bicylindrical model, and experiment.

Reflexion minima	ka	1D[a]	1D[b]	3D[c]	3D bicyl.	Expt.
1	0,185	0,27	0,27	0,00	8,11	8,11
2	0,366	1,09	1,09	6,28	1,09	2,46
3	0,54	2,59	2,59	2,59	0,00	0,00
4	0,73	1,37	1,37	6,85	0,41	1,37
5	0,9075	1,08	2,15	4,30	2,15	0,00
6	1,0884	2,70	1,40	1,86	0,93	2,33
7	1,2693	2,38	0,79	3,97	0,00	0,40
8	1,4502	2,43	0,69	12,50	0,69	0,69
9	1,6311	2,79	0,62	2,17	0,74	1,24

[a]Reference 11.
[b]Reference 10.
[c]Reference 13.

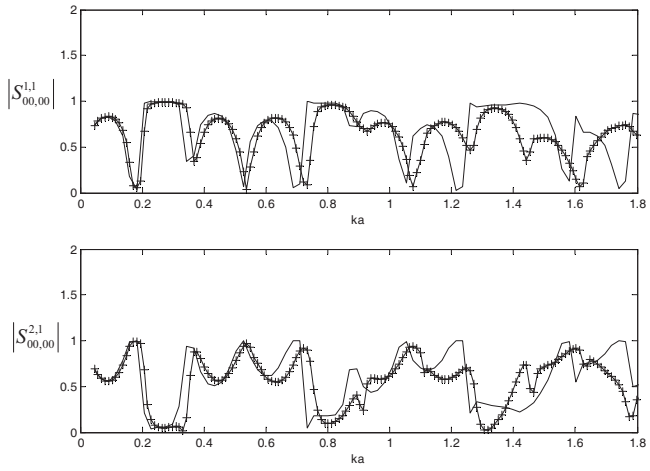


FIG. 10. Scattering coefficient of reflection $S^{1,1}_{00,00}$ and transmission $S^{2,1}_{00,00}$ of the mode (0,0) for two tube configuration; 3D (Ref. 13) (____) and FEM (-----).

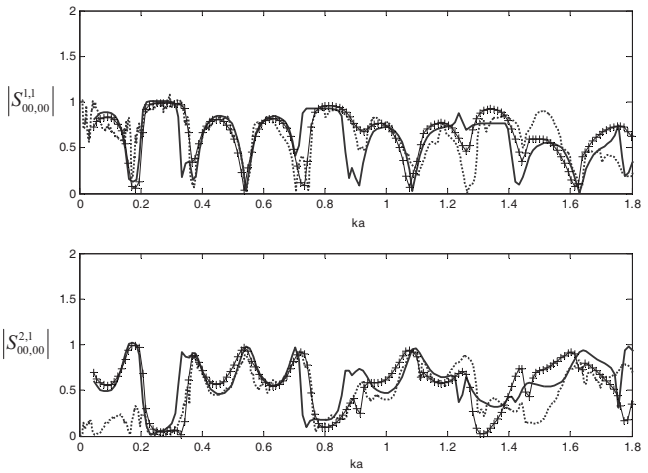


FIG. 11. Scattering coefficient of reflection $S^{1,1}_{00,00}$ and transmission $S^{2,1}_{00,00}$ of the mode (0,0) for two tube configuration; bicylindrical (—), FEM (--+--), and experiment (----).

generation of evanescent modes at the interface between a duct and a branched tube and that it can be corrected introducing, if the side-branch is long enough, a correction to the tube length in the 1D model. Indeed, for example,[10] a correction function versus ka was proposed for a side-branched tube mounted perpendicular to a cylindrical main pipe, both ducts having the same diameter. This length correction function has been added to the 1D model of the HQ system to get the results presented in Fig. 9, which shows a good agreement with FEM model on the location of the frequencies (Table I) associated with the minima of reflection. But a difference still remains on the amplitude of these coefficients. In Fig. 10, results on the modulus of the transmission coefficient from the left side to the right side $|S^{2,1}_{00,00}|$ and on the reflection coefficient from source side $|S^{1,1}_{00,00}|$ deduced from the 3D model[13] are compared with FEM predictions. The 3D analytical calculations were carried out on the basis of 20 modes in the calculation of Green's function to ensure the convergence of the solution. The results show that this 3D model better predicts the nonperiodic variation of both coefficients versus ka unlike to the 1D model. But a shift on the frequencies of minima of reflection and a disagreement on the amplitude have to be noticed, which can be probably explained because of the geometrical approximation $d/a \ll 1$, which is not valid in the problem (Fig. 3). Finally, the results of the variation of $|S^{2,1}_{00,00}|$ and $|S^{1,1}_{00,00}|$ versus ka deduced from the bicylindrical technique compared with FEM and experimental data are plotted in Fig. 11. A very good agreement between experimental and numerical results is shown except at low frequencies at $ka<0.2$ because of a low signal-to-noise ratio of the loudspeaker. Near $ka=1.25$ the measured values of $|S^{1,1}_{00,00}|$ and $|S^{2,1}_{00,00}|$ are, respectively, close to zero and 0.9 while the numerical ones are, respectively, close to 0.5 and 0.75. When frequency gets closer to $ka=1.841$, the (1,0) mode cut-off frequency, experimental and numerical results disagree because of the effect of the evanescent modes produced by the discontinuity. The local discrepancy is certainly due to minor geometrical defaults in the experimental set-up, such as added dissipation/transmission at the tube-duct junction, that might modify the amplitude of the coefficients just above the first duct cut-on frequency where the first transverse mode strongly excites the duct wall. Also, the inaccuracies observed in the measurement near the first duct cut-on frequency are due to the phase differences between the microphone signals, which become very small and lead to a decrease in the signal-to-noise ratio. Increasing the number of source and/or load configurations might improve the accuracy near the first cut-on frequency. Also, a more accurate phase calibration of the microphones, which would then be able to discriminate small phase differences, could be useful. It is noteworthy that the accuracy of the FEM results might be improved if one carefully accounts for viscous propagation effects, which are influent close to the duct cut-on frequencies.

The results deduced with the bicylindrical technique compared with the FEM data show (Fig. 11) that this technique not only reproduces as the 3D model[13] the nonperiodic variation of the coefficients versus ka but also localizes better the minima of the reflection coefficient deduced from the 1D model without any length correction (Table I). As expected, some differences remain between results deduced from FEM and bicylindrical model as, for example, near $ka=1.25$ and $ka=1.45$. Indeed, this model is based on nonvalid assumptions such as $\bar{v}_u$, the normal particle velocity, on the interface area S_u is constant. Indeed, the pressure variations are shown in Fig. 12 versus ka on the axis of the interface along which S_u and S_t, the surface of the tubes, intersect.

V. CONCLUSION

A 3D bicylindrical model was developed in order to take into account the effects of the saddle-shaped area at the interfaces between a branched tube and a main duct to compute the scattering matrix for a n-HQ tube configuration. Results deduced from this model were compared, in the plane wave frequency domain, with a 1D model with and without length correction as well as with experimental and numerical data obtained with a 2-HQ tube configuration having the same diameter as the main duct. Analysis of the results leads to the following conclusions.

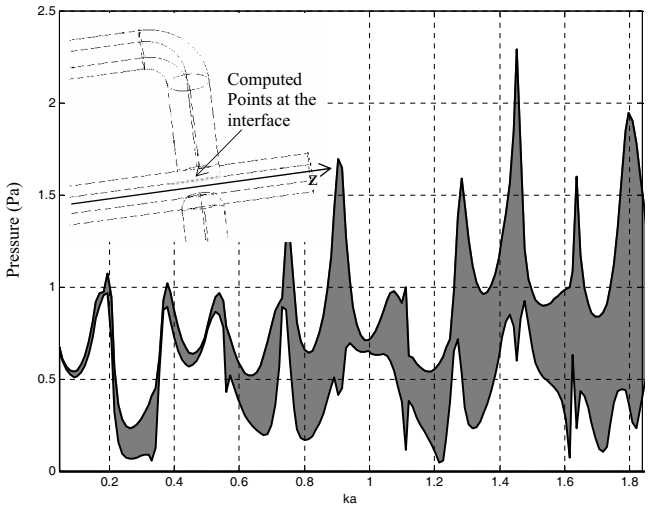


FIG. 12. Maxima and minima of the pressure versus adimensional frequency *ka*, calculated at ten axial positions on the interface by the FEM. The black zone represents the interval of the numerical values that takes the computed pressure at the positions on the interface.

- The 1D model[11] does not reproduce the nonperiodic variation with frequency of the reflection and transmission coefficients and the length correction[10] allows getting only the locations in the frequency domain of the minima in reflection but not their amplitude.
- The 3D model[13] reproduces the nonperiodic variation of the coefficients but does not lead to the exact locations in the frequency domain of the minima in reflection.
- The bicylindrical model reproduces the nonperiodic variation of the coefficients and leads to a better location in the frequency domain of the minima in reflection but still does not well correlate with the amplitudes.

In light of results, the computational cost of each of the various modeling approaches has to be discussed to evaluate the best method in practice. For well behaved problems, a grid of uniform mesh spacing (in each of the coordinate directions) gives satisfactory results. But for the geometry investigated, the solution is more difficult to estimate because of the singularities. One could use a uniform grid having a spacing fine enough so that the local errors estimated in these difficult regions are acceptable. But this approach is computationally extremely costly. For comparison, Table II presents the calculation time versus the number of tubes taking into account the modelization. For each configuration of tube, the mesh was chosen fine enough, particularly near the "saddle-shaped" discontinuities, to consider that the solution has converged. The results presented in the table show clearly that FEM technique is an interesting tool for a low number of tubes but is clearly an inadaptable technique with configurations of high number of tubes. Moreover, the analytical method is well adapted to compute results for optimization by changing easily the geometrical parameters as the diameter or the length of tube whereas a FEM requires a new computer-aided design in each case of parametrical test. From this point of view, the analytical method is closer to the industrial constraints, which still search for both accurate and optimized computation tools.

TABLE II. Calculation time (in seconds) of 1D, 3D, bicylindrical, and FEM codes versus the number of tube HQ.

No. of tubes	1D	3D	Bicylindrical model	FEM
1	0.47	9	12	2 512
2	0.53	18	29	4 280
3	0.54	33	58	8 949
10	0.55	403	628	30 397

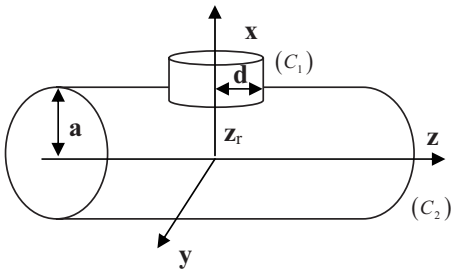


FIG. 13. Bicylindrical geometry.

Further works should be directed toward an improvement of the bicylindrical model, which should be extended to account for more realistic assumptions. In particular, one could use a quadric formulation (eventually in variational form) of the continuity conditions at the tube-duct interface that would account for the cross-sectional variations of the pressure and the velocity due to the local influence of the evanescent tube and duct modes.

ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support from the BDI-CNRS and SNECMA grant. The authors especially thank R. Marechal and J. Parzybut for their help during the experiments.

APPENDIX A: CALCULATION OF THE SADDLE-SHAPED AREA

In this study, to describe the analytical geometry, the authors use an *explicit representation*[16] to describe the bicylindrical surface as a set of points (x,y,z) satisfying an equation of the form $x=f(y,z)$. In the (x,y,z) space, they consider the surface corresponding to the intersection of (C_1) and (C_2) defined by

$$y^2+(z-z_r)^2 \le d^2 \quad (C_1),$$

$$x^2+y^2=a^2 \quad (C_2), \tag{A1}$$

where (C_1) is the boundary of the duct with radius a located in the Oz-direction and (C_2) is the tube with radius d located in the Ox-direction (cf. Fig. 13).

For the parametric vector of the space curve, the authors have three equations expressing x, y, and z in terms of two parameters u and v as follows:

$$x=X(u,v), \quad y=Y(u,v), \quad z=Z(u,v). \tag{A2}$$

Here the point (u,v) is allowed to vary over some 2D connected set D in the uv plane, and the corresponding points (x,y,z) trace out the surface in (x,y,z) space (Fig. 14). The

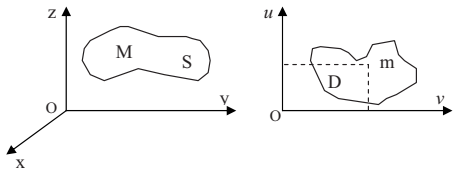


FIG. 14. Parametric representation of a 3D surface.

presence of the two parameters makes possible the 3D intersection surface to be described by only two degrees of freedom. If the authors introduce the radius vector $\boldsymbol{F}$ from the origin to a generic point (x,y,z) of the surface, they may combine the three parametric equations into one vector equation of the form

$$\boldsymbol{F}(u,v) = X(u,v)\cdot\boldsymbol{i} + Y(u,v)\cdot\boldsymbol{j} + Z(u,v)\cdot\boldsymbol{k} \tag{A3}$$

with u and v running on D. This is called the vector equation for the surface. The area of the elementary parallelogram spanned by $(\partial\boldsymbol{F}/\partial u)\Delta u$ and $(\partial\boldsymbol{F}/\partial v)\Delta v$ is the magnitude of the cross product

$$\left\| \frac{\partial \boldsymbol{F}}{\partial u}\Delta u \wedge \frac{\partial \boldsymbol{F}}{\partial v}\Delta v \right\|. \tag{A4}$$

After summing on all the elementary parallelograms and making Δu and Δv tending toward zero, the authors get the area S of the bicylindrical surface

$$S = \int\int \left\| \frac{\partial \boldsymbol{F}}{\partial u} \wedge \frac{\partial \boldsymbol{F}}{\partial v} \right\| du dv. \tag{A5}$$

Using Eq. (A1), the authors can give an explicit expression of y as follows:

$$x = \sqrt{a^2 - y^2}. \tag{A6}$$

To simplify the calculation, the authors take in a first time $z_r=0$ and they set $u=y$ and $v=z$. They define for every point (x,y,z) the application F as

$$(y,z) \xrightarrow{F} \begin{vmatrix} \sqrt{a^2-y^2} \\ y \\ z \end{vmatrix} \tag{A7}$$

and

$$\frac{\partial \boldsymbol{F}}{\partial y} \wedge \frac{\partial \boldsymbol{F}}{\partial z} = \begin{vmatrix} \dfrac{-y}{\sqrt{a^2-y^2}} \\ 1 \\ 0 \end{vmatrix} \wedge \begin{vmatrix} 0 & 1 \\ 0 & \dfrac{y}{\sqrt{a^2-y^2}} \\ 1 & 0 \end{vmatrix}. \tag{A8}$$

The authors deduce from Eqs. (A5) and (A8) the expression of S as follows:

$$S = \int\int_{\Delta} \frac{a}{\sqrt{a^2-y^2}} dz dy, \tag{A9}$$

where $\Delta=\{(y,z)/y^2\le a^2, y^2+z^2\le d^2\}$. Since $d\le a$, the domain Δ is obviously $\Delta=\{(y,z)/y^2+z^2\le d^2\}$; hence the area S becomes

$$S = \int_{y=-d}^{y=+d} \int_{z=-\sqrt{d^2-y^2}}^{z=+\sqrt{d^2-y^2}} \frac{a}{\sqrt{a^2-y^2}} dz dy \tag{A10}$$

and by symmetry

$$S = 4a\int_{y=0}^{y=d} \sqrt{\frac{d^2-y^2}{a^2-y^2}} dy. \tag{A11}$$

In the case where a function f is integrated over a bicylindrical section and $z_r\neq 0$, the authors are led to compute

$$I = \int_{y=-d}^{y=+d} \int_{z=z_r-\sqrt{d^2-y^2}}^{z=z_r+\sqrt{d^2-y^2}} f(y,z) \frac{a}{\sqrt{a^2-y^2}} dz dy. \tag{A12}$$

APPENDIX B: CALCULATIONS OF INTEGRALS ON A BICYLINDRICAL SECTION

1. Calculation of the incident pressure over the bicylindrical section

The calculation of the incident pressure over the bicylindrical section is given by

$$\bar{p}_u^{\mathrm{inc}}(w_u) = \boldsymbol{\beta}_u^+ \boldsymbol{\Gamma} \tag{B1a}$$

with $\boldsymbol{\Gamma}=\{\Gamma_{00}\cdots\Gamma_N\}$ the modal incident components and $\boldsymbol{\beta}_u^+ = \{\beta_{u,1}^+\cdots\beta_{u,N}^+\}$ where

$$\beta_{u,mn}^+ = J_{mn}(k_{mn}a)\frac{1}{S_u}\int_{S_u} e^{-im\theta}e^{-ik_{mn}^+ z} dS. \tag{B1b}$$

The authors deduce from Eq. (A12) the expression of $\beta_{u,mn}^+$ as

$$\beta_{u,mn}^+ = J_{mn}(k_{mn}a)\frac{1}{S_u}\int_{y=-d}^{y=+d}\int_{z=z_u-\sqrt{d^2-y^2}}^{z=z_u+\sqrt{d^2-y^2}} \frac{a}{\sqrt{a^2-y^2}} \times e^{-im\theta}e^{-ik_{mn}^+ z} dz dy \tag{B1c}$$

with

$$S_u = 4a\int_0^d \sqrt{\frac{d^2-y^2}{a^2-y^2}} dy \quad \text{and} \quad \theta = a\sin(y/a).$$

2. Calculation of Green's function over the bicylindrical section

Green's function is represented on the basis of the rigid-walled cylindrical duct eigenmodes $\Phi_{mn}(r,\theta) = J_m(\chi_{mn}/ar)e^{-im\theta}$ for a source located at (a,θ_s,z_s) as follows:

$$G^{\pm}(r,\theta,z|a,\theta_s,z_s) = \frac{i}{\pi a^2}\sum_{m=-Mg}^{Mg}\sum_{n=0}^{Ng} \frac{\Phi_{mn}(r,\theta)\Phi_{mn}(a,\theta_s)}{N_{mn}(k_{mn}^+ - k_{mn}^-)} \times e^{-ik_{mn}^{\pm}(z-z_s)}. \tag{B2a}$$

The authors deduce from Eq. (A12) the expression of the pressure radiated by a source located at (a,θ_s,z_s) on a point located at (a,θ_r,z_r) by integrating Green's function over the bicylindrical section

$$p(a,\theta_r,z_r|a,\theta_s',z_s') = -i\omega\rho v_0 \int_{y_s'=-d}^{y_s'=+d} \int_{z_s'=z_s-\sqrt{d^2-y_s'^2}}^{z_s'=z_s+\sqrt{d^2-y_s'^2}} \frac{a}{\sqrt{a^2-y_s'^2}} G(a,\theta_r,z_r|a,\theta_s,z_s)dz_s'dy_s' \tag{B2b}$$

and then,

$$p(a,\theta_r,z_r|a,\theta_s',z_s') = \frac{\bar{v}k\rho c}{\pi a^2}\sum_m \sum_n L_{mn} \int_{y_s'=-d}^{y_s'=+d} \int_{z_s'=z_s-\sqrt{d^2-y_s'^2}}^{z_s'=z_s+\sqrt{d^2-y_s'^2}} C(y_s')e^{-ik_z^+(z-z_s')}e^{-im(\theta_r-\theta_s')}dy_s'dz_s' \tag{B2c}$$

with

$$L_{mn} = \frac{J_m(\chi_{mn})^2}{N_{mn}(k_{mn}^+ - k_{mn}^-)} \quad \text{and} \quad C(y) = \frac{a}{\sqrt{a^2-y^2}}.$$

APPENDIX C: NUMERICAL COMPUTATION OF GREEN'S INTEGRAL ON A BICYLINDRICAL SECTION

This bicylindrical section's integral requires a numerical integration to be computed. Another solution is to develop it in terms of numerical series. It is the object of the following part.

To calculate Green's function on the bicylindrical section, the following simplified integral must be solved:

$$I(a,d) = 4a \int_{y=0}^{y=d} \int_{z=0}^{z=\sqrt{d^2-y^2}} \frac{e^{ik_z z}}{\sqrt{a^2-y^2}} e^{-im\theta} dz dy. \tag{C1}$$

1. Series development in the case $m=0$

First the special case $m=0$ is developed. A simplified expression is deduced as

$$I(a,d) = 4a \int_{y=0}^{y=d} \int_{z=0}^{z=\sqrt{d^2-y^2}} \frac{e^{ik_z z}}{\sqrt{a^2-y^2}} dz dy. \tag{C2}$$

This expression is developed in the terms of numerical series

$$f(z) = e^{ik_z z} = \sum_0^\infty \frac{(ik_z z)^n}{n!} = \sum_{n=0}^\infty a_n \tag{C3}$$

and

$$g(y) = \frac{1}{\sqrt{a^2-y^2}} = \left(\frac{1}{a}\right)\left(\frac{1}{\sqrt{1-\left(\frac{y}{a}\right)^2}}\right), \tag{C4}$$

$$g(z) = \frac{1}{a}\sum_{n=0}^\infty - \frac{(2n-1)}{n!\,2^{2n-1}}\left(\frac{y}{a}\right)^n = \sum_{j=0}^\infty b_j. \tag{C5}$$

The Cauchy's theorem is used for the product of two series for the functions f and g as

$$fg = \sum_{n=0}^\infty C_n = \left(\sum_{i=0}^\infty a_i\right)\left(\sum_{j=0}^\infty b_j\right) = \sum_{n=0}^\infty \sum_{k=0}^n a_k b_{n-k}, \tag{C6}$$

$$\frac{e^{ik_z z}}{\sqrt{a^2-y^2}} = \frac{-1}{a}\sum_{n=}^\infty \sum_{j=0}^n \left(\frac{(2j-1)!}{j!\,2^{2j-1}}\right)\left(\frac{(ik_z)^{(n-j)}}{a^j(n-j)!}\right) z^j y^{n-j}. \tag{C7}$$

To evaluate $I(a,d)$ expression (C7) must be integrated on the domain Δ. If the authors set

$$\frac{e^{ik_z z}}{\sqrt{a^2-y^2}} = \sum_{n=0}^\infty \sum_{j=0}^n \alpha(n,j) z^j y^{n-j}, \tag{C8}$$

then thanks to Lebesgue's theorem; it follows that

$$I(a,d) = -4a \sum_{n=0}^\infty \sum_{j=0}^n \alpha(n,j) \int_{y=0}^{y=d} y^{n-j} \frac{(d^2-y^2)^{(j+1)/2}}{j+1} dy. \tag{C9}$$

From this result, the following expression of the pressure field is deduced:

$$p(r,\theta,z|r_0,\theta_0,z_0) = -i\omega\rho v_0 I(a,d) \tag{C10}$$

2. Numerical methods, the case $m \neq 0$

In the case $m \neq 0$, the problem is not trivial. The difficulty comes from integral (C1), which cannot be developed as series and requires also a numerical integration.

[1] J. F. W. Herschel, "On the absorption of light by colored media, viewed in connection with the undulatory theory," London and Edinburgh Philosophical Magazine and Journal of Science **3**, 401–412 (1833).
[2] J. P. Smith and R. A. Burdisso, "Experimental investigation of the Herschel-Quincke tube concept on the honeywell TFE731," Report No. 60 NASA/CR 2002 211431, National Aeronautics and Space Administration Langley Research Center, Virginia Polytechnic Institute and State University of Mechanical Engineering Blacksburg, VA 24061, 2002.
[3] S. Griffin, S. Huybrechts, and S. A. Lane, "An adaptive Herschel-Quincke tube," J. Intell. Mater. Syst. Struct. **10**, 956–961 (1999).
[4] Y. Hwang, J. M. Lee, and S. J. Kim, "New active muffler system utilizing destructive interference by difference of transmission paths," J. Sound Vib. **262**, 175–186 (2003).
[5] E. P. Trochon, "A new type of silencer for turbocharger noise control," 2001-01-1436 SAE Conference on Noise and Vibration Control, Traverse City, MI (2001).
[6] I. McLean, "Optimized Herschel-Quincke acoustic filter," 2005-01-2360 SAE Conference on Noise and Vibration, Traverse City, MI (2005).
[7] M. El-Raheb and P. Wagner, "Acoustic propagation in rigid sharp bends and branches," J. Acoust. Soc. Am. **67**, 1914–1923 (1980).
[8] T. C. Redmore and K. A. Mulholland, "The application of mode coupling theory to the transmission of sound in the sidebranch of a rectangular duct system," J. Sound Vib. **85**, 323–331 (1982).
[9] V. Dubos, J. Kergomard, A. Khettabi, D. H. Keefe, J. P. Dalmont, and C. J. Nederveen, "Theory of the junction between a branched tube and a main guide using modal decomposition," Acust. Acta Acust. **85**, 153–169 (1999).

[10]Z. L. Ji, "Acoustic length correction of closed cylindrical side-branched tube," J. Sound Vib. **283**, 1180–1186 (2005).
[11]A. Selamet, N. S. Dickey, and J. M. Novak, "The Herschel-Quincke tube: A theoretical, computational, and experimental investigation," J. Acoust. Soc. Am. **96**, 3177–3185 (1994).
[12]G. W. Stewart, "The theory of the Herschel-Quincke tube," Phys. Rev. **31**, 696–698 (1928).
[13]R. F. Hallez, J. P. Smith, and R. A. Burdisso, "Control of higher-order modes in ducts using arrays of Herschel-Quincke waveguides," Proceedings of the Symposium from the 2000 International Mechanical Congress and Exposition, Orlando, FL, 5–10 Nov. 2000, pp. 87–100.
[14]R. F. Hallez and R. A. Burdisso, "Analytical modeling of Herschel-Quincke concept applied to inlet turbofan engines," Report No. NASA/CR-2002-211429, National Aeronautics and Space Administration Langley Research Center, Virginia Polytechnic Institute and State University Department of Mechanical Engineering, Blacksburg, VA 24061, 2002.
[15]M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[16]M. Apostol, "Multi-variable calculus and linear algebra, with applications to differential equations and probability," *Calculus*, 2nd ed. (Blaisdell, Waltham, MA, 1969), Vol. **2**.
[17]L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987).
[18]A. Sitel, J. M. Ville, and F. Foucart, "Multiload experimental procedure for measurement of acoustic scattering matrix of a duct discontinuity for higher order modes propagation conditions," J. Acoust. Soc. Am. **120**(5), 2478–2490 (2006).

Modeling subjective evaluation of soundscape quality in urban open spaces: An artificial neural network approach

Lei Yu and Jian Kang[a)]
School of Architecture, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

(Received 1 September 2008; revised 7 June 2009; accepted 28 June 2009)

This research aims to explore the feasibility of using computer-based models to predict the soundscape quality evaluation of potential users in urban open spaces at the design stage. With the data from large scale field surveys in 19 urban open spaces across Europe and China, the importance of various physical, behavioral, social, demographical, and psychological factors for the soundscape evaluation has been statistically analyzed. Artificial neural network (ANN) models have then been explored at three levels. It has been shown that for both subjective sound level and acoustic comfort evaluation, a general model for all the case study sites is less feasible due to the complex physical and social environments in urban open spaces; models based on individual case study sites perform well but the application range is limited; and specific models for certain types of location/function would be reliable and practical. The performance of acoustic comfort models is considerably better than that of sound level models. Based on the ANN models, soundscape quality maps can be produced and this has been demonstrated with an example.
 [DOI: 10.1121/1.3183377]

PACS number(s): 43.50.Rq, 43.50.Qp, 43.50.Sr, 43.66.Lj [BSF] Pages: 1163–1174

I. INTRODUCTION

The term soundscape, using the analogy to landscape, has entered the lexicons of a range of disciplines relating to the acoustic environment.[1,2] Soundscape studies have been paid more and more attention in both physical and aesthetic qualities of an environment.[3–15] Soundscape is also becoming an essential consideration in urban planning and design, where the subjective evaluation of sound level and sound preference are two main components in soundscape evaluation, and acoustic comfort is considered as an overall criterion.[8] While previous studies have acknowledged the importance of taking sound facet in formulating the quality of urban environment, especially in noise evaluation,[9,10,16–19] there is a recognized need to find efficient ways to transfer research results into the common practice of soundscape design. Therefore, the aim of this study is to develop computer-based models to aid urban planners and designers in understanding the potential users' evaluation of an expected soundscape in urban open spaces. With this aim, two types of computer model have been constructed, for the subjective evaluation of sound level and acoustic comfort, respectively.

In order to develop the computer models, the artificial neural network (ANN) technique has been introduced and employed.[20–22] The hypothesis of the study is that a well-trained ANN model derived from existing sites can be used to predict the subjective evaluation of soundscape in new urban open spaces with similar physical and social environments. ANN is a multilayered network, where the relationships between input variables and the output are crucial to determine the prediction performance.[23] It has already been applied in the field of acoustics, such as for sound source identification.[24] In addition to ANN, other models have also been explored in the field of soundscape and community noise.[25–27]

In this study, factors that may affect the subjective evaluation of sound level and acoustic comfort have first been statistically examined based on the data from large scale field surveys, and the results are used to select appropriate input variables for ANN models. Consequently, prediction models for the subjective evaluation of sound level and acoustic comfort have been developed. They are then used to produce soundscape quality maps, with a case study site as an example.

II. SOUND LEVEL AND ACOUSTIC COMFORT EVALUATION IN URBAN OPEN SPACES

A. Field surveys

The results of a series of field surveys have been used to gather the data for ANN models. The surveys were first undertaken in two urban open spaces in each of the following seven cities in Europe: Athens, Thessaloniki, Milan, Fribourg, Cambridge, Sheffield, and Kassel. Parallel surveys were carried out in five Chinese urban open spaces, two in Beijing, and three in Shanghai. In total over 9000 people in Europe and over 800 people in China were interviewed. The questionnaires[3,8,28] were initially developed in English, and then translated into other languages.[29] Table I summarizes the surveys in the 19 case study sites. It can be seen that they were generally located in four kinds of area, namely, city center, residential area, tourist spot, and railway station.

Since the emotional impact of an environment is basically contributed from two attributes, namely, the explicit attributes of physical surroundings and the implicit attributes of social aspects,[30,31] the investigated factors in the above

[a)]Author to whom correspondence should be addressed. Electronic mail: j.kang@sheffield.ac.uk

TABLE I. Overview of the case study sites.

		Case study sites			
Country	City	Code	Site name	No. of interviewees	Location
Germany	Kassel	1	Bahnhofsplatz	418	Railway station
		2	Florentiner	406	Tourist spot
Greece	Athens	3	Karaiskaki	655	City center
		4	Seashore	848	Tourist spot
	Thessaloniki	5	Kritis	777	Residential area
		6	Makedonomahon	1037	City center
Italy	Milan	7	IV Novembre	574	City center
		8	Piazza Petazzi	599	Residential area
Switzerland	Fribourg	9	Jardin de Perolles	888	Residential area
		10	Place de la Gare	1041	Railway station
United Kingdom	Cambridge	11	All Saint's Garden	459	Tourist spot
		12	Silver Street	489	Tourist spot
	Sheffield	13	Barkers Pool	499	City center
		14	Peace Gardens	510	City center
China	Beijing	15	Chang Chun Yuan Square	307	Residential area
		16	Xi Dan Square	304	City center
	Shanghai	17	Century Square	62	Tourist spot
		18	Nanjing Road Square	79	City center
		19	Xu Jia Hui Park	79	Residential area

field surveys are categorized accordingly, as shown in Table II, where the units and categories of measured and observed parameters and the scales for subjective evaluations are also given. The attributes are further classified as physical, behavioral, social/demographic, and psychological elements, for the convenience of analysis. For the subjective evaluation of acoustic comfort, five scales were used, from −2 (very uncomfortable) to 2 (very comfortable).

In addition to the factors directly relating to acoustic characteristics, a number of factors relating to other physical conditions and the subjective responses to these conditions were also considered, since these factors could affect people's mood, which would in turn be important for people to make judgments. For instance, a subject who is suffering from hot weather may evaluate soundscapes differently from someone who is under a comfortable thermal condition. More significantly, aural and visual interactions have been demonstrated in numerous studies.[3,32–34]

The physical factors listed in Table II were measured immediately before or after each interview using a multisensor measurement station, where the sound pressure level (SPL) was based on a 1 min measurement.

It is noted that in this study, the term "sound level" evaluation has been used, instead of using "loudness," "noisiness," "tranquility," or "quietness." This is because loudness and noisiness have rather specific definitions and calculation methods, referring to subjectively perceived sound and noise levels.[16–18] Tranquility and quietness often emphasize the positive aspect of a quiet environment. Sound level evaluation seems to be rather neutral, which is more appropriate for this study.

B. Analysis

The main objective of the analysis is to select suitable input variables for the ANN models. As ANN has a robust learning capability to model nonlinear relationships, many input variables can be used in the network as far as there are sufficient training samples. However, since the sample sizes vary considerably among different case study sites, it is important to limit the input variables so that the network size can be kept reasonable for a good prediction. On the other hand, if the input variables are selected too strictly, namely, only those factors that are highly related to the output are used, the advantage of ANN modeling compared to a simple multiple regression would not be significant. As a result, in this study, the level of significant correlations or differences, namely, p values derived from statistical analyses, is used for limiting the input variables, whereas a threshold in terms of correlation coefficient is not applied since this may limit the number of input variables too strictly. Such methods have also been used in other studies.[35]

The statistical analyses of the field survey results have been made using SPSS,[36] considering the Pearson/Spearman correlations (two-tailed) and Pearson chi-square for factors with 3+ scales, namely, taking both linear and nonlinear correlations into account; as well as mean differences (t-test, two-tailed) for factors with 2 scales.

In the following analysis both the sound level evaluation and acoustic comfort evaluation are regarded as output and the relating factors to each of them are examined. It is noted that for the acoustic comfort evaluation, the sound level evaluation is also used as an input variable. This is because

TABLE II. Factors considered in the surveys.

Attributes	Elements	Ref. No.	Attribute factors	Measures of the attributes
Explicit	Physical	Phy1	Season	1—winter; 2—autumn; 3—spring; 4—summer
		Phy2	Time of day	1—night > 21:00pm–8:59am; 2—evening: 18.00–20.59pm; 3—morning: 9.00am–11.59am; 4—afternoon: 15.00–17.59pm; 5—midday: 12.00–14.59pm
		Phy3	Air temperature	Measurement of air temperature: °C
		Phy4	Wind speed	Measurement of wind speed: m s^{-1}
		Phy5	Relative humidity	Measurement of relative humidity: %
		Phy6	Horizontal luminance	Measurement of horizontal luminance: Klux (Europe); lux (China)
		Phy7	Sun shade	0—interviewee not standing in the sun; 1—interviewee standing in the sun
		Phy8	Sound pressure level	Measurement of sound pressure level: dB(A)
	Behavioral	B1	Whether wearing earphones	0—not wearing earphone; 1—wearing earphone
		B2	Whether reading or writing	0—neither reading nor writing; 1—either reading or writing
		B3	Whether watching somewhere	0—not watching anywhere; 1—watching somewhere
		B4	Movement status	1—sitting; 2—standing; 3—playing with kids; 4—sporting
		B5	Frequency of coming to the site	Scales 1–5; 1—first time; 5—every day
		B6	Reason for coming to the site	1—the equipment/services of the site; 2—children playing and social meetings; 3—business/meeting/break; 4—attending social events; 5—passing by
		B7	Grouping: whether accompanied	0—none; 1—with 1 person; 2—with more than 1 person
Implicit	Social/demographic	S1	Age	1 < 12; 2 = 12–17; 3 = 18–24; 4 = 25–34; 5 = 35–44; 6 = 45–54; 7 = 55–64; 8 > 65
		S2	Gender	1—male; 2—female
		S3	Occupation	1—students; 2—working people; 3—others (e.g., unemployed and pensioners)
		S4	Education	1—primary; 2—secondary; 3—higher education
		S5	Residential status	0—nonlocal; 1—local
		S6	Sound level experience at home	Scales –2 to 2, with –2 as very quiet and 2 as very noisy
	Psychological	Psy1	Site preference	0—not like the site for certain reasons; 1—like the site
		Psy2	View assessment	Scales –1 to 1, with –1 as negative and 1 as positive
		Psy3	Heat evaluation	Scales –2 to 2, with –2 as very cold and 2 as very hot
		Psy4	Wind evaluation	Scales –2 to 2, with –2 as stale and 2 as too much wind
		Psy5	Humidity evaluation	Scales –2 to 2, with –2 as very damp and 2 as very dry
		Psy6	Brightness evaluation	Scales –2 to 2, with –2 as very dark and 2 as very bright
		Psy7	Overall physical evaluation	0—not comfortable; 1—comfortable
		Psy8	Sound level evaluation	Scales –2 to 2, with –2 as very quiet and 2 as very noisy

the former is regarded as the final outcome of soundscape quality, whereas the latter is one of its components. Correspondingly, in ANN modeling, the sound level evaluation model could be used as a sub-model of the acoustic comfort model.[3]

C. Factors relating to sound level evaluation

Based on the statistical analysis of the field survey results, factors relating to the sound level evaluation are summarized in Table III,[37,38] where marks * and ** indicate significant correlations or differences, with * representing $p<0.05$ and ** representing $p<0.01$; and the (—) signs indicate missing data. It can be seen that the relationship between the SPL (Phy8) and the subjective evaluation of sound level is rather strong as 15 out of 19 case study sites have been found with a significant level, marked with * or **, and the r values are generally greater than 0.30. The air temperature (Phy3), wind speed (Phy4), relative humidity (Phy5), and horizontal luminance (Phy6) are also important, with significant values in 5, 8, 7, and 7 of 19 case study sites, respectively. In comparison, the other physical factors, including season (Phy1), time of day (Phy2), and sun shade (Phy7), are less important for the sound level evaluation.

In Table III, it can be seen that some behavioral factors, including whether watching somewhere (B3), frequency of coming to the site (B5), and reason for coming to the site (B6), should be considered for the sound level evaluation due to the significant levels in over 5 of 19 case study sites. However, the importance of other behavioral factors, including B1 (whether wearing earphones), B2 (whether reading or writing), and B7 (grouping: whether accompanied), is very limited.

In terms of social/demographic factors, occupation (S3), education (S4), and the sound level experience at home (S6) are rather closely related to the sound level evaluation because at least 6 of 19 case study sites have significant values; while for other social/demographic factors, including age (S1), gender (S2), and residential status (S5), the importance

TABLE III. Effects of various factors on the sound level evaluation, in terms of the significance levels of the Pearson/Spearman correlations (two-tailed) for factors with 3+ scales except B6, where Pearson chi-square is used and the level of significance is shown in the table; and mean differences (t-test, two-tailed) for factors with 2 scales. Marks * and ** indicate significant correlations or differences, with * representing $p<0.05$ and ** representing $p<0.01$. (–) indicates missing data.

	Physical factors							
Site	Phy1	Phy2	Phy3	Phy4	Phy5	Phy6	Phy7	Phy8
1	0.04	0.04	0.09	0.02	−0.02	0.02	0.06	0.11(*)
2	0.05	0.05	0.11(*)	−0.09	−0.12(*)	0.03	0.07	0.21(**)
3	−0.25(**)	−0.26(**)	−0.23(**)	0.08(*)	0.21(**)	0.19(**)	−0.02	0.30(**)
4	−0.24(**)	−0.25(**)	−0.15(**)	−0.07(*)	0.07(*)	−0.05	−0.10	0.27(**)
5	0.04	0.03	0.02	−0.08(*)	−0.02	−0.16(**)	0.04	0.06
6	−0.10(**)	−0.10(**)	−0.05	0.05	−0.04	0.06	−0.05	0.14(**)
7	0.03	0.04	0.01	−0.00	0.05	0.13(**)	0.17	0.07
8	0.05	0.04	0.06	−0.12(**)	0.04	−0.10(*)	0.05	0.17(**)
9	−0.01	−0.01	−0.01	−0.07(*)	0.09(**)	−0.11(**)	0.18(**)	0.22(**)
10	0.03	0.03	0.06(*)	0.10(**)	−0.08(**)	0.07(*)	−0.11(*)	0.14(**)
11	0.02	0.03	−0.06	−0.12(*)	0.11(*)	0.08	−0.18(*)	0.12(*)
12	−0.07	−0.05	−0.12(**)	−0.08	0.01	−0.03	−0.21(*)	0.06
13	0.02	0.02	−0.04	−0.02	−0.05	0.03	−0.10	0.30(**)
14	−0.14(**)	−0.13(**)	−0.06	−0.06	0.04	0.07	0.06	0.43(**)
15	–	−0.02	−0.04	−0.15(**)	0.19(**)	0.08	–	−0.01
16	–	−0.08	−0.04	−0.01	0.02	−0.10	–	0.11
17	–	−0.01	−0.19	0.21	0.06	−0.11	–	0.77(**)
18	–	–	−0.01	0.12	0.04	−0.03	–	0.79(**)
19	–	−0.10	0.06	0.01	0.02	0.30(**)	–	0.80(**)
% of Sig.	28.6	22.2	26.3	42.1	36.8	36.8	28.6	73.7
Ratio	4/14	4/18	5/19	8/19	7/19	7/19	4/14	15/19

	Behavioral factors						
Site	B1	B2	B3	B4	B5	B6	B7
1	0.02	−0.05	0.07	−0.01	0.05	0.44	0.01
2	–	−0.02	0.08	−0.10(*)	0.08	0.01(*)	0.03
3	–	0.25	0.13(*)	0.05	0.00	0.06	0.07
4	–	0.12	0.30(**)	0.06	−0.10(**)	0.00(**)	0.12(**)
5	–	0.23	−0.31(**)	0.03	−0.11(**)	0.37	−0.03
6	0.26	0.26(*)	0.12	−0.03	0.10(**)	0.01(**)	−0.02
7	–	−0.16	−0.18(*)	0.02	0.15(**)	0.02(*)	−0.04
8	–	0.30	0.05	−0.12(**)	0.01	0.04(*)	0.10(*)
9	−0.07	−0.21	0.01	−0.04	−0.07(*)	0.78	0.00
10	0.38	0.07	0.00	0.00	0.00	0.55	−0.09(**)
11	–	0.15	0.20(*)	0.05	−0.06	0.00(**)	0.03
12	–	0.22	0.48(**)	0.06	−0.00	0.01(**)	−0.05
13	–	−0.20	0.21	−0.03	−0.00	0.76	0.06
14	–	0.46(**)	−0.49(**)	0.02	−0.04	0.83	0.10(*)
15	–	0.15	−0.08	−0.04	−0.10	0.40	0.09
16	–	−0.35	0.54	−0.06	−0.01	0.10	−0.10
17	–	−0.33	0.23	−0.05	0.13	–	0.00
18	–	0.16	0.14	−0.05	0.05	–	0.11
19	–	0.20	−0.09	0.00	−0.05	–	−0.04
% of Sig.	0.0	10.5	36.8	10.5	26.3	43.8	21.1
Ratio	0/4	2/19	7/19	2/19	5/19	7/16	4/19

	Social/demographic factors					
Site	S1	S2	S3	S4	S5	S6
1	−0.04	−0.15(*)	0.02	0.09	−0.12	−0.12(*)
2	−0.09	0.04	−0.04	0.02	−0.05	−0.06
3	0.07	−0.03	0.11(**)	−0.04	−0.11	0.05
4	−0.12(**)	0.00	−0.10(**)	0.06	0.06	0.09(*)
5	−0.06	−0.01	−0.11(**)	0.07(*)	0.08	0.07(*)
6	−0.12(**)	0.17(**)	−0.12(*)	0.12(**)	0.02	−0.01

TABLE III. (*Continued.*)

	Social/demographic factors					
Site	S1	S2	S3	S4	S5	S6
7	0.07	0.05	0.06	0.00	−0.05	0.02
8	−0.01	−0.08	0.00	−0.04	−0.20	0.18(**)
9	−0.09(**)	−0.03	−0.12(**)	0.12(**)	0.13(*)	0.04
10	0.04	−0.17(**)	0.03	0.08(**)	−0.09	0.03
11	−0.04	−0.04	−0.19(**)	0.06	0.08	−0.02
12	−0.06	−0.15	−0.01	0.10(*)	0.34(**)	−0.03
13	−0.00	−0.01	−0.03	−0.03	−0.08	0.33(**)
14	−0.12(**)	0.05	−0.12(**)	0.10(*)	−0.16	0.49(**)
15	0.00	−0.09	0.02	0.03	0.02	0.05
16	0.08	−0.01	−0.11	0.17(**)	−0.05	0.03
17	−0.06	−0.22	0.01	0.09	−0.33	−0.12
18	0.07	−0.05	−0.11	0.03	0.03	−0.15
19	0.09	0.09	0.07	0.23	0.30	0.21
% of Sig.	21.1	15.8	36.8	36.8	10.5	31.6
Ratio	4/19	3/19	7/19	7/19	2/19	6/19

	Psychological factors						
Site	Psy1	Psy2	Psy3	Psy4	Psy5	Psy6	Psy7
1	−0.12	−0.10(*)	0.04	−0.04	−0.05	−0.01	0.10
2	−0.04	−0.04	0.07	−0.10(*)	0.03	−0.02	−0.06
3	−0.19(**)	0.03	0.10(**)	−0.17(**)	−0.13(**)	0.09(*)	0.15
4	0.04	−0.04	−0.01	−0.08(*)	−0.09(*)	0.01	0.15
5	0.16(*)	−0.16(**)	−0.08(*)	0.08(*)	−0.07	−0.13(**)	0.26(**)
6	−0.31(**)	−0.18(**)	0.03	0.12	0.04	−0.09(*)	0.26(**)
7	−0.30(**)	0.03	−0.09(*)	0.10(*)	−0.02	0.05	−0.09
8	−0.15(*)	−0.05	0.05	−0.01	0.09(*)	−0.23(**)	0.26(**)
9	−0.14(*)	−0.07(*)	−0.04	−0.05	−0.05	−0.07	−0.03
10	−0.28(**)	−0.13(**)	0.09(**)	0.05	0.04	0.03	−0.04
11	−0.07	0.02	0.04	−0.05	0.09	−0.03	0.18
12	−0.21(*)	−0.04	−0.10(*)	0.02	0.05	0.06	−0.01
13	−0.13	−0.11(*)	−0.06	0.05	0.05	−0.02	−0.15
14	−0.25(**)	−0.09	−0.02	−0.11	−0.11	0.03	−0.01
15	–	−0.09	0.11	−0.12(*)	−0.07	−0.07	0.44(**)
16	–	−0.13(*)	0.15(**)	−0.01	−0.03	−0.19(**)	0.45(**)
17	–	−0.29(*)	0.13	0.35(**)	−0.09	−0.27(*)	−0.20
18	–	−0.14	0.12	−0.01	−0.14	−0.10	0.23
19	–	−0.03	−0.04	−0.03	0.06	−0.17	1.78(*)
% of Sig.	64.3	42.1	31.6	36.8	15.8	31.6	31.6
Ratio	9/14	8/19	6/19	7/19	3/19	6/19	6/19

for the sound level evaluation is limited as only 4, 3, or 2 case study sites have significant values, respectively.

The correlations between the psychological factors and the sound level evaluation are generally significant, as in terms of Psy1 (site preference), Psy2 (view assessment), Psy3 (heat evaluation), Psy4 (wind evaluation), Psy6 (brightness evaluation), and Psy7 (overall physical evaluation), 9 of 14, 8 of 19, 6 of 19, 7 of 19, and 6 of 19 case study sites show significant values, respectively.

D. Factors relating to acoustic comfort evaluation

The factors relating to the subjective evaluation of acoustic comfort were studied in seven case study sites in Sheffield and China. The importance of physical factors on the acoustic comfort evaluation is very limited except Phy8 (SPL) as significant correlations have been found in five of seven case study sites. Similarly, the behavioral and social/demographic factors also have limited importance for the acoustic comfort evaluation. The importance of psychological factors for the acoustic comfort evaluation is, however, much greater. In all the case study sites, the acoustic comfort evaluation is statistically significantly related to Psy2 (view assessment) and Psy6–8 (brightness, overall physical evaluation, and sound level evaluation), although for Psy1, 3, 4, and 5 (site preference, evaluation of thermal, wind, and humidity), the importance for the acoustic comfort evaluation is rather limited.

Overall, from the results of the sound level evaluation and acoustic comfort evaluation it can be seen that the relating factors are rather different in various case study sites.

III. ANN MODELING

For urban planners and designers it is important to communicate with the users.[39] In this section, therefore, the evaluations of users are simulated by computer models, representing a common attitude shared by a similar user group when using urban open spaces. Two models are considered, one for the subjective evaluation of sound level, and the other for the overall acoustic comfort evaluation.

A. Modeling framework based on ANN

ANN is a computer learning system, which can model nonlinear relationships. It was first provided by McCulloch and Pitts[40] in the theory of the processing unit. The mechanism is based on an understanding of the biological brain's working patterns. It introduces an idea of using silicon logic gates as microprocessors to mimic the brain structure. It can make predictions to similar questions from which they learnt.[41] The learning process of ANN is to develop weights between its processing elements, including various input nodes, hidden nodes, and output nodes. The weights stress the response strength. In the whole training process, the weights are constantly adjusted to reduce the differences between desired and actual responses. The adjustment process is not stopped until there is no further significant change. Based on its learning capability, ANN is considered to be suitable to build prediction models for soundscape evaluation.[20,21,42]

Two ANN software packages, QNET[43] and NEUROSOLUTIONS,[44] have been used in this study. Since their results generally show similar trendencies,[28] the results presented in this paper are based on the former only. QNET is a back-propagated multilayer forward neural modeling system. It has been successfully applied in a previous study to predict reverberation times and sound levels in urban open spaces.[45] In QNET the accuracy of prediction is determined by the training and test correlation coefficients, as well as the root-mean-square error (RMSE) considering both training and test. The correlation coefficients are calculated between the network outputs and the targets from actual data. The RMSE, and also the root-mean-square deviation, are to measure the differences between outputs and targets. ANN model could be trained well with a large number of training examples, but normally at least 10% of them have to be set aside for testing the model performance.[46]

In the soundscape ANN models presented below, for both sound level evaluation and acoustic comfort evaluation, the input variables include the subjective evaluations of physical conditions such as temperature, wind, humidity, and brightness, given the multiple relationships between various factors. While this is possible based on the field surveys in the model development in this study, if the ANN models are to be used at the design stage, those input variables will not be available. However, there have been established relationships between these physical conditions and their subjective evaluations,[29] which can be used in the soundscape ANN models at the design stage. Moreover, based on the field surveys, a series of ANN models can be developed for the subjective evaluation of these physical conditions, which could be used as sub-models for soundscape ANN models,[3] although this is not within the scope of this paper.

When constructing the ANN soundscape models, input variables were carefully selected based on the statistical analysis in Sec. II. The distributions of various factors, especially the outputs including the sound level evaluation and acoustic comfort evaluation, were examined and it was found that they were all distributed reasonably widely across the scale. Before processing any training, all data were normalized. In terms of model structure, based on pilot studies, the models below were designed with one input layer, one or two hidden layers, and one output layer.

B. Test of ANN model performance

Using the Makedonomahon Square in Thessaloniki (site 6) as an example, the prediction performance of ANN modeling was examined, considering the sound level evaluation. The 1037 samples were split into three sets, including training, test, and evaluation sets, taking 70%, 10%, and 20% of the total samples, respectively. The test set was for internally monitoring the network performance, whereas the evaluation set was for externally testing the network performance after it was fixed. After exploring a number of network structures, the optimal one was found to be with 2 hide layers and 11 hide nodes. The input variables included Phy1, 2, 8; B2, 5, 6; S1, 2, 3, 4; and Psy1, 2, 6, 7. The results showed that the training and test correlation coefficients were 0.69 and 0.44, and the training and test RMSE were 0.11 and 0.16, respectively, which are all acceptable.

In order to examine the performance of this model with new data, which were not used in the training process, the model was recalled and tested using the evaluation set. The paired-samples t-test was used to compare the predictions and the evaluation set, namely, 204 actual sample data from the field survey. The mean difference is only 0.068, and the difference is statistically insignificant, suggesting that the prediction performance of the ANN model was acceptable and reliable.

While in the case of Makedonomahon the sample size was rather large, and it was thus possible to use 20% of the samples for external testing of the model performance, for other case study sites, where the sample sizes were relatively small, it would be more efficient to use all the samples for model training. Therefore, in the following modeling process, no evaluation set was used.

C. Modeling the sound level evaluation

A general model was first explored, using the data from all the 19 case study sites, representing a variety of urban open spaces. According to the significance levels analyzed by combining all the case study sites into one data set, 16 factors were chosen as input variables, and a number of models using different hidden layers and nodes were constructed. However, none of them converged. This suggested that a general model including all kinds of urban open space was not feasible.

Models were therefore developed based on individual case study sites. Four sites from Europe were randomly cho-

TABLE IV. Model for the sound level evaluation based on individual sites.

Site	Network architecture					Results			
						Coefficient		RMSE	
	Input variables	Output variable	Hide layer	Hide node	Test sample size	Training	Test	Training	Test
2-Florentiner	Phy3, 5, and 8 B4 and 6 Psy4	Subjective evaluation of sound level	2	6	40	0.41	0.31	0.131	0.130
3-Karaiskaki	Phy1, 2, 3, 4, 5, 6, and 8 B3 S3 Psy1, 3, 4, 5, and 6		1	7	50	0.76	0.68	0.096	0.129
13-Barkers Pool	Phy8 S6 Psy2		2	4	50	0.45	0.41	0.123	0.142
14-PeaceGardens	Phy1, 2, and 8 B2, 3, and 7 S1, 3, 4, and 6 Psy1		2	7	50	0.72	0.61	0.109	0.138

sen, including the Kassel Florentiner Square (site 2), Athens Karaiskaki Square (site 3), Sheffield Barkers Pool (site 13), and Sheffield Peace Gardens (site 14). Table IV shows the network architecture and model results. As an example, in Fig. 1, the network architecture for the Kassel Florentiner Square (site 2) is illustrated. Again, each model was developed by exploring a number of model structures considering the number of hidden layers and hidden nodes, and the prediction results shown in Table IV are the ones from the optimal models. It can be seen that for the Athens Karaiskaki and Sheffield Peace Gardens models the predictions are rather good, with test coefficients over 0.6, but the models are less satisfactory for the Kassel Florentiner Square and Sheffield Barkers Pool, with test coefficients of about 0.3–0.4, which is possibly caused by the lower correlations between input and output variables.

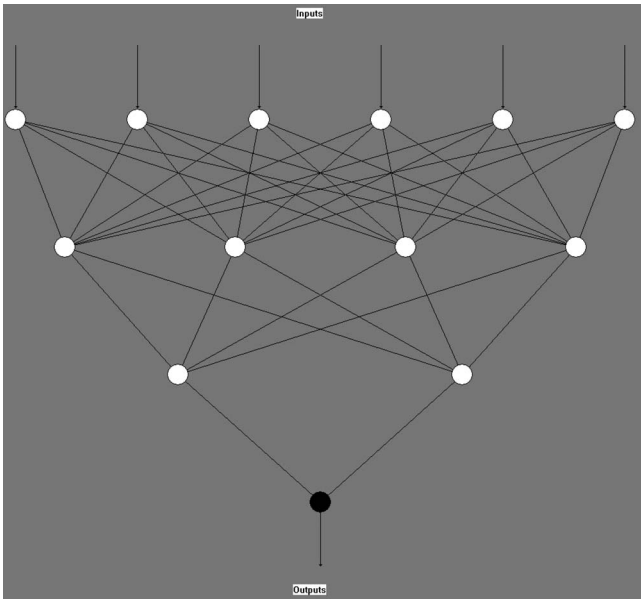


FIG. 1. Architecture of the sound level evaluation model for the Kassel Florentiner Square (site 2).

Encouraged by the results of individual models, another general model was explored using the common significant factors for all the 19 case study sites, where a factor was selected as an input variable if it reached a significant level for the sound level evaluation in at least 4 case study sites. For the five Chinese sites the data were combined due to the relatively small sample sizes for each case study site, and also, the input variables were slightly different since the questionnaire was modified. Again, a number of model architectures were explored, and the results from optimal models are shown in Table V. It can be seen that the test coefficients range from 0.22 to 0.66, generally considerably lower than those in Table IV, suggesting that such a general model may not be feasible.

Efforts were then made to develop models for urban open spaces with similar characteristics. The 19 case study sites were first classified as 4 types according to their locations/functions—as shown in Table I, 7 case study sites were located in city centers, 5 in residential areas, 5 at tourist spots, and 2 near railway stations. For each type some case study sites were grouped according to their city/country/continent. In total, eight models were developed, as shown in Table VI. For each model, the input variables were selected according to the significant levels of relevant factors in the case study sites in the group. From Table VI it can be seen that the best prediction result was achieved in the model with two Cambridge sites, which were both tourist spots. The test coefficient is 0.6, much higher than that based on all the five case study sites in Europe and China, which is only 0.31. For the three city center models based on Sheffield, Greece, and China, the test coefficients are 0.52, 0.48, and 0.45, respectively, which are also satisfactory.

Overall, the above results on the subjective evaluation of sound level suggest that a general model for all the case study sites is not feasible due to the complex physical and social environments in urban open spaces. Models based on individual case study sites perform well but their application

TABLE V. General sound level evaluation model.

	Network architecture					Results			
						Coefficient		RMSE	
Site	Input variables	Output variable	Hide layer	Hide node	Test sample size	Training	Test	Training	Test
1-Bahnhofsplatz			1	3	38	0.58	0.29	0.09	0.127
2-Florentiner			1	3	10	0.63	0.22	0.114	0.140
3-Karaiskaki			1	5	44	0.82	0.61	0.09	0.126
4-Seashore	Phy1, 2, 3, 4, 5, 6, 7, and 8		1	4	65	0.68	0.35	0.120	0.164
5-Kritis			1	8	76	0.68	0.46	0.114	0.144
6-Makedonomahon	B3, 5, 6, and 7		1	4	97	0.57	0.44	0.170	0.191
7-IV Novembre			1	4	50	0.61	0.40	0.135	0.173
8-Piazza Petazzi	S1, 3, 4, and 6		1	6	52	0.76	0.66	0.102	0.147
9-Jardin de Perolles		Subjective evaluation of sound level	2	5	74	0.49	0.30	0.122	0.126
10-Place de la Gare	Psy1, 2, 3, 4, 6, and 7		2	6	86	0.52	0.28	0.123	0.144
11-All Saint's Garden			1	3	25	0.86	0.55	0.08	0.146
12-Silver Street			1	2	23	0.73	0.53	0.102	0.125
13-Barkers Pool			1	2	25	0.64	0.22	0.103	0.141
14-Peace Gardens			1	2	25	0.79	0.47	0.103	0.157
China sites combined	Phy2, 3, 4, 5, 6, and 8 B3, 5, and 7 S1, 3, 4, and 6 Psy2, 3, 4, 6, and 7		1	5	60	0.70	0.53	0.09	0.118

range is limited. Specific models for certain types of location/function could be reliable and also practical.

D. Modeling the acoustic comfort evaluation

The acoustic comfort evaluation was only made for seven case study sites, two in Sheffield, and five in China. Similar to the model development for the sound level evaluation, a general model was first explored. Two approaches were used to select the input variables. Model 1 combined all the seven case study sites as one data set and factors were selected as input variables if they reached the significant level. In model 2 factors were selected as input variables if a significant level was achieved in at least two case study sites. Using the same input variables, two kinds of general model were made within model 2, one was for each case study site, except the three Shanghai sites, which were combined due to the small sample sizes, and the other was for the combined data from all the case study sites. In Table VII the network architecture and prediction results are shown for the two general models. It can be seen that the test coefficients are generally satisfactory, but are rather low for certain sites.

Models based on individual case study sites were then developed to further examine the prediction performance. Two typical case study sites were chosen, Sheffield Peace Gardens (site 14) and Beijing Xi Dan Square (site 16). In Table VIII the optimal network architecture and prediction results are shown. It can be seen that the prediction results are rather good for both models, especially for the Peace Gardens model, where the test coefficient reaches 0.79 and the RMSE is only 0.103.

Similar to the procedure for the sound level evaluation, acoustic comfort evaluation models were also developed based on case study sites with similar locations/functions. Two models were built for the city center locations, one for the two case study sites in Sheffield and the other for two case study sites in China. The optimal network architecture and prediction results are shown in Table IX. It can be seen that both models have rather good prediction performance. For the Sheffield model, compared to the individual model of the Peace Gardens, the test coefficient becomes slightly lower, by 0.11, whereas for the Chinese model, compared to the individual model of the Xi Dan Square (see Table VIII), the test coefficient is the same.

Compared to the sound level evaluation models, the prediction performance of acoustic comfort models is considerably better. This might mainly be caused by the role of input variables. Some factors are strongly correlated with the acoustic comfort evaluation, including Phy8, Psy2, and Psy6–8, whereas for the sound level evaluation only Phy8 is closely related. On the other hand, a considerable number of other factors are not much correlated with the acoustic comfort evaluation, whereas this is not the case for the sound level evaluation.

E. Comparison with OLR

The ordinal logistic regression (OLR) is a nonlinear statistical modeling technique. It is processed with logistic functions to yield a probability of a single output.[47] OLR has been successfully applied in many areas.

To compare with ANN models, a typical case study site, the Peace Gardens in Sheffield (site 14), was used to establish two OLR models, one for predicting the sound level evaluation and the other for predicting the acoustic comfort evaluation.[38] In the OLR models, the input variables were the same as those in the ANN models, and five outputs were obtained corresponding to the five evaluation scales. The average prediction accuracy over the five scales was 53% for the sound level evaluation model and 61% for the acoustic

TABLE VI. Sound level evaluation model in terms of locations/functions.

		Network architecture						Results			
								Coefficient		RMSE	
Location		Site	Input variables	Output variable	Hide layer	Hide node	Test sample size	Train	Test	Train	Test
City Centre	Sheffield	13-Barker Pool	Phy1, 2, 3, 4, 5, 6, 7, and 8 B3, 4, 5, 6, and 7	Subjective evaluation of sound level	2	10	110	0.58	0.52	0.145	0.164
		14-Peace Gardens	S1, 2, 3, 4, 5, and 6 Psy1 and 2								
	Greece	3-Karaiskaki	Phy1, 2, 6, 7, and 8 B7		1	3	80	0.60	0.48	0.121	0.130
		6-Makedonomahon	S3 Psy1 and 2								
	China	16-Xi Dan Square	Phy2, 3, 4, 6, 7, and 8 B2, 3, 4, and 5		1	4	60	0.58	0.45	0.101	0.130
		18-Nanjing road Square	S4 Psy2								
Residential	EU	5-Kritis	Phy1, 3, 4, 6, 7, and 8		1	5	180	0.42	0.34	0.138	0.146
		8-Piazza Petazzi	B3 S1, 3, 4, 5, and 6								
		9-Jardin de Perolles	Psy1 and 2								
	China	15-Chang Chun Yuan Square	Phy4, 5, and 8 B5		1	7	30	0.74	0.38	0.075	0.173
		19-Xu Jia Hui Park	S6 Psy2, 3, 5, and 7								
Tourist	Cambridge	11-All Saint's Garden	Phy3, 5, 7, and 8 B2, 3, 4		1	7	90	0.73	0.60	0.111	0.126
		12-Silver Street	S1, 3, 4, and 5 Psy1								
	Europe+China	2-Florentiner 4-Seashore	Phy2, 3, 5, and 8		1	9	208	0.49	0.31	0.135	0.150
		11-All Saint's Garden	B2, 3, and 4								
		12-Silver Street	S1, 3, and 5								
		17-Century Square	Psy2								
Railway	Germany	1-Bahnhofsplatz	Phy4, 5, and 8 B3, 4, 6, and 7		2	11	110	0.48	0.35	0.118	0.132
	Switzerland	10-Place de la Gare	S2 and 4 Psy1 and 2								

TABLE VII. Overall models for acoustic comfort evaluation.

		Network architecture				Results				
							Coefficient		RMS	
	Site	Input variables	Output variable	Hide layer	Hide node	Test pattern	Train	Test	Train	Test
Model 1	13-Barkers Pool			2	6	100	0.60	0.56	0.119	0.121
	14-Peace Gardens	Phy3, 4, 5, and 8								
	15-Chuang Chun Yuan	B4 and 7								
	16-Xi Dan	S1, 4, 5, and 6								
	17-Century Square	Psy2, 4, 5, 6, 7, and 8								
	18-Nanjing Road Square									
	19-Xu Jia Hui Park									
Model 2	13-Barkers Pool		Subjective evaluation of acoustic comfort	1	4	24	0.80	0.63	0.089	0.117
	14-Peace Gardens	Phy5 and 8		1	3	25	0.76	0.58	0.010	0.125
	15-Chuang Chun Yuan	B7		1	4	30	0.71	0.59	0.082	0.124
	16-Xi Dan	S6		2	5	27	0.64	0.37	0.107	0.109
	Shanghai sites combined	Psy2, 3, 6, 7, and 8		1	2	15	0.71	0.35	0.085	0.136
	All seven sites combined			2	7	110	0.59	0.59	0.118	0.124

TABLE VIII. Acoustic comfort evaluation model based on individual sites.

	Network architecture					Results			
						Coefficient		RMSE	
Site	Input variables	Output variable	Hide layer	Hide node	Test sample size	Train	Test	Train	Test
14-Peace Gardens	Phy1, 2, 4, 5, and 8 B7 S6 Psy2–4 and 6–8	Subjective evaluation of acoustic comfort	2	7	30	0.90	0.79	0.07	0.103
16-Xi Dan Square	Phy3, 5 S3, 4 Psy2, 3, and 6–8		2	5	25	0.74	0.59	0.09	0.122

comfort evaluation model, although at certain scales the accuracy was rather low. While the OLR and ANN results were not directly comparable, generally speaking, with ANN models better predictions were obtained, with a test correlation coefficient of 0.61 for the sound level evaluation and 0.79 for the acoustic comfort evaluation, as can be seen in Tables IV and VIII, respectively. Similar to ANN models, with OLR models the accuracy for the acoustic comfort evaluation was also better than that for the sound level evaluation.

IV. SOUNDSCAPE QUALITY PREDICTION MAP

The above ANN models can predict the sound level and acoustic comfort evaluation of individual receivers/zones in an urban open space, given that the physical factors such as SPL and user profiles could vary at different positions. Sound level and acoustic comfort evaluation maps can be produced accordingly, which would be a very useful tool for urban planners and designers. Therefore, following the models established above, soundscape quality mapping techniques have been developed.

While it is evident that a universal ANN model for all kinds of urban open space is not appropriate/feasible and models based on the data of individual case study sites are too specific, it is proposed that urban open spaces should be classified into certain types, taking into account the functions and locations of the urban open spaces, and for each type an ANN model for soundscape quality mapping can be developed and applied in practice. In this section, however, for the sake of convenience, the ANN models based on the Peace Gardens in Sheffield (site 14), as shown in Fig. 2, is used below as an example to demonstrate the mapping technique, although the model actually should be used to produce soundscape quality maps for other urban open spaces with similar locations/functions as the case study site used in developing the model.

In Fig. 3 is shown the prediction maps for the sound level evaluation and acoustic comfort evaluation in the Peace Gardens in terms of two age groups, 13–18 and >65. The SPL of each cell shown in the figure, which was used in the ANN models as an input variable, was calculated using noise mapping software CADNA,[48] with similar source conditions as those in the field survey. In other words, to a certain extent, the mapping results should be regarded as predictions for new situations, rather than representations of the current situation. Other conditions were assumed to be the same between the two age groups and were based on the overall situation in the field survey, so that the result in each cell could be regarded as the average evaluation of each age group. From Fig. 3 it can be seen that the 13–18 age group will generally feel quieter than the age group >65, whereas in terms of acoustic comfort, the evaluation of the two age groups is very similar. This corresponds to the results from the field survey,[3,8,49,50] although direct comparison between them is inappropriate, since the predictions in Fig. 3 are for new situations.

TABLE IX. Acoustic comfort evaluation model in terms of locations/functions.

			Network architecture					Results			
								Coefficient		RMSE	
Location		Site	Input variables	Output variable	Hide layer	Hide node	Test sample size	Train	Test	Train	Test
City Centre	Sheffield	13-Barkers Pool	Phy1, 2, 5, and 8 B7	Subjective evaluation of acoustic comfort	2	7	50	0.74	0.68	0.104	0.105
		14-Peace Gardens	S6 Psy2, 3, and 5–8								
	China	16-Xi Dan Square	Phy8 B4		2	5	30	0.66	0.59	0.102	0.122
		18-Nanjing Road Square	S3 and 4 Phy2, 3, 4, and 6–8								

FIG. 2. (Color online) The Peace Gardens, Sheffield, United Kingdom (site 14).

V. CONCLUSIONS AND DISCUSSIONS

With the data based on large scale field surveys, the importance of various physical, behavioral, social, demographical, and psychological factors has been examined in terms of the subjective evaluation of sound level and acoustic comfort. ANN models for predicting subjective evaluation of soundscape in urban open spaces have then been explored at three levels, and it has been found that for both sound level and acoustic comfort evaluation, a general model for all the case study sites is less feasible due to the complex physical and social environments in urban open spaces, reflected by the widely varied subjective evaluation of soundscape as well as different input variables required by ANN models among different case study sites; models based on individual case study sites perform well but the application is limited; whereas specific models for certain types of location/function would be reliable and practical. The performance of acoustic comfort models is considerably better than that of sound level models, mainly due to the relative importance of various input variables for the ANN models. With ANN models soundscape quality prediction maps can be produced, as demonstrated through an example.

While the usefulness of the ANN models has been demonstrated, it is noted that the test coefficients are generally not very high. A possible reason is that subjective evaluations are rather varied between individual users and this cannot be completely represented by computer models. On the other hand, further improvements could be made by establishing a more complex model structure, where a number of sub-models can be developed, including the evaluation of sound level (background sound), sound preference (noticed sounds, foreground sounds, or soundmarks), and the evaluation of other physical factors, such as the satisfaction level of the thermal, lighting, view, and overall physical environment. The outputs of the sub-models can then be used as the input of an overall model for the acoustic comfort evaluation. For this further carefully designed field studies would be useful.

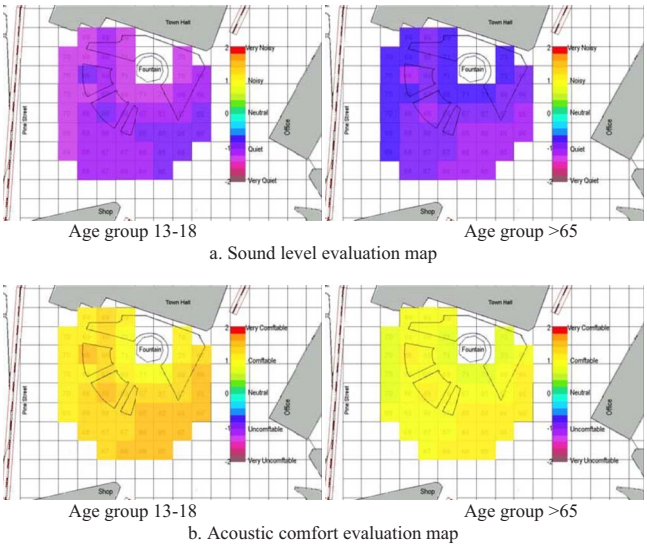


FIG. 3. (Color online) Prediction maps of sound level evaluation and acoustic comfort evaluation in the Sheffield Peace Gardens (site 14). The calculated SPL is shown in each cell.

ACKNOWLEDGMENTS

The data in this paper are mainly from two research projects funded by the European Commission and the British Academy. The authors are indebted to Dr. Robert F. Harrison, Dr. Mei Zhang, and Dr. Wei Yang, and other project partners for useful discussion.

[1]M. R. Schafer, *The Soundscape: Our Sonic Environment and the Tuning of the World* (Destiny Books, Rochester, VT, 1994).
[2]P. Hedfors and P. Grahn, "Soundscapes in urban rural planning and design," in *Northern Soundscapes—Yearbook of Soundscape Studies*, edited by R. M. Schafer and H. Järviluoma (Department of Folk Tradition, Tampere, Finland, 1998), Vol. 1, pp. 67–82.
[3]J. Kang, *Urban Sound Environment* (Taylor & Francis, London/Spon, London, 2006).
[4]B. Berglund, C. A. Eriksen, and M. E. Nilsson, "Exploring the perceptual content in soundscapes," in *Fechner Day*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst Science, Lengerich, Germany, 2001).
[5]A. Corbin, *Village Bells: Sound and Meaning in the 19th-Century French Countryside* (Columbia University Press, New York, 1998).
[6]T. J. Schultz, "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405 (1978).
[7]G. Bohme, "Acoustic atmosphere: A contribution to the study of ecological aesthetics," Soundscape: J. Acoustic Ecology **1**, 14–18 (2000).
[8]W. Yang and J. Kang, "Soundscape and sound preferences in urban squares: A case study in Sheffield," J. Urban Des. **10**, 69–88 (2005).
[9]D. Botteldooren and A. Verkeyn, "Fuzzy models for accumulation of reported community noise annoyance from combined sources," J. Acoust. Soc. Am. **112**, 1496–508 (2002).
[10]B. Schulte-Fortkamp, "The meaning of annoyance in relation to the quality of acoustic environments," Noise Health **415**, 13–18 (2002).
[11]B. Schulte-Forkamp, "Soundscape analysis in a residential area: An evaluation of noise and people's mind," Acta Acust. Acust. **92**, 875–880 (2006).
[12]B. De Coensel and D. Botteldooren, "The quiet rural soundscape and how to characterize it," Acta Acust. Acust. **92**, 887–897 (2006).
[13]M. E. Nilsson and B. Berglund, "Soundscape quality in suburban green areas and city parks," Acta Acust. Acust. **92**, 903–911 (2006).
[14]D. Dubois, C. Guastavino, and M. Raimbault, "A cognitive approach to urban soundscapes: Using verbal data to access everyday life auditory categories," Acta Acust. Acust. **92**, 865–874 (2006).
[15]C. Lavandier, "The contribution of sound source characteristics in the assessment of urban soundscapes," Acta Acust. Acust. **92**, 912–921 (2006).
[16]K. D. Kryter, *The Effects of Noise on Man* (Academic, New York, 1970).
[17]J. Goldstein, in *Community Noise*, edited by R. J. Peppin and C. W. Rodman (American Society for Testing and Materials, Philadelphia, PA, 1979), pp. 38–72.
[18]E. Zwicker and H. Fastl, *Psychoacoustics—Facts and Models* (Springer, Berlin, 1999).
[19]M. Zhang and J. Kang, "Subjective evaluation of urban environment: A case study in Beijing," Int. J. Environ. Pollut. (2009, in press).
[20]L. Yu and J. Kang, "Soundscape evaluation in city open spaces using artificial neural network," Proceedings of the UIA—XXII World Congress of Architecture, Istanbul, Turkey (2005).
[21]L. Yu and J. Kang, "Neural network analysis of soundscape in urban open spaces," J. Acoust. Soc. Am. **117**, 2591 (2005).

[22]L. Yu and J. Kang, "Integration of social/demographic factors into the soundscape evaluation of urban open spaces using artificial neural networks," Proceedings of the Inter-Noise, Honolulu, HI (2006).
[23]D. W. Patterson, *Artificial Neural Networks: Theory and Applications* (Prentice-Hall, Singapore, London, 1996).
[24]F. Phan, E. Micheli-Tzanakou, and S. Sideman, "Speaker identification using neural networks and wavelets," IEEE Eng. Med. Biol. Mag. **19**, 92–101 (2000).
[25]B. De Coensel, T. De Muer, and D. Botteldooren, "The influence of traffic flow dynamics on urban soundscapes," Appl. Acoust. **66**, 175–194 (2005).
[26]A. Verkeyn, D. Botteldooren, B. De Baets, and G. De Trè, in *Lecture Notes in Artificial Intelligence*, edited by T. Bilgic, B. De Baets, and O. Kaynak (Springer, Berlin, 2003), pp. 277–284.
[27]D. Botteldooren, A. Verkeyn, and P. Lercher, "A fuzzy rule based framework for noise annoyance modeling," J. Acoust. Soc. Am. **114**, 1487–1498 (2003).
[28]L. Yu, "Soundscape evaluation and ANN modelling in urban open spaces," Ph.D. thesis, School of Architecture, University of Sheffield, Sheffield, UK (2009).
[29]*Designing Open Spaces in the Urban Environment: A Bioclimatic Approach*, edited by M. Nikolopoulou (CRES, Attiki, Greece, 2004).
[30]A. P. Bell, C. T. Greene, D. J. Fisher, and A. Baum, *Environmental Psychology*, 4th ed. (Harcourt Brace College Publishers, Fort Worth, TX, 1996).
[31]G. Robert, *Environmental Psychology: Principles and Practice*, 2nd ed. (Allyn and Bacon, Boston, MA, 1997).
[32]J. L. Nasar, in *Public Places and Spaces*, edited by I. Altman and E. H. Zube (Plenum, New York, 1989).
[33]D. Alais, D. Burr, and S. Carlile, "Audiovisual interactions in the perception of space and time," Proceedings of the 19th International Congress on Acoustics, Madrid, Spain (2007).
[34]S. Yong-Gyu, J. Ji-Hyeon, K. Chan, and K. Sun-Woo, "Evaluation of human emotions depending on variations in audio-visual landscape elements in residual areas," Proceedings of the 19th International Congress on Acoustics, Madrid, Spain (2007).
[35]F. Y. Ling and M. Liu, "Using neural network to predict performance of design-build projects in Singapore," Build. Environ. **39**, 1263–1274 (2004).
[36]J. Pallant, *SPSS Survival Manual*, 2nd ed. (Open University Press, Buckingham, UK, 2005).
[37]L. Yu and J. Kang, "Effects of social, demographical and behavioral factors on the sound level evaluation in urban open spaces," J. Acoust. Soc. Am. **123**, 772–783 (2008).
[38]L. Yu, J. Kang, and R. Harrisons, "Mapping soundscape evaluation in urban open spaces with artificial neural networks and ordinal logistic regression," Proceedings of the 19th International Congress on Acoustics, Madrid, Spain (2007).
[39]A. Rapoport, *The Meaning of the Built Environment: A Nonverbal Communication Approach* (Sage, Beverly Hills, CA, 1982).
[40]W. S. McCulloch and W. H. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bull. Math. Biophys. **5**, 115–133 (1943).
[41]F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms* (Spartan Books, Washington, DC, 1962).
[42]B. Berglund, M. E. Nilsson, and P. Pekala, "Towards certification of indoor and outdoor soundscapes," Proceedings of the Inter-Noise, Prague, Czech Republic (2004).
[43]Vesta Services Inc., *Qnet Users Manual* (Winnetka, IL, 2000).
[44]NeuroDimension, Inc., *Neuro-Solutions5.0 Users Manual* (Gainesville, FL, 2008).
[45]L. Yu, "Predicting sound field and acoustic comfort in urban open spaces using neural networks," MSc thesis, School of Architecture, University of Sheffield, Sheffield, UK (2003).
[46]S. Rebano-Edwards, "Modelling perceptions of building quality—A neural network approach," Build. Environ. **42**, 2762–2777 (2007).
[47]L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models* (Springer-Verlag, New York, 1994).
[48]Data Kustik, *Cadna/A for Windows—User Manual* (Kustik, Munich, Germany, 2006).
[49]W. Yang and J. Kang, "Acoustic comfort evaluation in urban open public spaces," Appl. Acoust. **66**, 211–229 (2005).
[50]M. Zhang and J. Kang, "Towards the evaluation, description and creation of soundscape in urban open spaces," Environ. Plan. B: Plan. Des. **34**, 68–86 (2007).

Identifying acoustical coupling by measurements and prediction-models for St. Peter's Basilica in Rome

Francesco Martellotta
Dipartimento di Architettura e Urbanistica, Politecnico di Bari, via Orabona 4, 70125 Bari, Italy

(Received 27 May 2009; revised 3 July 2009; accepted 6 July 2009)

St. Peter's Basilica is one of the largest buildings in the world, having a huge volume resulting from the addition of different parts. Consequently, sound propagation cannot be interpreted using a conventional approach and requires experimental measures to be compared with statistical-acoustics and geometrical predictions in order to explain the interplay between shape, materials, and sound waves better. In previous research one of the most evident effects, the surprisingly low reverberation time, was believed to result from acoustical coupling phenomena. Taking advantage of more refined measuring techniques available today an acoustic survey was carried out and the results were analyzed using different methods, including Bayesian parameter estimation of multiple slope decays and directional energy plots, which showed that coupling effects actually take place, even though measured reverberation times were longer than those given in previous studies. In addition, experimental results were compared with geometrical- and statistical-acoustic models of the basilica, which showed that careful selection of input data and, in statistical models, the inclusion of phenomena such as direct sound radiation and non-diffuse energy transfer, allow obtaining accurate results. Finally, both models demonstrated that reduced reverberation depends more on increased absorption of decorated surfaces than on coupling effects.
 [DOI: 10.1121/1.3192346]

PACS number(s): 43.55.Br, 43.55.Gx [NX] Pages: 1175–1186

I. INTRODUCTION

Large reverberant spaces are frequently encountered in acoustical practice. Concert halls, auditoriums, and theaters may have volumes of tenths of thousands m^3, but the way the sound propagates inside them is well understood, provided that their shapes are reasonably proportionate.[1,2] Larger enclosed spaces are much less frequently found and, consequently, their acoustic properties are less investigated. Churches represent an important example of very large (and sometimes huge) buildings whose acoustic characteristics have been investigated since the early stages of acoustic science.[3,4] However, churches are not only large but, due to historical and architectural reasons, they also are complex, resulting from the combination of multiple volumes, which make sound propagation even more difficult to understand. In fact, large volumes are expected to be very reverberant, but the presence of richly decorated surfaces scattering the sound, combined with architectural complexity, mostly related to the presence of aisles, domes, or chapels, may sometimes contradict this simple expectation.

A remarkable example of this unusual behavior is St. Peter's Basilica in Rome, characterized by its huge volume of 480 000 m^3. In the early 1970s Shankland and Shankland[5] measured reverberation time in St. Peter's Basilica by means of a tape recorder and by ear and stopwatch. The results they obtained, although biased by the "subjective" approach, were quite surprising as they estimated a mid-frequency reverberation time of 7.1 s, a value lower than those observed in other smaller churches.[4–6] The authors explained this unexpected result as being the consequence of the weak acoustic coupling between the different parts of the church, a hypothesis that was later also used to explain similar discrepancies in Greek Byzantine churches.[7] Measurements carried out 20 years before, but with more refined techniques, in the basilicas of St. John Lateran and St. Paul outside the Walls in Rome[4] pointed out the existence of coupling phenomena in the first church, where different parts are connected by means of small apertures, also emphasizing (by comparing the two churches) the importance of sculptures and decorations in reducing the reverberation time. Measurements carried out in another of the largest worship buildings, St. Paul's Cathedral in London,[8] showed that despite its smaller dimensions (152 000 m^3) the measured reverberation time in unoccupied conditions is about 10.5 s. Later studies[9] showed, using a mathematical model, that coupling effects also take place in this church, confirming that the relationship between reverberation and coupling is not obvious and requires detailed investigation.

Even though the theoretical foundations of acoustic coupling were clearly stated by several researchers,[1,2,10] initially the identification of such phenomena from measured decay traces could only be made (in the best cases) by visual inspection.[4] Later on, the use of decay curves obtained from backward integrated impulse responses (IRs) (Schroeder plots) provided more detailed data, which, nonetheless, could hardly give reliable quantitative measures of the different decay rates. Concepts such as the "running reverberation"[9] or ratios between different portions of the reverberant decay[11] were used to describe the slope variation as a function of time. It was only the introduction of Bayesian parameter estimation[12–15] that opened up new perspectives in the research on acoustic coupling. In fact, in this way, a rigorous estimate of the different decay constants that characterize

multi-rate decay processes is possible, also allowing more accurate comparisons with theoretical models.

From this point of view the original two-room model[1] based on statistical-acoustics (SA) assumptions was first extended (by means of matrix notation) to a larger number of sub-rooms[2] and, more recently, the model was further generalized[16] by including a number of effects (such as direct sound radiation and non-diffuse transfer of energy), which allowed even more accurate analysis of the sound propagation in complex coupled systems. In addition, advancements in the geometrical-acoustic (GA) modeling[17] allowed improved prediction accuracy when using beam-axis/ray-tracing algorithms.

Taking advantage of the much more accurate measuring methods available today, allowing for three-dimensional (3D) sound field decomposition, flat frequency response, and high signal-to-noise ratios,[18] together with the possibility of better investigating coupling effects by means of Bayesian estimation, and to use these data to validate SA or GA models, a new survey was carried out in St. Peter's Basilica in order to analyze and distinguish the proportionate effects of surface absorption and acoustical coupling on reverberation.

II. THE ACOUSTIC SURVEY

A. Description of the church

The building of St. Peter's Basilica started in 1506, on the same site where the old Constantinian basilica had been built, in order to preserve the coincidence between the altar position and the tomb of the saint. The first architect appointed for the design of the new church was Bramante. After his death many others succeeded, but the current shape and the dome design were mostly due to Michelangelo's work, which started in 1547. The basilica was finally consecrated in 1626 after Carlo Maderno had modified Michelangelo's plans by adding the long nave (Fig. 1). Decorative works went on for several years. The basilica may be considered as the sum of five large volumes (the three braces of the original Greek cross, the main nave, and the domed crossing), joined together by means of the large openings of the crossing and by means of additional secondary volumes connected through smaller openings.

Its huge dimensions are characterized by a total length of 185 m, a nave width of about 26 m, a dome width of 41.5 m, and height (up to the lantern) of about 120 m. The resulting volume of the basilica, calculated by means of a computerized 3D model, is about 480 000 m^3, and the total surface area (assuming all the surfaces as flat) is about 80 000 m^2, 12 600 m^2 of which corresponds to the usable floor surface. Most of the surfaces are finished in plaster, marble, or stucco, but they are richly decorated with deep carvings (Fig. 2). The area of glazed surfaces is approximately 2000 m^2. Pews (with upholstered kneelers) are only installed in the choir, while the remaining floor areas are completely bare. At the center of the main nave a wooden barrier is installed to protect the central part of the floor.

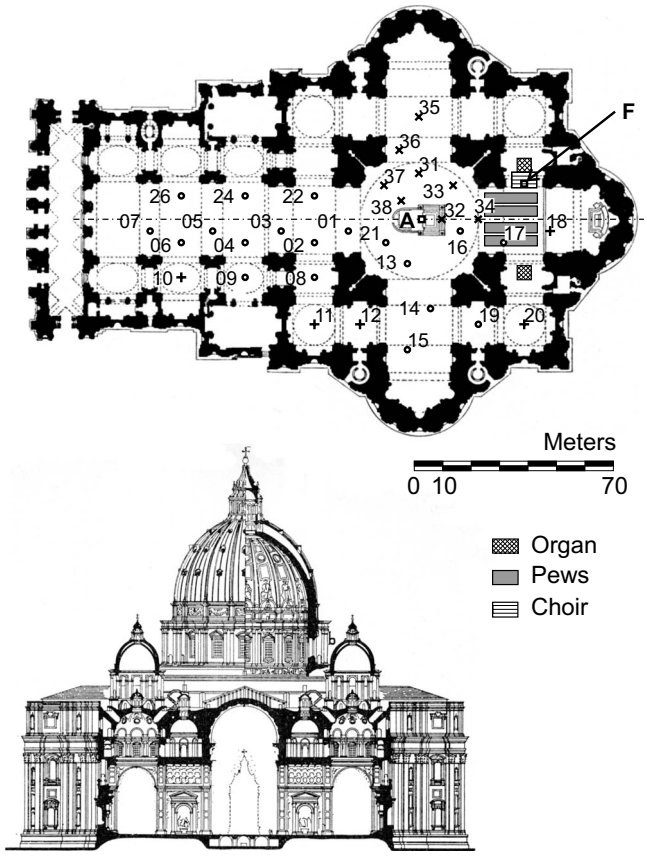


FIG. 1. Plan and section of the basilica, with indication of the source (A and F) and receiver (1–39) locations. (○) B-format microphone, (+) omnidirectional microphone, and (×) omni+figure-of-eight microphone.

B. Measurement techniques

Measurements were carried out in unoccupied conditions, during daytime between 9.00 and 10.30, just before the Wednesday Mass in St. Peter's Square. About 30 people (mostly security officers, priests, and other workers) could not leave the basilica during the measurements, but they were asked to perform only "silent" tasks in order to avoid interferences. Indoor air temperature was 21 °C, relative humidity was 60%, and both values remained constant during the measuring session. Measurements were carried out com-

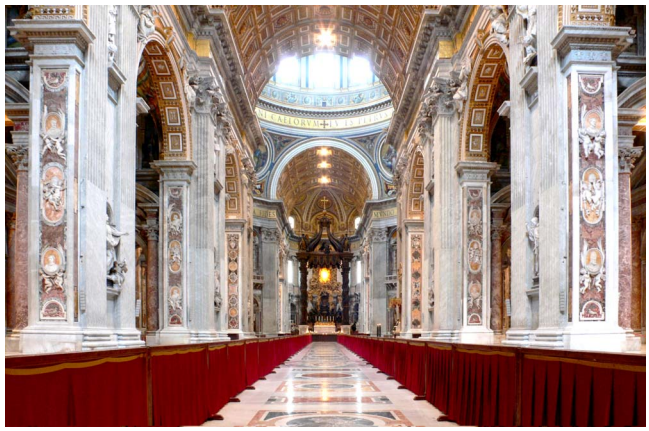


FIG. 2. (Color online) View of the interior of the church taken from the main entrance.

plying with ISO 3382 standard[19] also taking into account a set of guidelines specifically defined for churches.[20] Given the building dimensions and strict time limitations, the measurements were carried out using a special set-up. Two omnidirectional sound sources (a Look-Line D301 and a self-made dodecahedron made of twelve 120 mm loudspeakers), each one combined with an additional sub-woofer to cover low frequencies, were located, respectively, 2 m in front of the altar (position A) and in front of the organ on the left ("in cornu evangeli," position F), where there are also choir stalls installed. Source A was located 1.5 m above the altar floor, which is about 1.5 m higher than the nave floor. Source F was located 1.5 m above the risers of the choir stalls. Each sound source was fed by a different constant-envelope equalized sine sweep (40 s long) generated using MATLAB according to Müller and Massarani[18] so that the spectrum of the radiated sound was substantially flat from the 50 Hz to the 16 kHz third-octave bands. High-quality IRs were collected by using three measurement chains. The first chain included a B-format microphone (Soundifield Mk-V) and a binaural head and torso (B&K 4100D) connected to an Echo Audio Layla 24 sound card. The second chain included a Neumann TLM-127 with variable polar pattern, allowing the measurement of both omni-directional and figure-of-eight IRs, connected to an Echo Audio Fire 8 sound card. The third chain included an omni-directional microphone (GRAS 40-AR) connected to a portable digital audio tape recorder, used to get IRs in the farthest positions. In all the cases the room responses were recorded at a sampling rate of 48 kHz and 24 bit depth, to obtain, after deconvolution (performed using MATLAB), IRs with a signal-to-noise ratio, which even in the worst cases (receivers located more than 100 m from the source) allowed a "safe" calculation of reverberation time (T_{30}) based on at least 30 dB of decay even at the lowest frequencies. Globally, 33 receivers placed 1.2 m from the floor surface were distributed throughout the church according to the layout given in Fig. 1.

III. ANALYSIS OF EXPERIMENTAL RESULTS

A. Reverberation times

The analysis of experimental results started by taking into account conventional measures of T_{30} and early decay time (EDT) resulting from IRs measured in the locations distributed throughout the church in combination with both source positions (Fig. 1). The analysis of the average T_{30} as function of frequency (Fig. 3) shows the typical behavior observed in churches mostly finished in hard reflecting materials, characterized by long values at low frequencies (13.6 s) rapidly decreasing as the frequency grows due to air absorption, with a mid-frequency value of 9.9 s and a minimum of 2.2 s at 8 kHz. Average reverberation time is actually smaller than expected. In fact, given the huge volume, mid-frequency reverberation time should be well above 10 s (see, for example, St. Petronio Basilica in Bologna[6] having a volume of 170 000 m^3 and a mid-frequency T_{30} of 10.7 s). However, it is interesting to observe that measured values are longer than those reported in Ref. 5, especially at low frequencies.

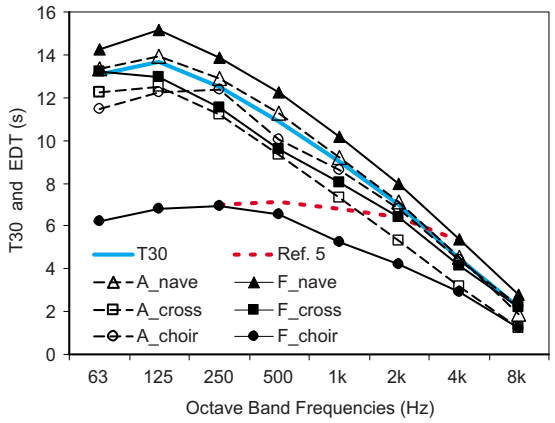


FIG. 3. (Color online) Octave-band values of average reverberation time (T_{30}) compared with measurements reported in Ref. 5, and EDT subdivided into different source and receiver locations.

Average EDT values as a function of frequency show negligible differences from T_{30}, being 0.2 s shorter at mid-frequencies. However, while standard deviation among measured T_{30} is very small (0.39 s at 1 kHz), the corresponding value for EDT is considerably larger (1.4 s at 1 kHz), suggesting a possible relationship with source placement and, above all, with source-receiver distance. When the source was in front of the altar (A) the average EDT values measured in the nave closely followed T_{30} behavior, while average EDT measured in the choir and in the crossing (the central volume where the braces converge) was slightly lower. The largest variations appeared when the source was located in the choir, showing average EDT values measured in the same subspace to be much lower (mid-frequency value being 5.9 s) than in the crossing (8.8 s), and in the nave (11.2 s). The latter were the largest values observed during the survey, being larger than those measured at the same receivers when the source was at A (10.2 s).

Given the measurement procedure used by Shankland and Shankland,[5] their reverberation times were expected to be more similar to EDT values (Fig. 3). However, apart from a better agreement at 1 and 2 kHz, moving toward low frequencies, no such relationship appears although the description of the measurement conditions seems substantially similar (no pews on the floor and no occupancy), and no significant changes have been made to the building since that survey. The discrepancy might be the result of different source placements, as values given in Ref. 5 correspond to both source and receivers in the nave, but they also give values measured in the crossing, and the small increase of just 0.2 s is well below the values measured in the present survey in the same conditions. So, it may be reasonably supposed that the "ear and stopwatch" method used in the measurements of the 1970s failed, especially at low frequencies, in such a huge volume where reaching adequate sound levels to measure T_{30} is a challenge even for today's instruments.

B. Analysis of spatial variations

A more detailed analysis of the nature of EDT differences was obtained by plotting mid-frequency T_{30} and EDT values as a function of source-receiver distance (Fig. 4). It

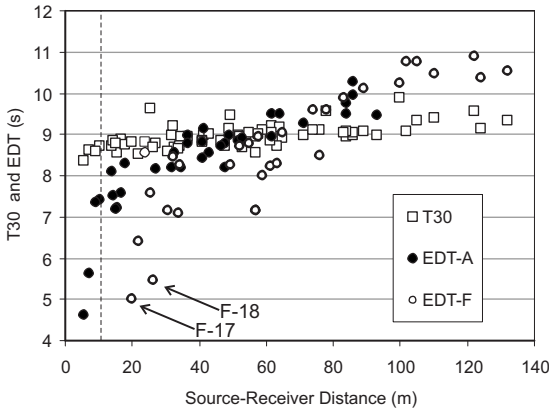


FIG. 4. Plot of T_{30} and EDT values at 1 kHz octave band, as a function of source-receiver distance. EDT values are further divided as a function of source placement.

can be observed that T_{30} shows only a mild increasing trend (with a slope of 0.7 s/100 m), while EDT shows a substantially different behavior, with a much steeper increase as a function of the distance, and a further dependence on the sound source location.

In fact, an increase in EDT as a function of distance is typically observed in churches as a consequence of the weak direct sound and early reflections arriving at the farthest points, which slow the transition toward the purely exponential decay of the diffuse sound energy.[21,22] However, in this case, two different trends appear, with a steeper slope when the source was in the choir (4.8 s/100 m) compared to the value observed when the source was in the crossing (2.7 s/100 m). Such a difference may depend on the larger number of obstacles that sound coming from the choir encounters during its propagation, affecting direct sound and early reflections. However, when both source and receiver are in the choir (and the direct sound path is unobstructed), EDT is markedly lower than at similar (equally unobstructed) combinations located in the crossing. This might well depend on early reflections differences, but a comparison of early decay traces (Fig. 5) shows that when both source and receiver are in the choir the initial part of the decay shows a long (about 8 dB) and markedly steeper slope. This provides evidence that EDT differences in the choir are not simply due to the

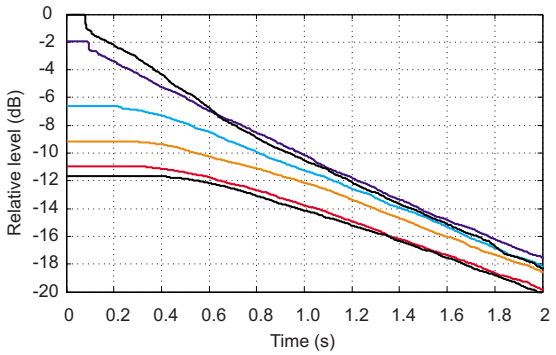


FIG. 5. (Color online) Plot of the early level decay for different receivers when the source is in the choir. Curves from top to bottom correspond to receivers 17, 21, 01, 03, 05, and 07. The topmost curve, corresponding to receiver 17, has a markedly different initial decay.

TABLE I. Summary of the results of Bayesian search of double slopes in measured IR.

S-R combination	E_{21} (dB)	$10\log(A_1/A_2)$ (dB)	T_1 (s)	τ_1 (s)	T_2 (s)	τ_2 (s)
F-17	97.0	1.0	3.2	0.044	8.9	0.151
F-18	133.6	3.2	3.3	0.058	8.9	0.108
F-34	105	−0.5	4.0	0.158	9.1	0.159
A-33	71.9	−5.0	1.3	0.071	8.4	0.071
A-21	68.0	−7.5	1.1	0.040	8.7	0.036
A-37	107.0	−3.0	0.8	0.025	8.3	0.051
A-38	66.3	−3.5	1.4	0.048	8.4	0.105

structure of early reflections but rather depend on diffuse-field variations between the given subspace and those to which it is connected.

C. Bayesian analysis

In order to investigate this hypothesis better, Bayesian analysis[12–15] provides a powerful and reliable tool to detect and quantify multiple slopes in IRs. Bayesian analysis is the most rigorous instrument for detecting double slopes in measured IRs and, hence, for investigating coupled-volume problems. The algorithm proposed by Xiang and Jasa[15] was implemented in MATLAB and several IRs were processed. Even though the procedure may be applied to a virtually unlimited number of decays, for the IR under test only double slopes gave the highest accuracy, evaluated by means of the Bayesian evidence E_{21}, as defined in Ref. 13. In order to create double slopes the best situation is when source and receiver are located in a subspace, which is less reverberant than the whole church. Consequently, source-receiver combinations located in the choir and in the crossing were investigated. Under these conditions the first (and shortest) decay (T_1) corresponds to the decay time of the subspace in which the source is located, while the second (T_2) corresponds to the coupled-system decay time. The relative amplitude of the two decays is evaluated by means of their logarithmic ratio as decay level difference. Finally, in order to estimate the accuracy of the decay time estimation, the procedure reported in Ref. 14 was followed, by calculating for each decay time the corresponding standard derivations (τ).

The following results are referred to the 1 kHz octave band (Table I), but similar results can be obtained at different bands. The highest estimation accuracy was obtained for combinations F-17 and F-18 (i.e., with both the source and receiver in the choir), providing a T_1 of 3.2 s and a T_2 of 8.9 s, in good agreement with the mean T_{30} value [Fig. 6(a)]. Standard derivation is small for the first decay (about 0.05 s) and somewhat longer for the second (about 0.13 s) but in both cases correspond to a 1.5% of the calculated value. Combination F-34 also showed a double slope, but its boundary position determined a T_1 of 4.1 s (with a standard derivation of 0.158 s) and a T_2 of 9.2 s. Results are less obvious out of the choir [Fig. 6(b)]. On average when the source and receivers were in the crossing T_1 was 1.3 s, while T_2 was 8.7 s. In all the cases the standard derivation is very small for both the decays. The first decay appears substan-

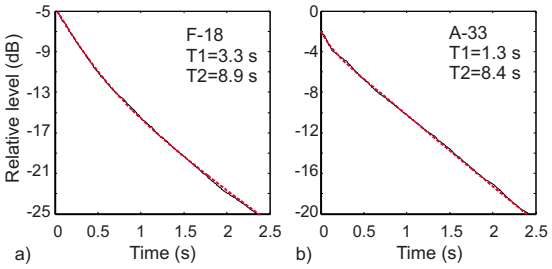


FIG. 6. (Color online) Comparison between measured decay (—) and multi-rate decay derived from Bayesian estimation (---) for combinations F-18 (a) and A-33 (b). Direct sound was excluded from each analysis.

tially shorter than in the choir, as a consequence of the wide openings. However, its magnitude is much smaller because of the large amount of acoustic energy flowing directly into adjacent sub-volumes. Consequently the coupled-system reverberation appears soon after the direct sound, resulting in EDT values generally longer than in the choir except at points very close to the source where the direct sound dominates. It is interesting to point out that even though Bayesian estimation shows that double decay rates take place, they barely affect T_{30}, confirming, as observed, that apart from a few cases the first decay expires soon.

D. Directional energy plots

The directional components of B-format measurements (X,Y,Z) were combined with the omni-directional component (W) to provide a 3D IR. Polar plots representing the energy content arriving from discrete directions represented by azimuthal and zenithal angles projected on the same plane were used to make the information more easily accessible. Then, also taking into account the time distribution, three time slices were considered: from 0 to 80 ms, from 80 to 200 ms, and from 200 to 1000 ms (Fig. 7). For each plot the level of the reflections was normalized assuming as a reference the maximum energy value for that slice.

A quantitative measure, which can be conveniently used in this analysis, is the directional diffusion δ as defined by Gover *et al.*,[23] so that a value equal to 0% corresponds to anechoic conditions, while 100% corresponds to perfectly diffuse sound field. In the present case δ was calculated with reference to the horizontal/azimuthal distribution (δ_h) and to the vertical/zenithal distribution (δ_v).

The analysis of combination F-17 [Fig. 7(a)] clearly shows that the first 80 ms are dominated by the direct sound and by few reflections in the horizontal plane, while in the vertical plane reflections appear weaker but quite diffuse. From 80 to 200 ms the dominance of the sub-volume is even clearer, with reflections mostly coming from the choir and very weak reflections coming from the crossing. From 200 to 1000 ms the reflections are more evenly distributed, but the dominant role of the choir is still evident as the reflections coming from the crossing are 5 dB weaker. This confirms that, as observed in Fig. 5(b), the coupled-system decay appears about 1 s after the direct sound.

It is interesting to observe, in comparison, the behavior of combination A-17 [Fig. 7(b)], which shows a clear dominance of the frontal reflections up to 200 ms, indicating that acoustic energy is mostly coming from the aperture, which connects the crossing, first in the form of direct sound and then as mostly diffuse reflections. In fact, from 200 to 1000 ms, diffuse reflections arrive from the choir side, providing the highest directional diffusion values.

Combination A-21 [Fig. 7(c)] is taken into account as an example of source and receiver located in the crossing, showing that within the first 80 ms the sound field is dominated by the direct sound and by reflections mostly located in the horizontal plane (in fact, the vertical distribution is quite oblong with important contributions coming from the floor). The directional diffusion is quite low, confirming that the large apertures prevent receivers located in this subspace from receiving strong early reflections. From 80 to 200 ms the reflections still come from the horizontal plane (mostly because of the high dome), and especially from the sides (identifiable as reflections from the pillars). From 200 to 1000 ms the sound field is more diffuse, but with a horizontal distribution, which suggests contributions from the choir and the transept, while reflections from the long nave are still lacking.

As a further comparison, it is useful to analyze combination A-05 [Fig. 7(d)], which shows a mostly frontal dominance up to 200 ms due to the combination of direct sound and diffuse energy arriving through the aperture. From 80 to 200 ms there are interesting contributions from the top, suggesting reflections from the barrel vaults. From 200 to 1000 ms reflections become more diffuse, but a clear axial dominance appears both in the horizontal and in the vertical plane suggesting a slow build-up of the purely reverberant field. Lack of strong lateral reflections may be explained as the result of the elongated geometry of the nave and of the wide and deep openings, which capture large amount of sound energy.

In conclusion the analysis of the directional energy plots allowed studying energy exchanges from a different perspective, giving information about which part of the room contributes to the decay at a given time.

IV. COMPARISON WITH THEORETICAL MODELS

A. Preliminary considerations

The above results only demonstrate that each subspace has its specific acoustic features, which influence the very beginning of the IR. In order to improve the knowledge of the coupling mechanism two parallel approaches were followed. First, a GA model was made using the CATT-ACOUSTIC V. 8.0h software, which implements a special algorithm to deal with coupled spaces.[17] Then, the generalized model of coupled subspaces proposed by Summers *et al.*[16] was applied.

However, both models share the definition of the space geometry and of surface characteristics. Room geometry was greatly simplified to take into account that geometric details that are small compared to wavelength are acoustically invisible or make the surface partly scattering (see Ref. 24, p. 176). Niches, sculptures, and decorations were replaced with simple flat surfaces having modified absorption and (only for GA) scattering properties. The need to modify absorption

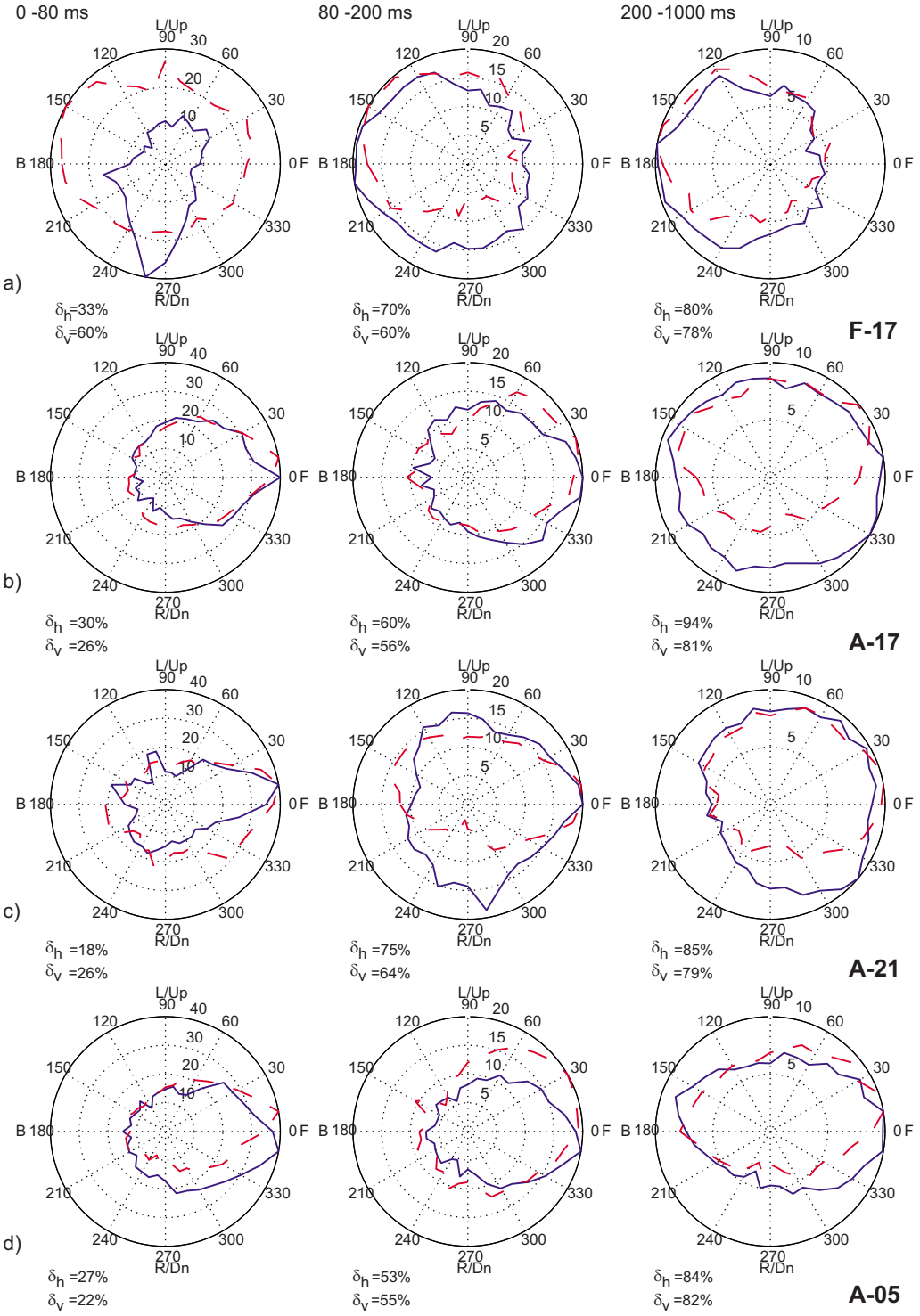


FIG. 7. (Color online) Polar plot of directional distribution of the energy content of the reflections at 1 kHz for different combinations of sources and receivers (F-17, A-17, A-21, and A-05). Reflection levels are normalized with reference to the maximum energy content for each time slice. (---) Energy level in the vertical plane and (—) energy level in the horizontal plane. Each receiver always points toward source A. δ_h=horizontal diffusion coefficient; δ_v=vertical diffusion coefficient; F=front, B=back, L/up=left/up, and R/Dn=right/down.

coefficients resulted from the increase in actual exposed area of decorated surfaces compared to the corresponding flat surface of the same material. Conversely, literature values[25,26] were assigned to pews, floors, glass, and wooden parts, which are normally assumed as flat (Table II). In order to better understand how conventional absorption coefficients had to be increased as a function of surface decorations further considerations were required.

Assuming Sabine's formula to be valid (neglecting coupling effects) the mean absorption coefficients (α) as a function of frequency, subtracting air absorption estimated using ISO 9613-1,[27] were calculated (Table III). However, taking into account that the flat floor surface is 12 600 m^2, the average absorption coefficient of the remaining surfaces is slightly higher, with a mid-frequency value of 0.10. Similar values are quite frequent in churches, especially when they

TABLE II. Summary of the absorption coefficients used in the GA and SA models.

	Fractional area (%)	Absorption coefficient per octave band (Hz)					
		125	250	500	1000	2000	4000
Marble floor (after Ref. 24)	17	0.01	0.01	0.2	0.02	0.03	0.03
Pews (after Ref. 24)	1	0.10	0.15	0.18	0.20	0.20	0.20
Glass (after Ref. 25)	3	0.35	0.25	0.18	0.12	0.07	0.04
Coffered vaults[a]	17	0.20	0.20	0.20	0.25	0.25	0.25
Dome and arches[a]	19	0.04	0.04	0.05	0.05	0.06	0.06
Sculptures[a]	3	0.12	0.12	0.15	0.15	0.18	0.18
Richly decorated marble/stuccos[b]	40	0.04	0.05	0.06	0.07	0.08	0.08

[a]Determined by comparison with similar surface treatments found in other churches.
[b]Determined from iterative calibration of GA model.

are richly decorated and characterized by a complex geometry, which determines large surface-to-volume ratio. The availability of a large set of acoustical and geometrical data[6] showed that over a sample of 56 different churches the minimum mid-frequency α (observed in seven churches) was 0.04, while the maximum was 0.14. The less absorbing group of churches was characterized by scarcely decorated plaster walls, hard stone floors (covered by few pews), and flat reflecting ceilings. The most absorbing group of churches was characterized by either richly decorated surfaces, coffered ceilings or large pew areas, or combinations of them. Taking into account the architectural features observed in the basilica, a mid-frequency α equal to 0.10 appears a perfectly suitable value.

However, assuming a uniform absorption on all the surfaces except the floor is not realistic as the decorative patterns are not evenly distributed. In fact, the exposed surface may be considerably different and, given the dimension of the church (and proportionally of the decorations, sometimes approaching half a meter in depth), the increased area is likely to be effective even at medium and low frequencies.

Comparisons between similar churches differing by just one element (provided it covered a reasonably large area) allowed estimation of the absorption coefficients of the given element. Thus, for a wooden coffered ceiling (like that found in the Cathedral of Taranto, or in the Basilica of Santa Maria Maggiore in Rome), α varies from 0.40 up to 0.60 at mid-frequencies, with small variations as a function of frequency. For a coffered dome finished in plaster (like that of the church of the Gran Madre di Dio in Turin) α varies between 0.20 at 125 Hz and 0.23 at 4 kHz. Finally, decorated surfaces, even though they are made of marble or plaster (like those found in the church of San Lorenzo in Turin, or in the church of San Luca e Martina in Rome), show a mid-frequency α varying between 0.08 and 0.10 according to the richness of the decorative pattern (and consequently of the exposed surface), while the variations as a function of frequency are relatively small. Taking into account these results it seemed that absorption coefficients should be increased over all the frequency bands including low frequencies. It is interesting to observe that this "wide-band" behavior and the increased absorption values are well compatible with the absorption coefficients of baroque woodcarvings measured in a reverberant chamber.[28]

In order to understand this phenomenon better and improve the accuracy of the estimation a preliminary investigation was carried out using scale model testing in a 1:20 reverberant chamber. A 1/8 in. microphone (GRAS 40DP) with a flat frequency response (± 1 dB) up to 30 kHz was used as the receiver. A spark generator with adjustable energy discharge and air gap was used as the sound source. The measurements presented here were made in air with numerical corrections for absorption according to ISO 9613-1.[27] Given the above set-up the full scale frequency range was 125–1000 Hz.

TABLE III. Calculation of mean absorption coefficient α as a function of octave bands, using Sabine's formula and assuming the whole basilica as a single volume. α_{res} is assumed as the average absorption coefficient obtained by subtracting the contribution of the marble floor (having a surface of 12 600 m^2 and the absorption coefficients reported in Table II).

	Frequency (Hz)					
	125	250	500	1000	2000	4000
T_{30} (s)	13.6	12.5	10.9	8.9	6.9	4.5
A_{tot} (m^2)	5604	6125	7029	8558	11 142	16 982
A_{air} (m^2)	106	315	656	1122	2 378	7 136
$A_{tot}-A_{air}$ (m^2)	5498	5810	6373	7436	8 764	9 846
α	0.07	0.08	0.08	0.10	0.11	0.13
α_{res}	0.08	0.09	0.10	0.11	0.13	0.15

TABLE IV. Summary of the results of the measurements of absorption coefficients carried out in 1:20 scale model reverberant chamber. Results report only relative variations compared to a flat sample of the same material used to reproduce the decorative patterns.

		Relative increment per octave band (Hz)			
Sample type	Surface ratio	125	250	500	1000
Fluted pillar	2:1	1.5	1.5	1.3	1.3
Simple coffered pattern	2:1	1.3	2.1	2.0	1.7
Decoration	n.a.	1.5	1.6	1.9	2.0

Three gypsum models were made, two (approximately) reproducing decorative patterns typically found in churches and the third reproducing a simplified coffered pattern. To account for the different acoustic behavior of materials at high frequencies all the absorption measurements were expressed in relative terms by comparison with a flat sample of the same material. Curved diffusers made of polypropylene sheets were used in the chamber model to ensure a reasonably diffuse sound field in all the conditions, and consequently ensure that the observed variations in reverberation time could be attributed to differences in sound absorption. The measurements show (Table IV) that the absorption coefficients were increased by a minimum of 30% up to 50% at 125 Hz, and from a minimum of 30% up to 100% at 1 kHz, doubling the absorption of the flat surface. Given the relative simplification adopted to model the samples, those results were used as floor values, allowing greater increases when dealing with more richly decorated surfaces.

The previous results suggested assigning to coffered vaults, moderately decorated surfaces (domes and arches), and sculptures the absorption coefficients given in Table II. However, the absorption coefficients of the largest surface (about 30 000 m^2, 40% of the total surface), covered by richly decorated marble and stuccos, were finally determined by iteration, during the calibration of the GA model (see below), in order to have the best match between measured and predicted values. The resulting absorption coefficients were those given in the last line of Table II, showing an increase of 50%–75% with respect to the reference value, in good agreement with both the findings of the scale model measurements and the values derived from measurements in other churches.

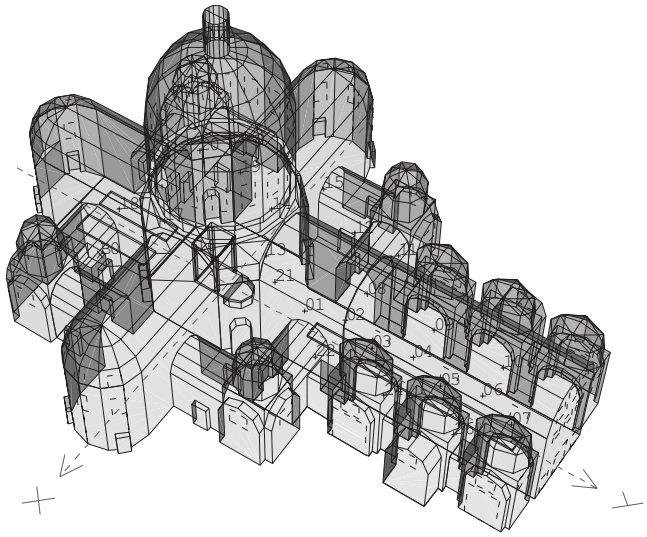


FIG. 8. GA model used to simulate the acoustics of the basilica.

B. GA model

Taking into account some geometrical measurements carried out in the field, combined with architectural drawings available in literature, a simplified 3D model was realized using about 1500 planes (Fig. 8). Source and receiver placement were arranged in order to correspond to actual positions used during the acoustic survey. As the late-part ray-tracing algorithm[17] can be considered as a Monte Carlo approximation to the exact solution and, consequently, it suffers from run-to-run fluctuations, which can be reduced by using a large number of rays, 1×10^6 rays were used in this case, while the truncation time was assumed to be equal to 14 s, slightly above the longest measured T_{30}. In addition, final results were given as averages over five different predictions for each source position together with the corresponding 95% confidence intervals (Table V). The latter were very small (corresponding to about 1.2% of the predicted value), suggesting that number of rays and truncation time were properly chosen.

Different surface treatments were accurately simulated taking advantage of photographs and field observations. Absorption coefficients were assigned as explained in Sec. IV A, while scattering coefficients were assigned taking into account the dimension of surface irregularities compared to

TABLE V. Summary of the octave-band values of reverberation time measured and predicted using different formulas, GA model, and SA model of coupled spaces. GA values are given with the corresponding confidence interval resulting from run-to-run fluctuations.

	Reverberation time (T_{30}) per octave band (Hz)					
	125	250	500	1000	2000	4000
Measured	13.6	12.5	10.9	8.9	6.9	4.5
Sabine's formula (single room)	11.4	10.9	9.8	8.0	6.5	4.1
Eyring's formula (single room)	11.8	11.4	10.1	8.2	6.7	4.2
GA model	13.4 ± 0.21	12.6 ± 0.15	11.0 ± 0.13	8.9 ± 0.11	7.1 ± 0.08	4.4 ± 0.04
SA model Sabine (coupled)	13.5	12.7	11.3	9.2	7.3	4.4
SA model Eyring (coupled)	13.0	12.2	10.9	8.8	6.9	4.2

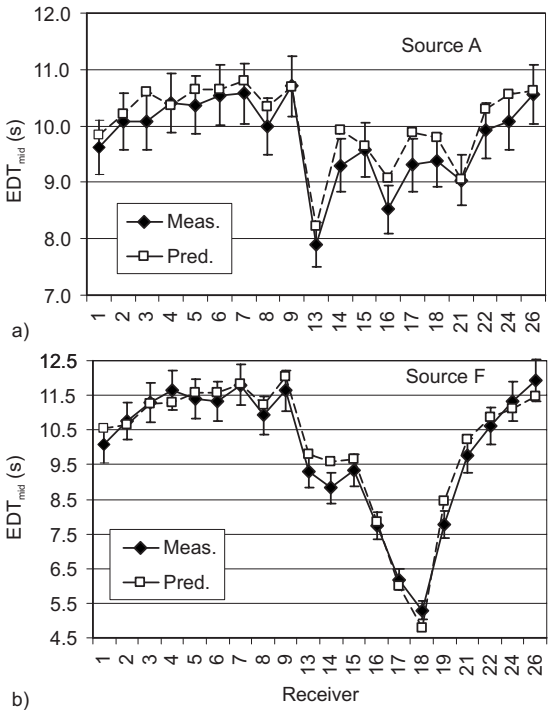


FIG. 9. Plot of EDT averaged over 500 and 1000 Hz octave bands measured (—) at different receivers compared with corresponding values predicted with GA model (---). Error bars correspond to JND for given parameter (equal to 5% of measured value).

wavelength. Flat or scarcely decorated surfaces were assigned scattering coefficients varying from 0.12 at 125 Hz to 0.17 at 4 kHz, including a linear increase of 0.01 to account for frequency dependence. Shallow decorated surfaces were assigned higher scattering coefficients varying linearly from 0.20 at 125 Hz to 0.40 at 4 kHz. Finally, pews, sculptures, and coffered vaults were assigned scattering coefficients varying from 0.30 at 125 Hz up to 0.80 at 4 kHz.

The prediction accuracy was estimated by calculating average rms errors over each source-receiver combination and expressing it in terms of just noticeable difference (JND), corresponding for T_{30} and EDT to 5% of the measured value.[29] With reference to source A the measured T_{30} were matched almost perfectly, with an octave-band error below one-half JND from 125 Hz to 4 kHz. The resulting mid-frequency values of EDT [Fig. 9(a)] were also predicted with good accuracy (average error of 0.9 JND), as well as other acoustic parameters not discussed here, such as strength (average rms error of 0.6 dB) and clarity (average rms error of 1.5 dB), suggesting good reliability of the model. The comparison between measured and predicted values with reference to source F [Fig. 9(b)] was carried out using the same settings, leading to the following average errors (at mid-frequency) of 1.0 JND for T_{30} and 0.8 JND for EDT. Other parameters were also predicted with reasonable accuracy with an average rms error of 0.4 dB for strength and 2.0 dB for clarity.

The GA model finally predicted T_{30} and EDT with good accuracy and showed that T_{30} values calculated with either Sabine's or Eyring's formula assuming the whole church as a single large room and using the same absorption coefficients were lower (Table V). As the basilica is the combination of different volumes, Sabine's formula was expected to fail. However, contradicting Shankland and Shankland's hypothesis,[5] the underprediction of T_{30} proved that the coupled-space model required absorption coefficients greater than those resulting from application of the above formulas to the whole volume.

Bayesian analysis applied to predicted IRs located in the choir showed results in agreement with those measured, with only slightly smaller values, indicating that at 1 kHz T_1 for this subspace is about 2.8 s (average $\tau_1=0.018$), while T_2 is 8.2 s (average $\tau_2=0.16$). When the source was in the crossing the steep initial decay was difficult to detect, probably because much of the radiated energy propagated to the adjacent subspaces, so that it was rapidly overtaken by the coupled-system decay.

C. SA model

The final comparison was carried out by applying a model of coupled subspaces, i.e., dividing the whole interior volume into subspaces and applying the classical diffuse-field equations to each of them, assuming they are connected by apertures through which acoustic energy may be exchanged.

In order to mathematically represent the decay of sound in acoustically coupled spaces, the nonstationary processes of sound energy decay are considered, following either steady state or impulse excitations. The sound energy decay in the whole interior, divided into m acoustical subspaces, is described by a system of m sound energy balance equations[16]

$$V_i(d\varepsilon_i/dt) = -cA_i\varepsilon_i/4 + \sum_j cS_{ij}(\varepsilon_j - \varepsilon_i)/4, \quad (1)$$

where $i=1,\ldots,m$, c is the sound speed, ε_i denotes the average sound energy density in the ith subspace, V_i is the volume of the ith subspace, and A_i is the equivalent absorption area of the ith subspace calculated according to Sabine's model as $S_i\bar{\alpha}_i+4mV_i$, where S_i and $\bar{\alpha}_i$ are, respectively, the total surface area and the geometrically averaged absorption coefficient of the ith subspace and $4mV_i$ is the propagation loss due to air. The coupling area between subspace i and adjacent subspace j is denoted $S_{i,j}$.

Following the approach of Summers *et al.*,[16] the above equation may be modified to account for using Eyring absorption exponent α_i' instead of $\bar{\alpha}_i$. Defining η_i as the ratio $\alpha_i'/\bar{\alpha}_i$, Eq. (1) may be modified by multiplying A_i by η_i.

The resulting system of linear differential equation (1) can be presented in matrix form and solved by finding the corresponding eigenvalues δ_i and eigenvectors ε_0^i (see Ref. 16 for mathematical details), and finally determining the constant terms from initial conditions.

One of the main issues to be addressed when solving coupled-volume problems by means of statistical analysis is the general validity of the basic assumption of statistical models. This means that inside each subspace the sound field must be reasonably diffuse. Furthermore, it is known that SA models are most accurate when applied to systems that are not strongly coupled. In strongly coupled systems the acous-

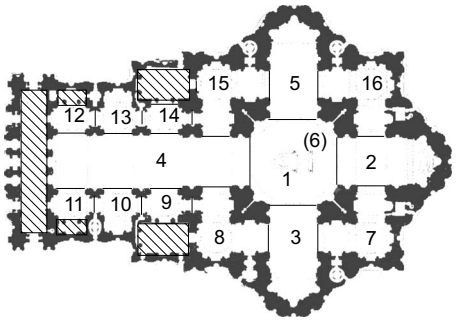


FIG. 10. Subdivision of the basilica into 16 subspaces. Subspace 6 refers to the dome. Dashed areas correspond to subspaces separated by doors and not included in the model.

tic energy is equally distributed, while in weakly coupled systems the subsystems behave as if they were isolated. Intermediate cases may be described by a theory that accounts for the exchange of energy between the spaces. A criterion to define how coupled two spaces are is therefore highly necessary. According to Kuttruff[2] SA models are applicable when energy lost via coupling is not substantially larger than the energy lost via absorption. Cremer and Muller[1] provided a coupling factor defined as

$$k_i = \frac{S_{ij}}{A_i + S_{ij}}, \tag{2}$$

giving strong coupling when $k_i \approx 1$ and weak coupling when $k_i \approx 0$. Further causes of discrepancies may appear as the subspaces become more strongly coupled. In fact, large apertures cause anisotropy in the sound field due to the establishment of a net flow of energy and alteration of the free-path distribution.[16]

Taking into account that surface decorations and the geometry ensure a high level of diffusion inside each subspace, the second most important issue to be evaluated is the coupling level, which should not be too high to prevent the above mentioned problems. In order to satisfy this condition the whole volume of the church was subdivided into 16 subspaces (Fig. 10) and, according to Eq. (1), for each subspace volume, total surface area, total absorption, and coupling area were calculated. Calculations were performed by using the same geometrical information and absorption coefficients derived from the GA model. The coupling factor, calculated according to Eq. (2), varied between 0.34 and 0.63, reasonably below the limit of 1, which corresponds to perfect coupling and potential lack of accuracy in the model solution.

According to the approach of Summers *et al.*,[16] when the sound source was located in room i, a fraction of its radiated power P was assumed to directly propagate into adjacent sub-rooms according to the fraction of the total solid angle subtended by the coupling surface S_{ij} as viewed from the source. Assuming a point source located close to the actual source positions the fraction was calculated by numerical integration for subspaces 1 (which radiates toward subspaces 2, 3, 4, 5, and 6) and 2 (which radiates toward subspaces 1, 4, 7, and 16).

In addition, Ref. 16 reports that one problem that is strictly related to complex spaces with multiple connections is a non-diffuse transfer of energy, which determines radiation of sound energy through the coupling aperture that is distinct from the energy density of the reverberant field of the room into which it radiates. In this way reverberant energy may be directly transferred between rooms that are not adjacent, depending on the radiation shape factor between the two apertures. In the present case, this correction is particularly useful when the sound source is located in the choir, as a significant part of its reverberant energy is likely to be transferred to the subspaces connected to the crossing without being reflected. This fraction was given by radiation shape factors that were calculated (according to Eq. 21 in Ref. 16) using numerical solving based on a contour double integral formula. As a consequence the energy fractions directly radiated from subspace 2 to subspaces 3, 4, 5, and 6 are, respectively, 0.116, 0.121, 0.116, and 0.101. The fraction directed toward subspace 4 was reduced to 0.08 to take into account the masking effect of Bernini's baldachin. The corresponding coupling surfaces were consequently rearranged according to Eq. (22) in Ref. 16.

Surfaces were assigned the same absorption coefficients used in the GA model and calculations were performed by using both Sabine's and Eyring's absorption exponents. Results derived from the application of the SA model were in good agreement with those derived from the GA model. In fact, as reported in Table V, it can be observed that values calculated with Sabine's formula are best matched with low frequency values (where mean absorption is lower), while those calculated with Eyring's formula are best matched with medium and high frequencies. In all the cases the reverberation times resulting from the coupled-volume model are longer than the corresponding values calculated assuming the whole space as a single room.

Even though the time constants of the 16 exponentials contributing to each decay curve are well known, in order to compare the results predicted by the SA model with those derived from Bayesian estimation, the latter algorithm was applied to determine the initial decay times corresponding to each subspace, when the sound source is inside the same space. Calculations were carried out at the frequency of 1 kHz and taking into account decays obtained using Eyring's correction. The results show that both in the crossing (where $T_1 = 1.8$ s) and in the choir (where $T_1 = 2.8$ s), predictions are in quite good agreement with results deriving from measurements.

When the source was in the crossing [Fig. 11(a)] much of the radiated energy propagated to the adjacent subspaces, consequently the steep initial decay is not as evident as in other subspaces, being rapidly overtaken by the coupled-system reverberation. When the receiver was located in the choir or in the nave the initial decay is slower (as a consequence of the energy exchange between adjacent subspaces) and rapidly fades into the coupled-system reverberation. When the source was in the choir [Fig. 11(b)] the initial decay appears clearly when the receiver was in the same subspace, while in the crossing the shorter initial decay is barely perceptible. When the receiver was in the nave, the initial energy is negligible, so the initial slope is near zero.

Even though the decay curves show a non-linear behav-

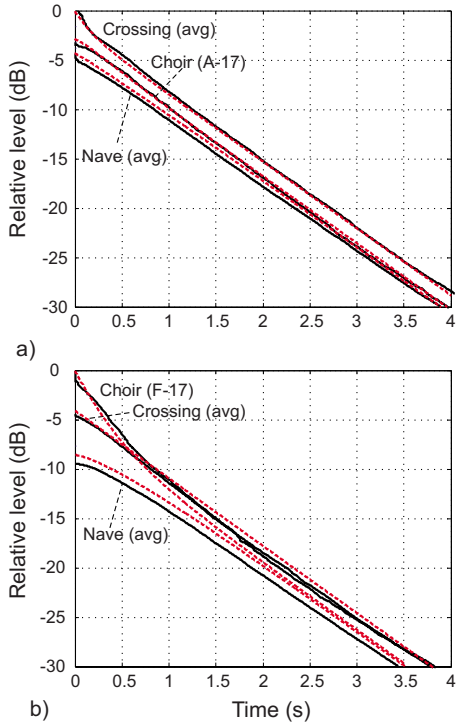


FIG. 11. (Color online) Comparison between early decay curves predicted using the SA model (- - -) and those obtained by averaging measured values at points representative of the same source-receiver combination (——). (a) Sound source located in the crossing in front of the altar. (b) Sound source located in the choir.

ior (suggesting that T_{30} and EDT should be treated carefully), a final comparison between the subspace averages of the above mentioned parameters and those predicted by the GA and SA models at the same frequency shows (Fig. 12) good agreement (with differences generally below 5%), confirming the accuracy of both the prediction-models.

A comparison between values calculated with and without solid angle correction for direct radiation and non-diffuse energy transfer proved that, as suggested by Summers *et al.*,[16] their inclusion is useful in giving accurate estimation of the initial decay, leading to better EDT predictions.

Given the good accuracy provided by the simple SA model of coupled volumes it is particularly interesting to analyze sound behavior in subspaces where, due to time limitations, no sound sources were placed during the survey. A sound source located in the diagonal chapel (subspace 8) would determine, in the same subspace, a $T_1=2.3$ s with the coupled-volume reverberation ($T_2=8.8$ s) appearing more than 15 dB after the direct sound. This implies that EDT (equal to 2.9 s) is strongly influenced by T_1, while T_{30} becomes less significant as the decay curve within the measuring interval is markedly non-linear. Similarly, in one of the side chapels (subspace 10), $T_1=1.7$ s, and T_2 appears after 20 dB, so EDT=2.5 s, while T_{30} is not significant. The latter results suggest that the acoustics of the secondary volumes of the basilica, frequently used (even today) for daily celebrations, may provide better conditions for speech intelligibility.[30]

All the above considerations refer to the unoccupied conditions. A large congregation would add absorption, reducing the coupling factor [see Eq. (2)] and finally leading to a weaker coupling. This means that energy exchange between volumes would be reduced, emphasizing differences among them. As this configuration is actually the most important for the listener, a simple calculation was made using the SA model and assuming a seated audience (1.5 person/m^2, $\alpha=0.99$ at 1 kHz) distributed over 70% of the floor area of the four braces and the crossing. Results show that the coupled-system reverberation at 1 kHz would be lowered to 6.1 s. EDT in the nave would be 5.5 s with the source on the altar, and 6.6 s with the source in the choir. A sound source located in the crossing and then in the choir would determine in each subspace, respectively, EDTs of 4.4 and 3.1 s.

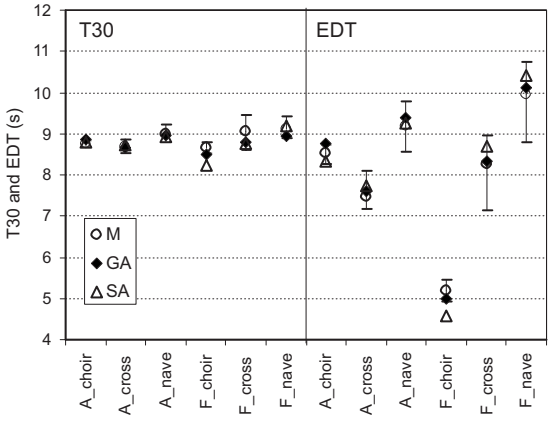


FIG. 12. Plot of average values of T_{30} and EDT measured at 1 kHz octave band in different subspaces and with different source positions compared to values predicted with both GA and SA models. Error bars represent variations between measured values.

V. CONCLUSIONS

The analysis of experimental data showed that reverberation times measured in St. Peter's Basilica are actually shorter than generally observed in churches of even smaller dimensions. In addition, significant differences appeared among EDT values, suggesting important acoustic differences between subspaces. Bayesian parameter estimation suggested that coupling effects actually take place and the analysis of the directional distribution of reflected energy allowed visualizing these effects.

The application of GA and SA models showed that St. Peter's Basilica actually behaves as a system of coupled volumes in which the acoustic conditions may significantly vary from subspace to subspace, according to source and receiver placements. Both models show that the resulting reverberation time is longer than that predicted using either Sabine's or Eyring's formula assuming the whole space as a single room volume. This implies that, given the substantial similarity of the surface treatment in the main nave and in the side chapels, in order to obtain the measured reverberation time the above mentioned surfaces should have increased absorption properties, mostly depending on the increased exposed surface due to the high degree of decoration, contradicting Shankland and Shankland's hypothesis[5] that the lower reverberation might result from coupling. Conversely,

coupling effects explain the dependence of reverberation on source and receiver position, leading to sound perception, which may be quite different from point to point, demonstrating how this building is capable of sounding not only like a big cathedral during solemn celebrations but also like a small parish church during daily services celebrated in the secondary volumes. Further investigations are needed in order to better clarify the mechanism of increased absorption shown by decorated surfaces.

ACKNOWLEDGMENTS

The author would like to thank His Eminence Cardinal Angelo Comastri, President of the Fabric of St. Peter's, Archpriest of St. Peter's Basilica in the Vatican for granting access to the church. He is also grateful to Professor Ettore Cirillo for encouraging and supporting this research, to Michele D'Alba for his kind cooperation in organizing and carrying out the survey, and to Rendell Torres and Jason Summers for their help in reviewing the paper.

[1]L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982), Vol. **1**.
[2]H. Kuttruff, *Room Acoustics*, 3rd ed. (E & FN Spon, London, 1991).
[3]V. O. Knudsen, "The effect of form on the reverberation of sound in rooms," J. Acoust. Soc. Am. **3**, 314 (1932).
[4]A. C. Raes and G. Sacerdote, "Measurement of the acoustical properties of two Roman basilicas," J. Acoust. Soc. Am. **25**, 954–961 (1953).
[5]R. S. Shankland and H. K. Shankland, "Acoustics of St. Peter's and Patriarchal Basilicas in Rome," J. Acoust. Soc. Am. **50**, 389–395 (1971).
[6]E. Cirillo and F. Martellotta, *Worship, Acoustics, and Architecture* (Multiscience, Brentwood, UK, 2006).
[7]A. Trochidis, "Reverberation time of Byzantine churches of Thessaloniki," Acustica **51**, 299–301 (1982).
[8]T. H. Lewers and J. S. Anderson, "Some acoustical properties of St. Paul's Cathedral, London," J. Sound Vib. **92**, 285–297 (1984).
[9]J. S. Anderson and M. Bratos-Anderson, "Acoustic coupling effects in St. Paul's Cathedral, London," J. Sound Vib. **236**, 209–225 (2000).
[10]C. F. Eyring, "Reverberation time measurements in coupled rooms," J. Acoust. Soc. Am. **3**, 181–206 (1931).
[11]B. Harrison and G. Madaras, "Computer modeling and prediction in the design of coupled volumes for a 1000-seat concert hall at Goshen College, Indiana," J. Acoust. Soc. Am. **109**, 2388(A) (2001).
[12]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1424 (2001).
[13]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," J. Acoust. Soc. Am. **113**, 2685–2697 (2003).
[14]N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayesian decay time estimation," J. Acoust. Soc. Am. **117**, 3707–3715 (2005).
[15]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**, 3744–3749 (2006).
[16]J. E. Summers, R. R. Torres, and Y. Shimizu, "Statistical-acoustics models of energy decay in systems of coupled rooms and their relation to geometrical acoustics," J. Acoust. Soc. Am. **116**, 958–969 (2004).
[17]J. E. Summers, R. R. Torres, Y. Shimizu, and B. I. Dalenback, "Adapting a randomized beam-axis-tracing algorithm to modelling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**, 1491–1502 (2005).
[18]S. Müller and P. Massarani, "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**, 443–471 (2001).
[19]ISO-3382, "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters," ISO, Geneva, Switzerland, 1997.
[20]F. Martellotta, E. Cirillo, A. Carbonari, and P. Ricciardi, "Guidelines for acoustical measurements in churches," Appl. Acoust. **70**, 378–388 (2009).
[21]E. Cirillo and F. Martellotta, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**, 232–248 (2005).
[22]F. Martellotta, "A multi-rate decay model to predict energy-based acoustic parameters in churches," J. Acoust. Soc. Am. **125**, 1281–1284 (2009).
[23]B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of directional properties of reverberant sound fields in rooms using a spherical microphone array," J. Acoust. Soc. Am. **116**, 2138–2148 (2004).
[24]M. Vorländer, *Auralization* (Springer-Verlag, Berlin, 2008).
[25]J. Meyer, *Kirchenakustik* (Verlag Erwin Bochinsky, Frankfurt am Main, Germany, 2003).
[26]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers* (Spon, New York, 2004).
[27]ISO 9613-1, "Acoustics—Attenuation of sound during propagation outdoors—Part 1: Calculation of the absorption of sound by the atmosphere," ISO, Geneva, Switzerland, 1993.
[28]A. P. O. Carvalho, M. Lencastre, and V. Desarnaulds, "Sound absorption of 18th-century baroque woodcarving in churches," Proceedings of the Inter-Noise 2002, Dearborn, MI, August (2002).
[29]I. Bork, "A comparison of room simulation software—The 2nd round robin on room acoustical computer simulation," Acust. Acta Acust. **86**, 943–956 (2000).
[30]B. Yegnanarayana and B. S. Ramakrishna, "Intelligibility of speech under nonexponential decay conditions," J. Acoust. Soc. Am. **58**, 853–857 (1975).

Investigation of acoustically coupled enclosures using a diffusion-equation model[a)]

Ning Xiang,[b)] Yun Jing, and Alexander C. Bockman
Graduate Program in Architectural Acoustics, School of Architecture, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 5 June 2008; revised 11 June 2009; accepted 12 June 2009)

Recent application of coupled-room systems in performing arts spaces has prompted active research on sound fields in these complex geometries. This paper applies a diffusion-equation model to the study of acoustics in coupled-rooms. Acoustical measurements are conducted on a scale-model of two coupled-rooms. Using the diffusion model and the experimental results the current work conducts in-depth investigations on sound pressure level distributions, providing further evidence supporting the valid application of the diffusion-equation model. Analysis of the results within the Bayesian framework allows for quantification of the double-slope characteristics of sound-energy decays obtained from the diffusion-equation numerical modeling and the experimental measurements. In particular, Bayesian decay analysis confirms sound-energy flux modeling predictions that time-dependent sound-energy flows in coupled-room systems experience feedback in the form of energy flow-direction change across the aperture connecting the two rooms in cases where the dependent room is more reverberant than the source room.
 [DOI: 10.1121/1.3168507]

PACS number(s): 43.55.Br, 43.55.Ka [LMW] Pages: 1187–1198

I. INTRODUCTION

A number of recently built performing arts venues contain one or more reverberant auxiliary rooms that are connected to the primary room (audience chamber) in such a way that there is an exchange of acoustic energy through an opening connection (coupling aperture). Such configurations constitute systems of coupled-rooms. In addition to enabling variable acoustics, one motivation behind use of coupled-rooms is the creation of particular kinds of non-exponential sound-energy decays,[1–3] so-called double-slope decays.[4] When sound-energy decays consist of a double-slope process with a rapid initial portion followed by a slower late portion, they are believed to simultaneously realize the desirable, yet competing perceptual attributes of clarity and reverberance. The subject of this paper is to apply a modeling technique based on diffusion equations to study acoustics in coupled-rooms.

Geometrical-acoustics (Refs. 5 and 6) and wave-acoustics approaches[7,8] have been recently used for numerically or computationally modeling the behavior of sound fields in coupled-rooms. The wave-acoustics methods are considered the more rigorous of the approaches as they more fully represent the phenomenology of sound fields. However, they only yield analytic solutions for a few simple cases and are problematic in high-frequency regimes due to the high computational load. By contrast, geometrical-acoustics, in which sound fields are modeled by ensembles of non-interacting classical phonons, is an inherently high-frequency approach. Though it does not necessarily allow for a greater number of enclosure geometries to be solved numerically, it is well suited to various computational implementations. Further, geometrical-acoustics methods provide a rigorous base from which to develop statistical models of sound fields. In recent work,[6] the cone-tracing algorithm with randomized tail corrections used by the commercial software CATT-ACOUSTIC was modified in order to model coupled-rooms.

Due to its computational efficiency in comparison with existing geometrical-acoustics and wave-theoretical approaches, transport/diffusion theory has been actively applied in urban noise propagation and room-acoustics predictions.[9–18] The calculation load of diffusion-equation modeling on currently available personal computers is on the order of seconds for steady-state simulations and 1 min or less for transient simulations for all numerical examples discussed in this paper. In a recent paper, Billon *et al.*[12] applied diffusion equations in acoustically coupled-rooms. In their work, Billon *et al.*[12] compared experimentally measured results in two coupled empty rooms with predicted results using the diffusion-equation model. The boundary conditions used for the diffusion equation in their work have certain limitations. As reported by Valeau *et al.*,[11] when the absorption assigned is as high as 0.2, discrepancies were found. Furthermore, the double-slope decays modeled by their diffusion model are not properly quantified.

This paper advances the efforts by Billon *et al.*[12] in several significant ways. First, this paper illustrates predicted sound-energy flows in two coupled-rooms. The correlation between the energy flow decay function and the steady-state energy decay function is discussed. Second, this work em-

[a)]This work is dedicated to Professor Jens Blauert on the occasion of his 70th birthday. Aspects of this work have been presented at WESPAC IX, 2006 in Seoul, Korea, at the International Symposium on Room-Acoustics August 2007 in Seville, Spain, and at the 155th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. 123, 3910(A) (2008)].

[b)]Author to whom correspondence should be addressed. Electronic mail: xiangn@rpi.edu

ploys an eighth scale-down system of two coupled-rooms to investigate the relevant aspects. Finally, the double-slope decays from the numerically modeled results and from the acoustical scale-modeling results are quantified within a Bayesian framework. The Bayesian probabilistic framework has been shown to estimate not only the decay parameters from a Schroeder decay model, but also to quantify uncertainties of decay time estimates and the interrelationship between multiple decay times.[19]

This paper is structured as follows: Section II briefly presents the governing equations of the diffusion-equation model including boundary conditions. Section III focuses on the sound pressure level (SPL) distributions and sound-energy decays in the coupled spaces. The experimental results obtained from the scale-model coupled-rooms are compared with the simulation results. Section IV elaborates on the energy flow in coupled spaces, including both the time-dependent energy flow directions and energy flow decays. Section V concludes the paper.

II. DIFFUSION-EQUATION MODELS

The diffusion-equation model is based on the assumption that the room(s) under study contains scattering objects that uniformly scatter the sound and have the same mean free path length of the room when walls are considered diffusely reflecting. Under this circumstance, walls are replaced with the scattering objects. Following the physical analogy with the diffusion of particles in a scattering medium, and assuming sound particles travel along straight lines at sound speed c and follow a certain statistical process (usually happens after early reflections) in the room(s) under investigation with diffusely reflecting walls,[10] the sound-energy density $w(\mathbf{r},t)$ as a function of location $\mathbf{r}$ and time t is then governed by the diffusion equation, which describes the energy flow from a high-density area to low-density area

$$\frac{\partial w(\mathbf{r},t)}{\partial t} - D\nabla^2 w(\mathbf{r},t) + cmw(\mathbf{r},t) = q(\mathbf{r},t), \quad \in V, \tag{1}$$

where $D=\lambda c/3$ is termed the diffusion coefficient with λ being the mean free path. The subroom denoted by domain V has a source term $q(\mathbf{r},t)$, which is zero for any subdomain where no source is present. The term $cmw(\mathbf{r},t)$ accounts for air dissipation in the room(s) with m being the absorption coefficient of air,[18] and can be extended to account for absorption due to scattering objects inside the room(s).[13] Equation (1) is the interior equation in domain V, being subject to the boundary condition on the interior surface S,

$$D\frac{\partial w(\mathbf{r},t)}{\partial n} + cAw(\mathbf{r},t) = 0, \tag{2}$$

where n is the out-going normal of the wall surface. The absorption term A can take the following forms:

$$A_S = \frac{\alpha}{4}, \tag{3a}$$

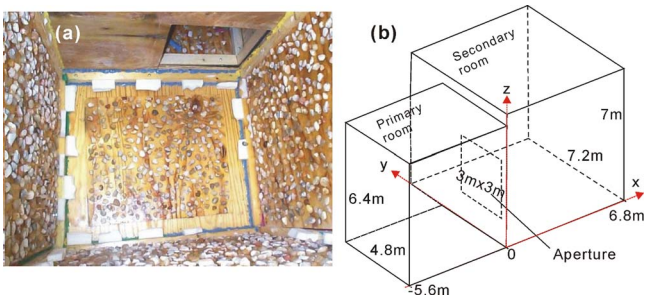


FIG. 1. (Color online) Eighth scale-model of two coupled-rooms. An opening window (coupling aperture) connects the larger (secondary) room to the main (primary) room. (a) Photograph of the modeled coupled-rooms. Wall surfaces are treated acoustically to create diffuse reflections. (b) Dimensions of the two rooms. The aperture size and position may be changed for different investigations. The dimensions of the two rooms are given in full-scale.

$$A_E = \frac{-\ln(1-\alpha)}{4}, \tag{3b}$$

or

$$A_M = \frac{\alpha}{2(2-\alpha)}. \tag{3c}$$

The term A_S has been used in room-acoustics predictions since Ollendorff in 1969.[9] More recently, Jing and Xiang[14] and Billon *et al.*[15] independently proposed the absorption term A_E in Eq. (3b) for modeling cases where some surfaces of the room under test have a high absorption. Diffusion equations using the absorption terms A_S or A_E in Eqs. (3a) and (3b) are designated as the diffusion-Sabine model or diffusion-Eyring model, respectively.[15] The Sabine-diffusion model[11,16] is only suitable for room surfaces with absorption coefficients not larger than 0.2. The Eyring-diffusion model fails to predict acoustic behavior in a room whenever a portion of room surfaces features a high absorption coefficient of 1.0; in this case, the Eyring-diffusion model suffers from a singularity problem. Most recently, Jing and Xiang[16] proposed the term A_M in Eq. (3c); they demonstrated that the diffusion equation with this modified absorption term in the boundary condition is theoretically grounded and can model high absorption for a small portion of surfaces. In addition, the diffusion-equation model inherently assumes that overall absorption in rooms under test must not be high.[11,16]

III. SOUND PRESSURE DISTRIBUTIONS AND ENERGY DECAYS

Preliminary experimental verification of the diffusion-equation model for coupled spaces has been made in Ref. 12. This section further compares the diffusion-equation model with experimental results, conducted in a scale-model of two coupled-rooms, for both SPLs and energy decay characteristics.

A. Experimental model

The coupled-rooms are implemented in a 1:8 scale-model made of 2 in. thick plywood as shown in Fig. 1(a). The interior surfaces have been covered with rocks, adhered

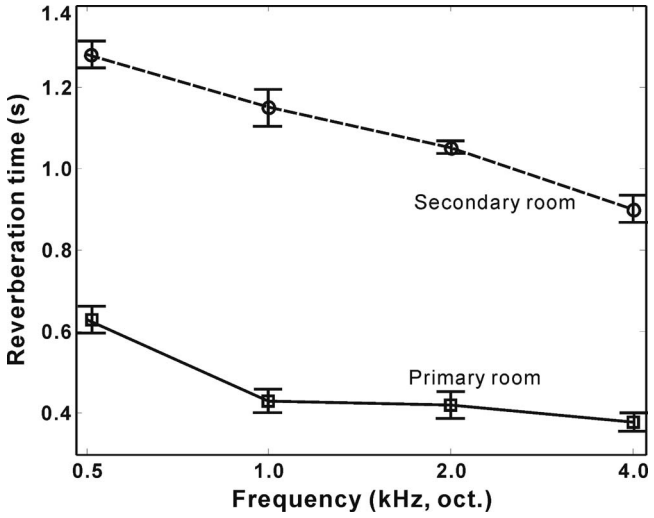


FIG. 2. Spatially averaged values of natural reverberation times measured at five different locations in the eighth scale-model rooms shown in Fig. 1(a), when the two rooms are acoustically separated (single-rooms). The bars indicate range of data.

to the wood surfaces with glue, to promote diffuse reflections. Figure 1(b) illustrates the dimensions of the rooms in Cartesian coordinates, given in full-scale. The room where a miniature dodecahedron loudspeaker system is placed to excite sound fields is referred to as the main (primary) room, while the other room, connected by a transparent area, is referred to as the secondary room. The transparent area connecting the two rooms is referred to as the (coupling) aperture, and its size is 3×3 m^2 throughout the experiments. In the scale-model experiments, the omni-directional sound source at location $(-5, 2.4, 1.3)$ m radiates a continuous maximal-length sequence signal of length 2^{19}-1 points in one period at a sampling frequency of 100 kHz, averaged over ten repetitions, while a 1/4 in. microphone is used as a receiver. Room impulse responses are measured throughout all experiments. The miniature dodecahedron loudspeaker system can cover a frequency range up to 32 kHz. With a scale factor 1:8, the measured room impulse responses contain useful information up to 4 kHz in full-scale. The glued rocks on interior surfaces can be considered diffusely reflecting above 1 kHz in full-scale.

When decoupled, reverberation times of each room, termed *natural reverberation times* and denoted as T_1' and T_2' to distinguish them from the decay times T_1 and T_2 of the double-slope decays in coupled-rooms, are determined first separately at five different locations as shown in Fig. 2 as spatially averaged values. The absorption coefficients of the wall materials are estimated from the averaged reverberation time by inverting the Eyring equation. No drying air or nitrogen[20] is used to eliminate the air dissipation in the volume, since the purpose of this measurement is to verify the diffusion-equation model, not to model any real space. The obtained absorption coefficients of the wall materials are expected to include the air absorption. Therefore, the air dissipation term in the diffusion-equation model is ignored. The overall averaged absorption coefficients are listed in Table I.

TABLE I. Absorption coefficient of the materials used in the scale-model at different octave bands (in full-scale).

	500 Hz	1 kHz	2 kHz	4 kHz	1.5–4 kHz
Primary room	0.21	0.28	0.30	0.32	0.30
Secondary room	0.14	0.16	0.17	0.19	0.17

B. SPL distributions

The recent work of Billon *et al.*[12] has reported reasonable agreement when comparing the SPL distribution along one straight line perpendicularly across the aperture. The SPL results reported in Ref. 12 show better agreement for higher-frequency bands (1 and 4 kHz), than for the lowest one (250 Hz). This work compares the SPL distributions in the two coupled-rooms and across the aperture between the scale-modeling experiments and the diffusion-equation modeling in aspects beyond the previous work. The diffusion model is implemented within two quasi-cubic rooms coupled through the coupling aperture as shown in Fig. 1(b). Equations (1), (2), and (3c) were solved by a finite-element method using a total of 8000 linear Lagrange-type mesh elements. The mean free paths of the primary and secondary rooms amount to 3.6 and 4.6 m, and the mean free times λ/c are 10.5 and 13.4 ms, respectively. The mean free time is the time determined by one mean free path, λ.

The steady-state SPL distribution is determined after solving for $w(\mathbf{r})$ from Eqs. (1), (2), and (3c). Figure 3 illustrates the sound source location, meshing elements, and the aperture along with the SPL distributions using a half-transparent pseudo-color scale.

For a better understanding of the SPL distributions near the aperture, Fig. 4 illustrates two sets of acoustical measurement results at 1.5–4 kHz (broad-band) in the scale-model coupled-rooms in comparison with the predicted SPL values by the diffusion model. The direct sound is excluded from the experimental data, leaving only the reverberant sound. The comparison along two straight lines $(0.32\ \text{m}\leq y \leq 4\ \text{m})$ parallel to the aperture shows agreement to a degree similar to that reported in the work of Billon *et al.*[12] To be precise, the vertical axis in Fig. 4 is scaled to detail variations within a 10-dB range. Most experimental results show deviations smaller than 1-dB, with maximal variation of 2.4 dB occurring at a position in the secondary room behind the wall. The reasonable agreement between the modeled and

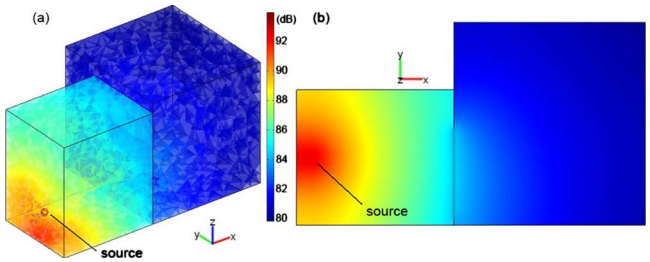


FIG. 3. (Color online) SPL distributions by the diffusion model in the scale-model coupled-rooms as shown in Fig. 1. (a) Three-dimensional presentation with a half-transparent pseudo-colored scale showing finite-element meshing. (b) Two-dimensional presentation in the X-Y plane.

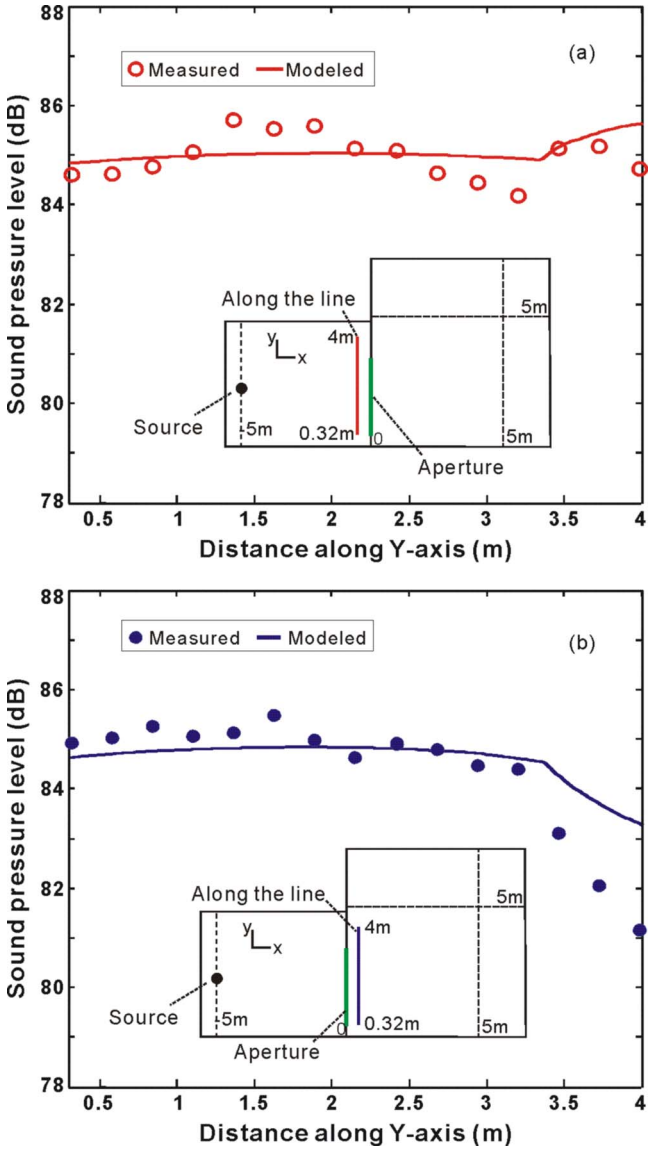


FIG. 4. (Color online) Comparison of steady-state sound pressure level distributions by the diffusion model with experimentally measured results at 1.5–4 kHz in the scale-model coupled-rooms as shown in Fig. 1. At 15 receiving positions SPLs are evaluated from measured room impulse responses along y-axis at $x=-0.05$ m and $x=0.05$ m, respectively. (a) Comparison in the primary room at $x=-0.05$ m along a line parallel to x-axis. (b) Comparison in the secondary room at $x=0.05$ m along a line parallel to y-axis. The maximum deviation of 2.4 dB is observed at a location behind the aperture opening.

measured SPLs indicates that the modified diffusion model implemented within this work can be used for detailed discussion on the SPL distributions and other aspects of sound fields.

One important aspect is the transition of the SPLs caused by the receiver location changing from a region close to the opening of the coupling aperture to the solid wall, and the transition of the SPLs when the receiver goes across the aperture from the primary room into the secondary room. In Fig. 3 the SPL distributions in pseudo-color scale across the aperture need further elaboration. Figure 5 illustrates SPL distributions by the modified diffusion model. Six different curves in each room are plotted on each side of the aperture as a function of y across the entire width of the rooms at x

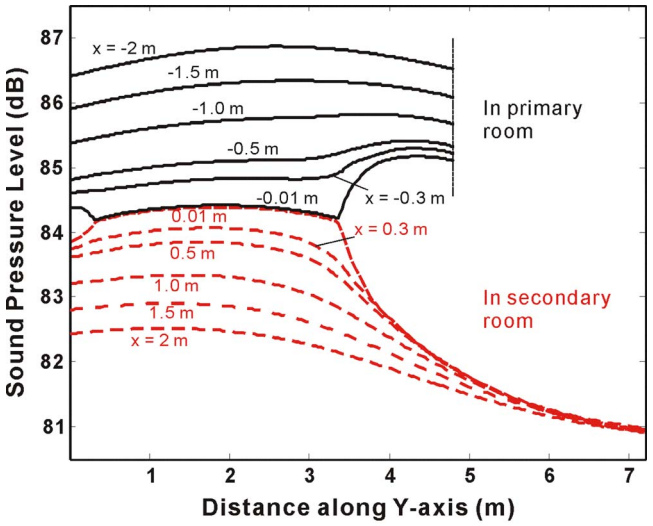


FIG. 5. (Color online) Steady-state sound pressure level distributions predicted by the diffusion model for the scale-model coupled-rooms as shown in Fig. 1. Six different curves are shown as a function of y across the entire width of the rooms at $x=\pm0.01$, ±0.3, ±0.5, ±1.0, ±1.5, ±2.0 m, and $z=3$ m [see Fig. 1(b) for definition of the coordinate].

$=\pm0.01$, ±0.3, ±0.5, ±1.0, ±1.5, ±2.0 m, and $z=3$ m [see Fig. 1(b) for definition of the coordinate]. As for SPLs at $x=\pm0.01$ m, being close to the wall featuring the aperture, the SPLs near the edges of aperture from the aperture opening to the solid wall undergo a sudden change while the SPLs from the primary room to the secondary room within a region close to the aperture opening exhibit continuities. Similar results are also found by solving the diffusion-equation model when the secondary room is equally or more absorbent than the primary room. In the primary room, the SPLs toward the solid wall beyond the edge of the aperture increase abruptly, while in the secondary room they decrease abruptly.

C. Sound-energy decays

For time-dependent solutions of the diffusion equation, the source term $q(\mathbf{r},t)$ on the right-hand side of Eq. (3) can be assigned as[11]

$$q(\mathbf{r}_s,t)=E_0\delta(t), \tag{4}$$

where $\delta(t)$ is the Dirac impulse, assuming an omnidirectional sound source with a sound-energy density E_0 at $t=0$ throughout a tiny, but finite volume V_s occupied by the source located at $\mathbf{r}=\mathbf{r}_s$. The solution of Eqs. (1), (2), and (3c) at any location $\mathbf{r}$ in subdomain V excluding V_s represents an energy (density) room impulse response $w(\mathbf{r},t)$, exclusive of the direct sound. Its energy-time function (ETF) [usually known as energy-time curves (ETCs)] can be expressed as[21]

$$L_p(\mathbf{r},t)=10\log_{10}\left(\frac{w(\mathbf{r},t)\rho c^2}{P_{\text{ref}}^2}\right), \tag{5}$$

where ρ is the air density, c is speed of sound, and P_{ref} equals 2×10^{-5} Pa. For sound-energy decay analysis, however, Schroeder-integration[22] of the energy room impulse response yields an approximation of the *steady-state sound-energy decay*

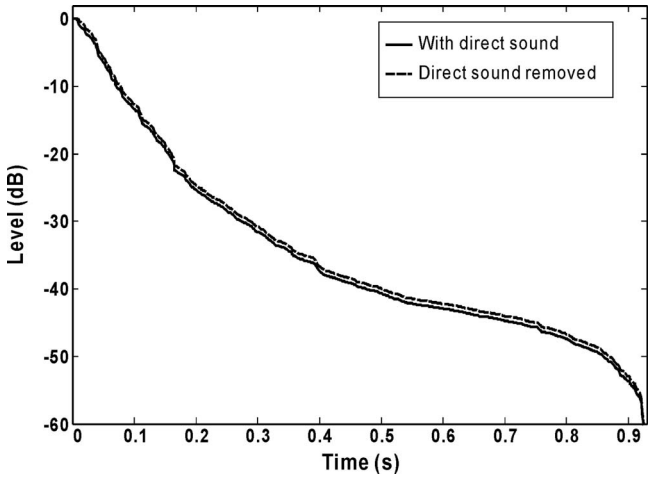


FIG. 6. Comparison of normalized Schroeder decay functions evaluated from a measured room impulse response in the scale-model rooms with that when the direct sound portion (first 5 ms) is removed.

$$d(\mathbf{r},t) \approx \frac{1}{N(\mathbf{r})}\int_t^\infty w(\mathbf{r},\tau)d\tau, \tag{6}$$

with

$$N(\mathbf{r}) = \int_0^\infty w(\mathbf{r},\tau)d\tau. \tag{7}$$

Such "steady-state derived sound-energy decay" describes the process resulting from the switching off of the source after the enclosure under investigation arrives at its steady-state.[22] Although the "approximation" in Eq. (6) implies that the energy room impulse response predicted by the diffusion model actually excludes the portion of direct sound,[11] the Schroeder-integration will still result, approximately, in the desired sound-energy decay. Figure 6 shows a comparison of Schroeder decay functions from a room impulse response experimentally measured in the scale-model rooms with that when the direct sound portion (first 5 ms) is removed. As indicated in Fig. 6, the general slope of the decay curves will remain, while the curve with removed direct sound is only slightly shifted.

Figure 7 illustrates two predicted ETF curves using Eq. (5) and their Schroeder decay functions in Eq. (6). The receiver positions (R_1,R_2) are also shown. Using the diffusion model, the Schroeder decay functions as plotted in Fig. 7 are generally smoothed curves; a slight shift of the decay curves will not cause significant estimation errors in the relevant decay-parameter estimation.

An ETF, essentially an energy room impulse response, is a time-derivative of the Schroeder decay function, while the steady-state derived energy decay is the Schroeder-integration of the energy room impulse response expressed in Eq. (6), termed Schroeder decay functions.[23,24] For the single-slope case, the decay process can be predominantly modeled by a single exponential decay term with its specific decay slope, the time-derivative or time-integration of which still remains a single exponential decay function. For double-slope decay process, though, essentially expressed by a linear superposition of two exponential terms with two distinctly different decay constants, the Schroeder-integration will yield decay terms with different decay parameters as defined in the following (Sec. III D).

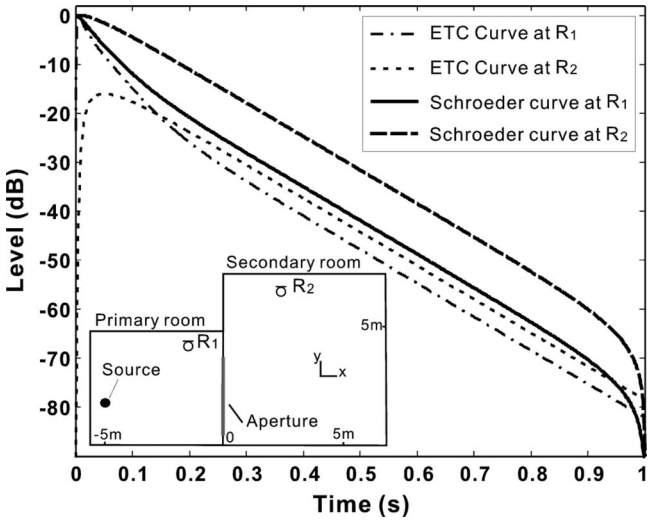


FIG. 7. Two energy-time curves and the corresponding, normalized Schroeder decay curves calculated using Eq. (6), predicted from the diffusion model. A sketch is included to show the receiver positions (R_1,R_2), respectively, in the modeled coupled-rooms.

Within proportionate spaces with single-slope decay, slopes of ETF curves (ETC) are similar to those of Schroeder decay functions, but when received at positions far from the sound source in elongated rooms,[11] flat rooms,[16] or at positions in the secondary coupled-room, where the direct sound cannot directly reach the sound receiver, the ETF curves will show a profound convex curvature at the initial part (see the ETF curve at R_2 in Fig. 7). A direct usage of the ETF curves for reverberation time evaluation, particularly for early decay time estimation, will lead to biased results. More importantly, for energy decay analysis in coupled-rooms, where double-slope decay characteristics are often expected, the difference will be profound as shown in Fig. 7. At receiver R_1 the sound energy exhibits double-slope decay characteristics. A comparison between the energy-time curve at R_1 and its corresponding Schroeder decay curve shows a clear difference at the initial part, resulting in different decay parameters.

Another option for obtaining the desired sound-energy decay from the diffusion model is to assign a switch-off signal to the source term $q(\mathbf{r}_s,t)$; however, this is much more time-consuming.

D. Quantifying double-slope characteristics

To quantify double-slope characteristics of Schroeder decay functions either experimentally measured or diffusion-model predicted, Xiang and Goggans[4] first proposed two decay times and a decay-level difference ΔL. Using these parameters as calculated by the software developed by Xiang and Goggans,[4,19] Bradley and Wang[25] carried out subjective tests using the data generated by a geometrical-acoustics based modeling software. More recently, Meissner[7] also used similar parameters to quantify his modeled energy decays based on wave-equation based method. Figure 8(a) illustrates

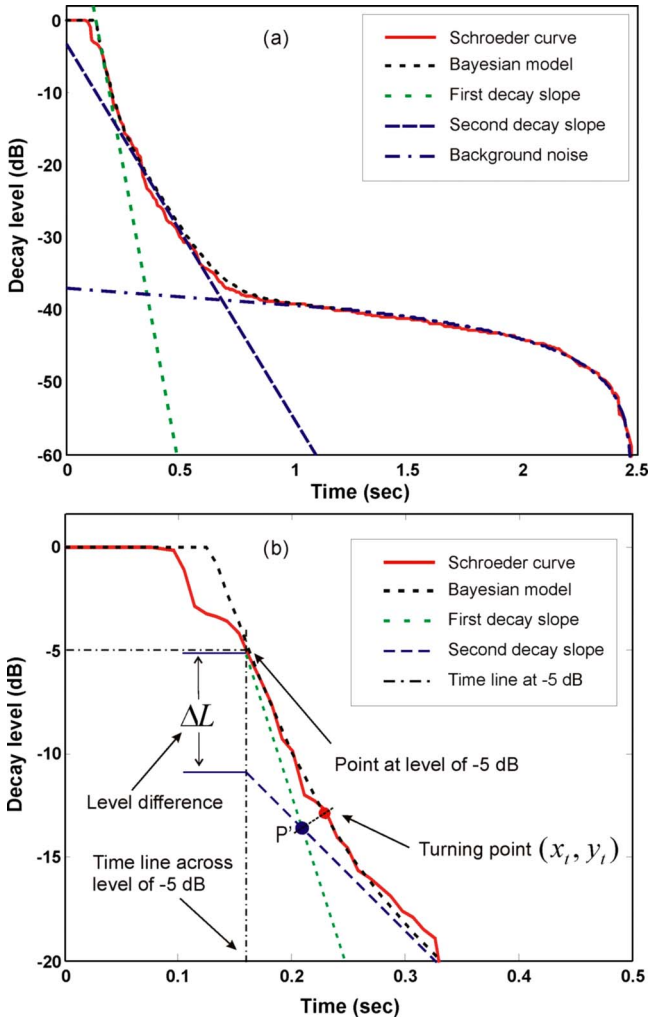


FIG. 8. (Color online) (a) Comparison between a normalized double-slope decay function measured in the scale-model system of two coupled-rooms with its normalized Schroeder decay model. (b) Definition of the decay-level difference ΔL(in decibel) and the turning point.

a double-slope decay function measured in the scale-model system of two coupled-rooms, in comparison with its Schroeder decay model.

Bayesian analysis[4,19,26] yields decay parameters associated with the three model terms in this case

$$F(\mathbf{A},\mathbf{T},t) = A_0(t_{\max} - t) + A_1 \exp\left(-\frac{13.8t}{T_1}\right) + A_2 \exp\left(-\frac{13.8t}{T_2}\right), \quad (8)$$

as plotted in Fig. 8(a) with $t_{\max}$ being the time limit corresponding to the upper limit of Schroeder integration. Figure 8(b) illustrates a magnified view of the first 20 dB. As recommended by ISO 3382,[27] the first 5 dB should be excluded for analysis purposes, particularly using the experimentally measured data; because the early portion of the energy decay represents only a small number of short-delay paths, it cannot be modeled by exponential decays. The data analysis from experimental measurements illustrated in Fig. 8 (as well as all other data analyses throughout this work) is undertaken from −5 dB to the end of the data record. Bayesian analysis is able to provide the model parameters in Eq. (8); namely, the amplitude parameters (A_1, A_2), decay times (T_1, T_2), associated uncertainties, and mutual dependence. A_0 is associated with the background noise in the room impulse response, actually a nuisance parameter, being a necessary part of the model, but irrelevant for the energy decay analysis. In addition to the individual decay times, architectural acousticians are interested in the relative amplitude A_2 with respect to A_1 for double-slope cases, rather than their individual values.[4] The decay-level difference ΔL_S in decibels, first used by Xiang and Goggans,[4] is defined as

$$\Delta L_S = 20 \log_{10}(A_1/A_2)|_{S[\mathrm{dB}]} \quad (\mathrm{dB}). \quad (9)$$

Figure 8(b) illustrates this definition and further indicates the reason that the reverberation analysis is usually undertaken using the data segment from $S=-5$ dB point along the Schroeder decay function, particularly from the experimentally measured data or from geometrical-acoustics modeling as recommended by ISO 3822. The decay-level difference ΔL_S quantifies how low the second decaying process characterized by T_2 is relative to the first one of T_1. The value S, marking the starting point along the decay function, is subject to choice. Other than $S=-5$ dB, purposely taken to exclude a small, but bumpy portion owing to discrete nature predominately associated with early reflections, S can take on a value to exclude an even smaller portion at the beginning of the decay function, such as $S=-0.5$ dB for those modeled by the diffusion-equation model as discussed in Sec. III C. These decay functions exhibit a smooth curve after one "mean free time."

In Fig. 8(b) two straight lines corresponding to two decay slopes in logarithmic presentation can be determined as follows:

$$y_j = a_j + b_j t_k, \quad (10)$$

with $a_j = 10 \log_{10}(A_j)$, $b_j = -10(13.8/T_j) \log_{10} e$, and $j=1,2$.. Bayesian decay-parameter estimation in the case of a double-slope decay yields two straight lines corresponding to the two decay slopes, which, in general, will not cross at a point coincident with the turning point of the data (Schroeder curve) and the model decay curve. Rather, the crossing point $P'(x_0, y_0)$ given by

$$x_0 = (a_2 - a_1)/(b_1 - b_2), \quad y_0 = (a_2 b_1 - a_1 b_2)/(b_1 - b_2) \quad (11)$$

will generally be lower in level. The turning point $P_t(x_t, y_t)$ is defined to be a point on the decay model curve, to which the crossing point (P') has the minimum distance

$$\sqrt{(x_t - x_0)^2 - (y_t - y_0)^2} \to \min. \quad (12)$$

Two decay times or decay ratio (T_2/T_1), along with the level difference (ΔL in decibel) are relevant decay parameters, sufficient for quantifying double-slope characteristics of sound-energy decays. In addition, the estimated coordinate of the turning point (x_t, y_t), particularly, the time instant associated with x_t, will approximately show the turning from the first decay process specified by its decay time T_1 to the second one.

Successful application of Bayesian analysis to the decomposition of the validated Schroeder decay model [Eq.

TABLE II. Decay parameters of the Bayesian decay analysis from both measured and diffusion-equation simulated results in the primary room with the receiver at (−2.7, 3, 3) m. Standard deviations Std_1 and Std_2 associated with decay times T_1 and T_2 are obtained from Bayesian uncertainty estimations (Ref. 19).

Band (kHz)	Data	T_1 (s)	Std_1 (s)	T_2 (s)	Std_2 (s)	Level difference ΔL (dB)	Turning point (ms)
1.0	Measured	0.32	3.65×10^{-3}	0.93	2.17×10^{-2}	5.31	86.4
	Simulated	0.32	2.5×10^{-4}	0.93	3.31×10^{-3}	5.36	50.0
2.0	Measured	0.31	1.79×10^{-3}	0.81	2.83×10^{-2}	5.22	86.4
	Simulated	0.32	2.4×10^{-4}	0.86	3.55×10^{-3}	5.39	60.0
4.0	Measured	0.29	2.13×10^{-3}	0.69	3.87×10^{-2}	6.31	91.2
	Simulated	0.30	2.4×10^{-4}	0.78	4.54×10^{-3}	6.29	50.0
1.5–4.0 (broad-band)	Measured	0.30	1.27×10^{-3}	0.81	2.31×10^{-2}	5.64	85.4
	Simulated	0.31	2.4×10^{-4}	0.86	2.75×10^{-3}	5.75	65.0

(8)] of sound-energy decay, as discussed in this paper, demonstrates that other methods for evaluating non-exponential decays, including solely visual inspections, and those which compare linear fits of different portions of logarithmic decay functions (e.g., T_{15} vs T_{20}, or T_{10} vs T of an arbitrarily chosen portion, e.g. between −30 and −40 dB) are scientifically questionable. Those quantifiers cannot generally provide a unique description for a non-exponential decay consisting of a linear combination of exponential decay functions. Thus, despite their ease of implementation, such techniques should no longer be practiced.

Table II lists relevant decay parameters, including two individual decay times $T_{1,2}$, their standard derivations $\mathrm{Std}_{1,2}$, the decay-level difference ΔL, and the time instant of the turning point between 1 and 4 kHz (octave bands) as well as 1.5–4 kHz broad-band, obtained from a measured room impulse response at a specific location (−2.7, 3, 3) m, which is close to the aperture opening in the primary room of the scale-down model as shown in Fig. 1. Bayesian decay analysis[19,26] has been applied to the Schroeder decay functions derived from the single octave-bandpass filtered and the broad-band room impulse responses, resulting in estimates of decay times, decay time uncertainties (quantified by their respective standard derivations), and decay-level differences. The decay parameters (decay times and decay-level differences in Table II) reveal double-sloped decaying characteristics in the primary room. Comparisons between measured results in scale-models and the simulation results listed in Table II indicate that the most reasonable agreement is found in two decay times associated with broad-band results. Standard deviations of two decay times indicate that uncertainties of measured decay times are one order of magnitude higher than those of the decay times estimated from the modeled decay functions. Discrepancies are found in individual octave-band analysis, which implies that diffusion-equation models are particularly valid in broad-band (1.5–4 kHz) modeling. Better agreement of the diffusion-equation model with the experimental results for high frequencies (1 and 4 kHz) has also been reported by Billon *et al.*[12]

The time location of turning points suffers from weaker agreement with experimental values consistently larger than those predicted by the diffusion-equation model. In addition to extra noise and the bumpy nature of measured results as potential sources of errors, the deviations of the measured values from the diffusion-equation modeled ones should be interpreted with caution. While the diffusion-equation model is based on the assumption of sufficient mixing of sound particles, such an assumption is not valid theoretically for a time interval within one mean free time, as pointed out by Morse and Feshbach,[28] and most recently by Valeau *et al.*[11] In contrast, the acoustic measurements of the room impulse responses contain full information including the direct sound, discrete early reflections, and reverberation tails. Sound-energy propagation within the discrete early-reflection portion cannot be correctly predicted by the diffusion-equation model. In considering the consistently larger turning point times evaluated from experimental results, the differences (especially in 1.5–4 kHz), on an order of two mean free times (2×10.5 ms), may suggest that the diffusion-equation model may be considered valid for the sound-energy prediction after at least two mean free times. The more consistent results from broad-band evaluations of SPL distributions, decay times, and level differences also suggest that the diffusion-equation model can be used in coupled-volume systems for high-frequency, broad-band prediction.

IV. SOUND-ENERGY FLOWS

According to Fick's law,[28] the gradient of the sound-energy density $w(\mathbf{r},t)$ at position $\mathbf{r}$ and time t in the room under investigation causes the sound-energy flow vector $\mathbf{J}$

$$\mathbf{J} = -D\,\nabla\, w(\mathbf{r},t), \tag{13}$$

with D the diffusion coefficient as given in Eq. (1). This section discusses the energy flow in the coupled spaces in terms of diffusion-equation modeling. Emphasis is given to the direction of the energy flow and level of the energy flow as it decays.

A. Energy flow directions

This section first discusses the energy flow directions, describing three cases. The natural reverberation times in the primary room are chosen to be smaller, the same as, and larger than the one in the secondary room, respectively (see Table III). The natural reverberation times are reverberation times in either of two rooms when standing alone as a single-room. Absorption coefficients in case 1 are from the mea-

TABLE III. Approximate values of the natural reverberation times in two rooms for energy flow demonstrations.

	Case 1	Case 2	Case 3
T_1 (s) (primary room)	0.45	0.45	0.45
T_2 (s) (secondary room)	1.05	0.45	0.30

surement at 2 kHz (see Table I). Absorption coefficients are adjusted for other cases so that the secondary room could be less or equally reverberant than the primary room. The room geometry remains the same as in the previous paragraphs.

A time-dependent solution of the diffusion equation [Eqs. (1), (2), and (3c)] using the source term in Eq. (4) at each observation point in the rooms initially leads to an energy-density impulse response; the energy flow is calculated according to Eq. (13) for each time step (5 ms), termed impulse-response derived energy flow in the following. Figure 9 illustrates the impulse-response derived energy flow directions for several representative time steps. The first time instant is 20 ms, chosen to be on order of two "mean free paths" in the rooms to ensure the validity of the diffusion equation.[11,28] The first column of Fig. 9 shows the case where the natural RT in the primary room is smaller, and energy feedback is found around $t=100$ ms by tracing the energy flow directions. The energy feedback implies that the sound-energy flows from the secondary room back to the primary room. The feedback energy dominates the decay process in the primary room but with a slower decay rate after $t=100$ ms. A double-sloped energy decay is, therefore, expected. For the other two cases, no energy flow has yet been found, indicating that the energy feedback depends on the overall decay rates in both rooms. Most recently Jing and Xiang[29] visualized the energy flow directions in form of two- and three-dimensional animations for the three different cases.

The steady-state sound-energy flow-direction changes are obtained by applying Eq. (13) to the steady-state derived energy (density) decay. Figure 10 illustrates the steady-state derived energy flow directions; only the first case ($T_1' < T_2'$) is shown since the energy feedback is present in this case. As opposed to the impulse-response derived sound-energy flow, the energy feedback occurs at different times along the steady-state sound-energy flow decay curve.

B. Energy flow decays

This section studies the energy flow decays. To generate the steady-state derived sound-energy flow decay, an assignment of a switch-off signal to the source term by

$$q(\mathbf{r}_s,t) = E_0\zeta(t), \tag{14}$$

with

$$\zeta(t) = \begin{cases} 1, & t \le 0 \\ 0, & t > 0, \end{cases} \tag{15}$$

will yield the steady-state derived energy decay function in solving the diffusion equation [Eqs. (1), (2), and (3c)]. Physically, the sound source is turned on for a long-enough period of time and is then switched off at a time point referred to as 0 ms, the solution of which is called the *switch-off* energy flow decay. In the numerical implementation, it requires a time-dependent solution already before $t=0$ in order to ensure the system arrives at the steady-state. Generally at least twice the computational load is needed in comparison with the time-dependent solution of the energy room impulse response, from which the so-called impulse-response derived sound-energy flow decay is derived. The energy flow level is defined[29] as

$$J_L(r,t) = 10\log_{10}\left\{\left[\frac{\partial w(r,t)}{\partial x}\right]^2 + \left[\frac{\partial w(r,t)}{\partial y}\right]^2 + \left[\frac{\partial w(r,t)}{\partial z}\right]^2\right\}^{1/2}. \tag{16}$$

The diffusion coefficient D in Eq. (13) is not considered since only the relative amplitude in each room is of major concern.

The energy flow decay shows a flipping-over characteristic in a certain area around the aperture; i.e., when the energy flow amplitude decays to a certain level, it ascends slightly and decays again in a different decay rate. The energy flow decay curve features two points worth mentioning: A "dip" is followed by a "peak." Figure 11 illustrates a typical energy flow curve obtained at $(-2, 2, 3)$ m and another typical curve without the dip at $(-2, 4, 3)$ m. The first receiver is close to the aperture opening while the second one is not. In this example, the absorption coefficients at 2 kHz are used to assign the boundary conditions for the modified diffusion-equation model.

The dip can be related to the turning point (see Sec. III D). After applying the Bayesian analysis[4,26] to the steady-state derived sound-energy decay function, two resulting decay slopes are used to estimate the turning point [see Eq. (12)]. Figure 11(b) illustrates the Bayesian model curve, slope-decomposition, and the turning point. Comparing the dip in Fig. 11(a) with that in Fig. 11(b), the turning point along the time-axis is close to the dip on the energy flow decay curve. The turning point is at $t=61$ ms while the dip is at $t=59$ ms. So far tested within the scope of the current work, the time occurrence of the turning point on the energy decay curve and that of the dip in the energy flow seem approximately correlated, which has also been found in other receiver locations as long as the feedback energy passes these locations, and dominate the primary energy decay. Table IV lists their specific time instants. In Bayesian analysis, the decay function can be decomposed into two exponential decays terms. The turning point then represents the intersection of two straight decay lines. Both the dip and the turning point indicate the time when the second energy decay starts to dominate the first energy decay. According to Table IV, the dip occurs around the turning point, sometimes slightly before/after the turning point. This is expected to be a result of uncertainties intrinsically residing in Bayesian analysis. The uncertainties quantified by $\mathrm{Std}_{1,2}$ of decay times $T_{1,2}$ also imply estimation uncertainties of the turning point.

Figure 12 illustrates the steady-state derived sound-energy flow decay in comparison with the impulse-response

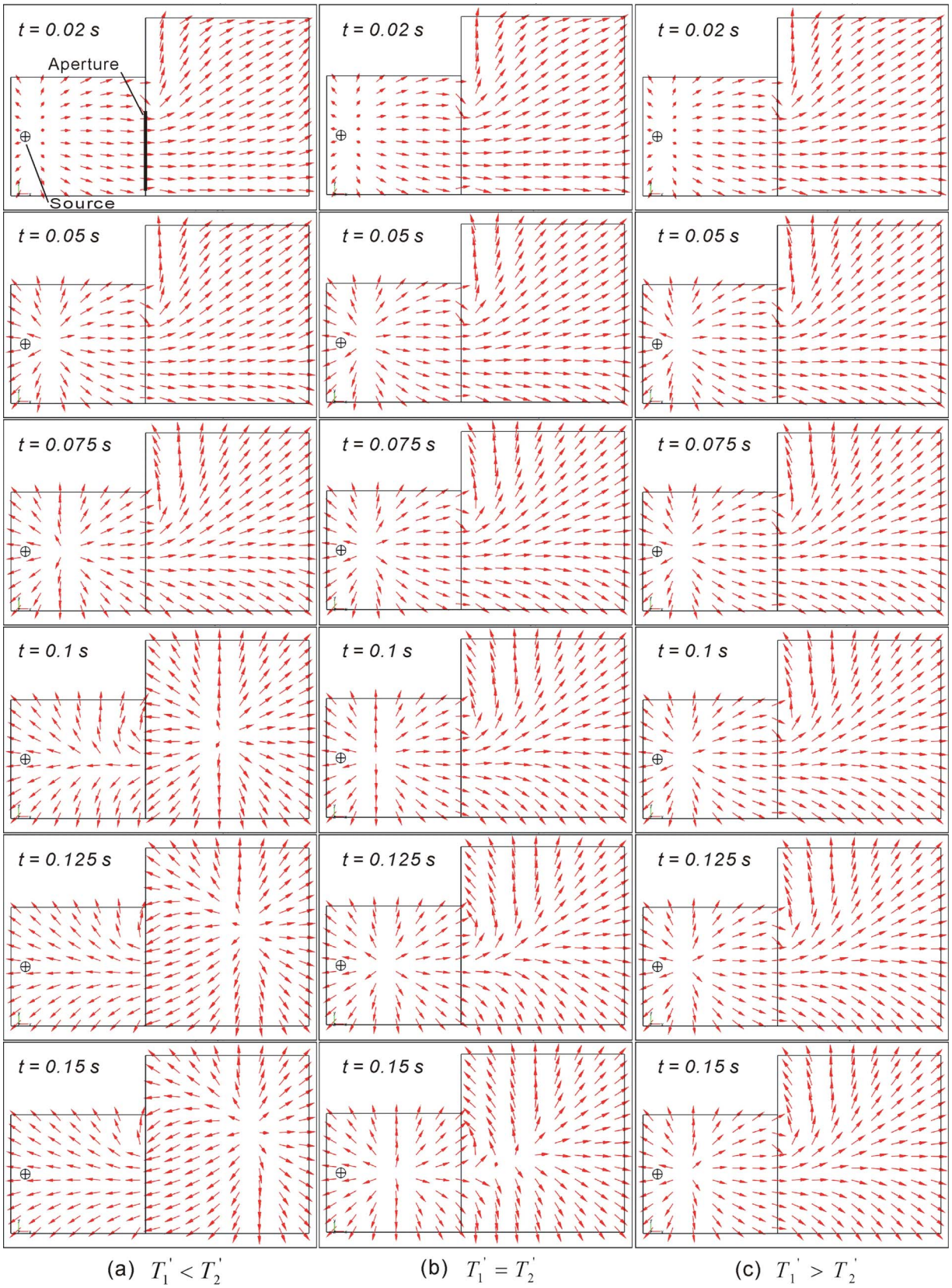


FIG. 9. (Color online) Two-dimensional mapping of impulse-response derived sound-energy flow vectors [Eq. (13)] for six different snapshots on x-y plane at $z=3$ m in the coupled-rooms; the dimension is given in Fig. 1(b). Three different acoustics conditions are characterized by the natural reverberation times (T_1', T_2') in two rooms. (a) $T_1' < T_2'$, (b) $T_1' = T_2'$, and (c) $T_1' > T_2'$.

derived sound-energy flow decay at receiver position (−2, 2, 3) m. To obtain the impulse-response derived sound-energy flow, an impulse source signal is used instead. The dips appear at different times, indicating the difference between impulse-response derived decays and steady-state derived decays, and the necessity of using Schroeder-integration for

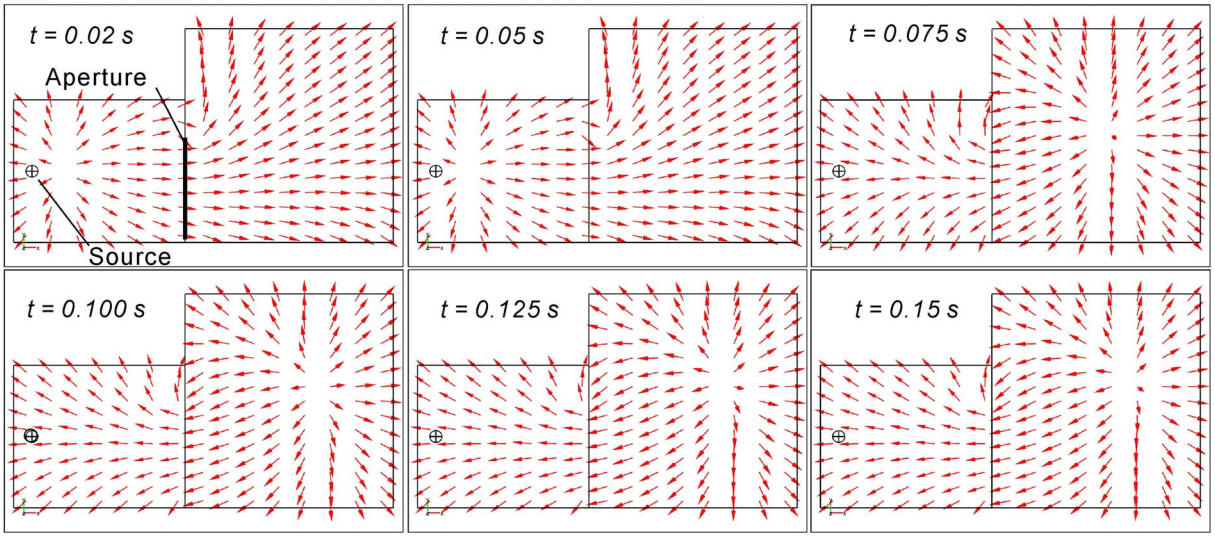


FIG. 10. (Color online) Two-dimensional mapping of steady-state derived sound-energy flow vectors [Eq. (13)] for six different snapshots on x-y plane at $z=3$ m in the coupled-rooms, the dimension is given in Fig. 1(b). Overall acoustic condition is characterized by the natural reverberation times (T_1', T_2') in two rooms: $T_1' < T_2'$.

the sound-energy decay analysis. From both impulse-response and steady-state derived sound-energy flow decays, it is found that the time when the energy flow reverses its direction is correlated with the time when the dip appears on the energy flow decay curve. For instance, in the impulse-response situation, the energy flow at receiver position (−2, 2, 3) m reverses between $t=75$ ms and $t=100$ ms (closer to $t=100$ ms) as shown in Fig. 9(a), while the dip appears at 95 ms.

Finally, the physical meaning of the dip in the energy flow decay curve is explained here. The sound energy initially flows from the primary room to the secondary room since the sound-energy density in the primary room is stronger. If the sound energy decays faster in the primary room than in the secondary room, the change in sign of the energy gradient across the aperture will, at some future point, indicate flow back to the primary room. This phenomenon manifests itself in magnitude plots such as in the energy flow decay curve, as a dip [see Fig. 11(a)].

The reversal of energy flow direction is due to the energy feedback. When feedback dominates the primary energy decay, flow directions reverse. Physically, the direction changes continuously. Thus, the energy flow magnitude decreases gradually until sound-energy flow from the secondary room dominates; it reverses direction (a dip appears) and increases beyond the dip. Eventually, the energy flow reaches a local peak value [see Fig. 11(a)]; beyond the local peak, the energy flow decays further. This flow-direction reversal and the dip in the energy flow decay process cannot exist in cases in which the primary room's natural reverberation time is longer than that of the secondary room.[29]

This section has discussed the sound-energy flows determined using the gradient of the time-dependent sound-energy density when solving the diffusion equation. Upon assignment of the source term using Eq. (4), the solution of the diffusion equation delivers an energy (room) impulse response, whose gradient, apart from a constant with a minus sign, is termed the impulse-response derived sound-energy

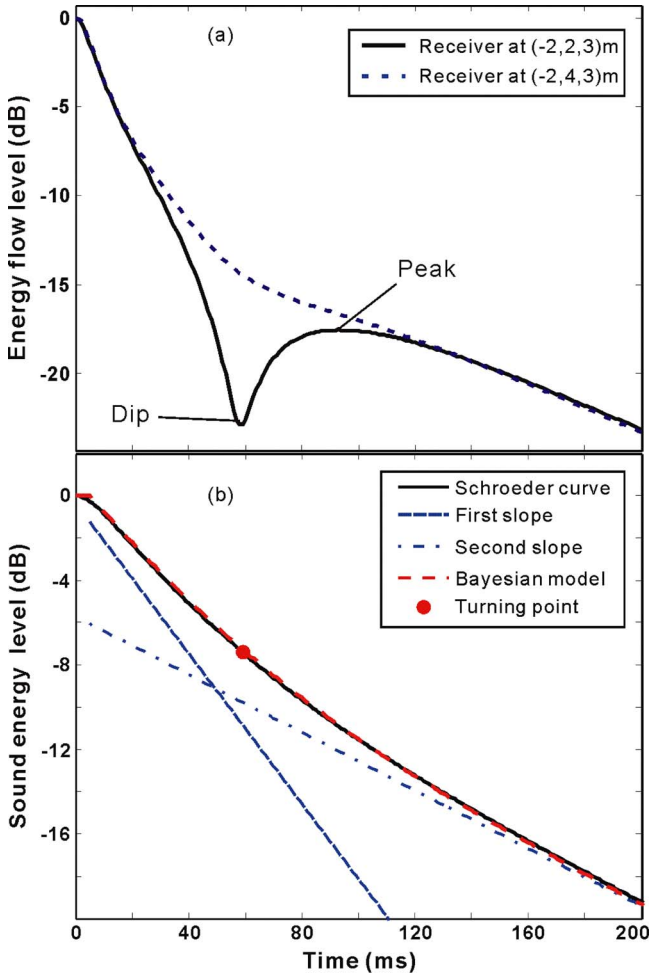


FIG. 11. (Color online) The normalized energy flow decays and the turning point along the normalized energy decay. (a) Energy flow decay at (−2, 2, 3) m with dip and peak and energy flow decay at (−2, 4, 3) m without dip and peak. (b) Bayesian decay decomposition as well as the turning point of the energy decay curve obtained at (−2, 2, 3) m. The Schroeder decay curve is predicted by the diffusion model while the Bayesian model results from the Bayesian decay-parameter estimation.

TABLE IV. Relationship between the turning point of the sound-energy decay and the dip location along the time-axis of the sound-energy flow decay.

Receiver position (m)	Turning point (ms)	Dip (ms)
−2, 3, 3	63	61
−1, 3, 3	56	61
−1, 2.5, 3	56	59
−2, 2.5, 3	58	58

flow. Steady-state derived sound-energy decay functions can also be derived upon assignment of the source term using a switch-off function; its gradient is termed, in this section, steady-state derived sound-energy flow. The energy flow-direction reversal is expressed as a dip in the energy flow decay function. The steady-state derived sound-energy flow decays need to be used to correlate the dip with the "turning point" in the Schroeder decay functions, which are the steady-state derived sound-energy decay functions.

V. CONCLUDING REMARKS

This paper discusses previously undiscovered, appealing characteristics predicted by a diffusion-equation model to two coupled spaces. Experimental results are first compared with the simulation results in terms of the SPL distribution along a line parallel and close to the aperture in both the primary room and the secondary room. Experimental results from the sound-energy decay analysis are also compared with the simulation results. Both SPL distributions and energy decay analysis show more consistent results, particularly in some high-frequency bands (1 and 2 kHz) and broad-band (1.5–4 kHz) data. Less agreeable results are found when comparing the time instants of turning points between the experimental and modeled results. The consistency of such differences hints at an intrinsic feature of the diffusion-equation model that suggests exclusion by the model of both the direct sound and discrete early reflections. While previous work proves the validity of the diffusion-equation model by showing the SPL distribution along lines across the aperture, this work contributes additional evidence supporting the application of the diffusion-equation model, particularly in high-frequency, broad-band modeling. This work also reveals differences between scale-model measurements and predictions from the diffusion equation when evaluating absolute time occurrence of the turning from the first decay process to the second one.

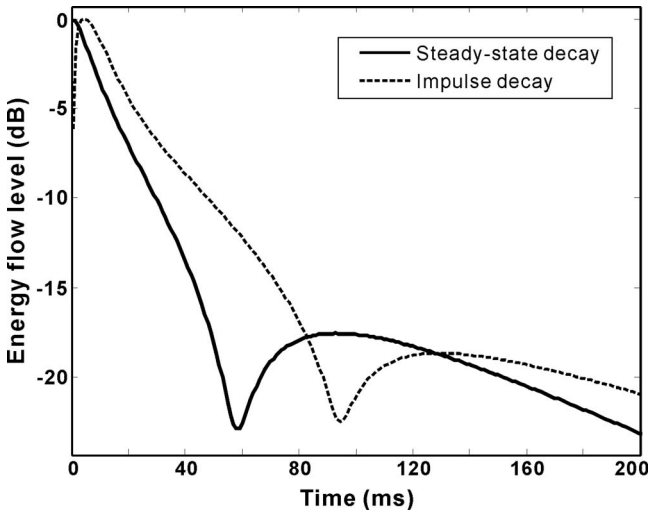


FIG. 12. Normalized steady-state derived energy flow decay and normalized impulse-response energy flow decay at the same receiver position (−2, 2, 3) m. The steady-state (so-called *switch-off*) energy flow decay is calculated by assigning the source to be a switch-off function while the impulse-response energy flow decay is calculated by assigning the source an impulsive function.

The time-dependent sound-energy flow is also studied, including the direction changes and the amplitude decay. The diffusion equation is inherently suitable to generate the energy flows that may be derived from the gradient of an efficiently obtained sound-energy density distribution for the room under study. The energy flow-direction changes show the distinct energy feedback when the natural reverberation time in the primary room is smaller than that in the secondary room. The energy flow-direction reversal is expressed as a dip on the energy flow decay curve, which correlates with the turning point on the double-sloped sound-energy decay extracted from the Bayesian analysis. In addition, both impulse response and steady-state, switch-off response are investigated for the time-dependent energy flow, revealing the intrinsic difference between these two responses.

All results discussed in this paper have been restricted to the examples given up to now. Particularly the scale-model experiments within the scope of this study are limited up to 4 kHz in full-scale. A generalization must be carefully examined in future investigations. Future research is expected in following directions: (1) studying the location dependence of double-sloped energy decays with respect to sound source-receiver arrangement relative to the coupling aperture; (2) studying the aperture effect on the energy decay in coupled spaces by systematically changing the aperture size, shape, and location; (3) experimental comparison and verification in real-sized spaces, real concert halls with coupled volumes; (4) refining the valid frequency range/limit for the diffusion-equation model; (5) investigation into the absolute time occurrences of turning points in energy decay in the diffusion-equation model; and (6) experimental verification of energy flows, which may pose experimental challenges.

ACKNOWLEDGMENTS

The authors are grateful to Professor William Siegmann, Professor Joel Plawsky, Dr. Jason Summers, Dr. Vincent Valeau, Dr. Christopher Jaffe, and Mr. Tomislav Jasa for their stimulating discussions. The authors would like to thank Zuhre Su and Rolando de la Cruz for their effort in collecting experimental data on the 1:8 scale-models. The authors are also grateful to Associate Editor Professor Lily Wang and the anonymous reviewers whose constructive comments greatly helped in improving clarity from the early version of the manuscript.

[1]J. C. Jaffe, "Selective reflection and acoustic coupling in concert hall design," *Proceedings of the Music and Concert Hall Acoustics, MCHA 1995*, edited by Y. Ando and D. Noson (Academic, New York, 1995), pp.

85–94.
[2]R. Johnson, E. Kahle, and R. Essert, "Variable coupled volume for music performance," *Proceedings of the Music and Concert Hall Acoustics, MCHA 1995*, edited by Y. Ando and D. Noson (Academic, New York, 1995), pp. 372–385.
[3]J. C. Jaffe, "Innovative approaches to the design of symphony halls," Acoust. Sci. & Tech. **26**, 240–243 (2005).
[4]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1424 (2001).
[5]J. E. Summers, R. R. Torres, and Y. Shimizu, "Estimating mid-frequency effects of aperture diffraction on reverberant-energy decay in coupled-room auditoria," Build. Acoust. **11**, 271–291 (2004).
[6]J. E. Summers, R. R. Torres, Y. Shimizu, and B.-I. L. Dalenbäck, "Adapting a randomized beam-axis-tracing algorithm to modelling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**, 1491–1502 (2005).
[7]M. Meissner, "Computational studies of steady-state sound field and reverberant sound decay in a system of two coupled rooms," Cent. Eur. J. Phys. **5**, 293–312 (2007).
[8]Ch. Thompson, "On acoustics of a coupled space," J. Acoust. Soc. Am. **75**, 707–714 (1984).
[9]F. Ollendorff, "Statistical room-acoustics as a problem of diffusion: A proposal," Acustica **21**, 236–245 (1969).
[10]J. Picaut, L. Simon, and J. D. Ploack, "A mathematical model of diffuse sound field based on a diffusion equation," Acust. Acta Acust. **83**, 614–621 (1997).
[11]V. Valeau, J. Picaut, and M. Hodgson, "On the use of a diffusion equation for room-acoustic prediction," J. Acoust. Soc. Am. **119**, 1504–1513 (2006).
[12]A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," J. Acoust. Soc. Am. **120**, 2043–2054 (2006).
[13]V. Valeau, M. Hodgson, and J. Picaut, "A diffusion-based analogy for the prediction of sound fields in fitted rooms," Acust. Acta Acust. **93**, 94–105 (2007).
[14]Y. Jing and N. Xiang, "A modified diffusion equation for room-acoustic predication (L)," J. Acoust. Soc. Am. **121**, 3284–3287 (2007).
[15]A. Billon, J. Picaut, and A. Sakout, "Prediction of the reverberation time in high absorption room using a modified-diffusion model," Appl. Acoust. **69**, 68–74 (2008).
[16]Y. Jing and N. Xiang, "On boundary conditions for the diffusion equation in room-acoustic prediction," J. Acoust. Soc. Am. **123**, 145–153 (2008).
[17]C. Foy, V. Valeau, A. Billon, J. Picaut, and A. Sakout, "An empirical diffusion model for acoustic prediction in rooms with mixed diffuse and specular reflections," Acust. Acta Acust. **95**, 97–105 (2009).
[18]A. Billon, J. Picaut, C. Foy, V. Valeau, and A. Sakout, "Introducing atmospheric attenuation within a diffusion model for room-acoustic predictions (L)," J. Acoust. Soc. Am. **123**, 4040–4043 (2008).
[19]N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayesian decay time estimation," J. Acoust. Soc. Am. **117**, 3707–3715 (2005).
[20]N. Xiang and J. Blauert, "Binaural scale modelling for auralization and prediction of acoustics in auditoria," Appl. Acoust. **38**, 267–290 (1993).
[21]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1981).
[22]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).
[23]L. Cremer, H. A. Mueller, and T. J. Schultz, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982).
[24]L. Nijs, G. Jansens, G. Vermeir, and M. van der Voorden, "Absorbing surfaces in ray-tracing programs for coupled spaces," Appl. Acoust. **63**, 611–626 (2002).
[25]D. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," J. Acoust. Soc. Am. **118**, 1480–1490 (2005).
[26]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**, 3744–3749 (2006).
[27]ISO, "Acoustics—Measurement of the reverberation time of rooms with reference to other parameters," 3382 (1997).
[28]Ph. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1955).
[29]Y. Jing and N. Xiang, "Visualizations of sound energy across coupled rooms using a diffusion equation model," J. Acoust. Soc. Am. **124**, EL360–365 (2008).

The variance of the discrete frequency transmission function of a reverberant room[a)]

John L. Davy[b)]
School of Applied Sciences, RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, Australia

(Received 28 May 2009; revised 23 June 2009; accepted 30 June 2009)

This paper first shows experimentally that the distribution of modal spacings in a reverberation room is well modeled by the Rayleigh or Wigner distribution. Since the Rayleigh or Wigner distribution is a good approximation to the Gaussian orthogonal ensemble (GOE) distribution, this paper confirms the current wisdom that the GOE distribution is a good model for the distribution of modal spacings. Next this paper gives the technical arguments that the author used successfully to support the pragmatic arguments of Baade and the Air-conditioning and Refrigeration Institute of USA for retention of the pure tone qualification procedure and to modify a constant in the International Standard ISO 3741:1999(E) for measurement of sound power in a reverberation room. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3184568]

PACS number(s): 43.55.Cs, 43.55.Br, 43.55.Nd, 43.50.Cb [RLW] Pages: 1199–1206

I. INTRODUCTION

This paper gives an equation for the relative covariance of the transmission function of a reverberant room. The equation depends on the distribution of the modal frequency spacings. This paper describes experimental measurements of the modal frequency spacings in a reverberation room and the analysis of these measurements which indicate that the Gaussian orthogonal ensemble (GOE) is a good model of the modal frequency spacings.

The 1996 version of the draft international standard ISO/DIS 3741, "Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for reverberation rooms," (ISO, 1996) deleted the room qualification procedure for the measurement of discrete frequency components. The alternative multiple source position method was retained. This paper shows that there was an error in the constant in the equation for determining the number of source positions in the retained alternative multiple source position method. It also shows that the multiple source position method is not sufficient at low modal overlap. Thus the room qualification procedure needed to be reinstated. The arguments in this paper were presented to the ISO Working Group which was revising and combining ISO 3741:1988(E) and ISO 3742:1988(E) (ISO, 1988). This resulted in the room qualification procedure for the measurement of discrete frequency components being reinstated and the error in the constant being corrected in ISO 3741:1999(E) (ISO, 1999).

The measurement variance can be split into source position, receiver position, and room variance. The room variance depends on the distribution of modal spacings. Earlier theoretical and numerical calculations used the Poisson or "nearest neighbor" distributions. Both these distributions produce nonzero room variance. The GOE distribution, which is currently believed to be correct, produces zero room variance at high modal overlap. At low modal overlap, the GOE and nearest neighbor distributions produce room variance values which tend toward the nonzero values produced by the Poisson distribution.

II. THEORY

The transmission function of a reverberation room is defined to be the square of the modulus of the ratio of the reverberant field sound pressure at a point in the room to the volume velocity of the sound source. The case considered is where LN measurements of the transmission function are made from each of N sources positions to each of L receiver positions and the LN measurements are averaged before further statistical calculations are made. These further statistical calculations would typically be the calculation of means, variances, or covariances across excitation frequency or room shape. Theoretical work by Lyon (1969), Davy (1981b), and Weaver (1989a) has shown that if the transmission function is averaged over an array of N source positions and L receiver positions, the relative covariance of the averaged transmission function at two angular frequencies which differ by θ is given by

$$\mathrm{relcov} = \phi(\theta)\left\{\frac{1}{LN} + \frac{1}{M}\left[\left(\frac{K-1}{N}+1\right)\left(\frac{K-1}{L}+1\right) - C\left(\frac{2}{LN}+1\right)\right]\right\}, \quad (1)$$

where

[a)]Portions of this work were presented in "The distribution of modal frequencies in a reverberation room," Science for Silence—Proceedings of Inter-Noise 90 Conference, edited by H. G. Jonasson, Gothenburg, Sweden, 13–15 August 1990, Vol. 1, pp. 159–164 and in "The variance of pure tone reverberant sound power measurements," Fifth International Congress on Sound and Vibration, University of Adelaide, Adelaide, Australia, 15–18 December 1997.

[b)]Electronic mail: john.davy@rmit.edu.au. Also at CSIRO Materials Science and Engineering, P.O. Box 56, Highett, Victoria 3190, Australia.

$$\phi(\theta) = \frac{1}{\left[1 + \left(\frac{\theta}{2\gamma}\right)^2\right]}, \tag{2}$$

$$M = 2\pi n \gamma, \tag{3}$$

$$K = \frac{\langle p^4(x)\rangle}{\langle p^2(x)\rangle^2}, \tag{4}$$

γ is the decay rate of the modal amplitudes in nepers per unit of time, n is the modal density in number of modes per unit of angular frequency, $p(x)$ is the modal amplitude as a function of position x in the room, and C is a function of the distribution of the modal frequency spacings. The angular brackets $\langle\ \rangle$ in Eq. (4) denote the average value over position x in the room. ϕ is Schroeder's (1987a, 1987b) frequency autocorrelation function with angular frequency as the argument and M is the statistical modal overlap which is the product of the modal density with the statistical bandwidth of the modes. The statistical bandwidth of a mode is twice the effective or noise bandwidth of the mode and π times the half power or 3 dB bandwidth of the mode. For a rectangular parallelepiped room with rigid walls, K is equal to $(3/2)^3$, $(3/2)^2$, or (3/2) for oblique, tangential, or axial modes, respectively. C is equal to 0, 1/2, or 1 for Poisson, nearest neighbor, or GOE distributions of modal frequency spacings. Legrand *et al.* (1995) (Legrand and Mortessagne, 1996) showed that, while Schroeder's (1987a, 1987b) frequency autocorrelation function is correct for the Poisson case, it needs to be replaced with $(1-(\theta/2\gamma)^2)/[1+(\theta/2\gamma)^2]^2$ in the GOE version of the covariance of the real part of the input impedance case (the number of receiver positions L equals infinity). Because this paper is only concerned with $\theta=0$, this correction will not be considered further in this paper.

Equation (1) is only correct for the nearest neighbor and GOE cases if the statistical modal overlap is not too low. As the statistical modal overlap tends to zero, the relative covariance tends to that given by the Poisson version of Eq. (1), where $C=0$ regardless of the distribution of the modal frequency spacings (Lyon, 1969; Davy, 1981b; Weaver, 1989a; Lobkis *et al.*, 2000; Langley and Cotoni, 2005). This is because the actual distribution of modal spacings only has an effect on the relative covariance of the pure tone transmission function if the modal responses are likely to overlap significantly in the frequency domain. Equation (1) does not include the increase in the theoretical relative variance due to the variability of the decay rates of the modes (Burkhardt and Weaver, 1996) because this increase usually makes the agreement between theory and experiment worse. A good review of this research area is given in Sec. 3.2.4 of Tanner and Sondergaard, 2007.

The Poisson or exponential distribution of modal frequency spacings results if the modal frequencies are distributed independently of each other. If the mean value is normalized to 1, the probability density function is e^{-x}, the cumulative distribution function is $1-e^{-x}$, and the fraction of values in the bin from x to y is $e^{-x}-e^{-y}$. The fluctuations of the pure tone transmission function of a reverberant room over frequency are also distributed according to the Poisson or exponential distribution (Schroeder, 1987b).

The other two distributions result if the modal frequencies repel each other. The nearest neighbor distribution was adopted by Lyon (1969) from early experimental work on the distribution of energy levels in atomic nuclei, and used mainly because it simplifies the mathematics since only the exponential of x and not x^2 is involved. If the mean value is normalized to 1, the probability density function is $4xe^{-2x}$, the cumulative distribution function is $1-(1+2x)e^{-2x}$, and the fraction of values in the bin from x to y is $(1+2x)e^{-2x}-(1+2y)e^{-2y}$.

The GOE distribution of spacings does not have an elementary function representation. According to Weaver (1989b), it can be well approximated by the Rayleigh or Wigner distribution. If the mean value is normalized to 1, the probability density function of the Rayleigh or Wigner distribution is $(\pi x/2)\exp(-\pi x^2/4)$, the cumulative distribution function is $1-\exp(-\pi x^2/4)$, and the fraction of values in the bin from x to y is $\exp(-\pi x^2/4)-\exp(-\pi y^2/4)$. The fluctuations of the square root of the pure tone transmission function of a reverberant room over frequency are also distributed according to the Rayleigh or Wigner distribution (Schroeder, 1987b). According to Weaver (1989a), recent studies on the distribution of the spacings of energy levels of atomic nuclei suggest than the GOE spacing distribution should apply to the spacings of modal frequencies in reverberation rooms. Equation (2) of Lyon, 1969 is the probability density function for the Rayleigh or Wigner distribution. Note that the constants in this Eq. (2) of Lyon, 1969 are not correct if $\langle E\rangle$ is interpreted as the mean of the distribution. The mean of the distribution given by Eq. (2) of Lyon, 1969 is $\sqrt{\pi/2}\langle E\rangle$.

The conditional modal density $n(\omega_l|\omega_m)$ is the modal density at ω_l given that there is a mode at ω_m. If the modes are distributed independently of each other, $n(\omega_l|\omega_m)$ equals the modal density at ω_l, namely, $n(\omega_l)$. If the modal density is normalized to 1, the conditional modal density $n(\omega_l|\omega_m)$ can be written as $1-Y(\omega_l-\omega_m)$, where Y is the two point cluster function. $Y(x)$ is equal to 0, $e^{-4|x|}$, or $s^2(x)-J(x)D(x)$, respectively, for the exponential, nearest neighbor, or GOE spacing distributions. D is the derivative of s, $s(x)$ equals $\sin(\pi x)/(\pi x)$, and

$$J(x) = \int_0^x s(y)dy - \mathrm{sgn}(x)/2. \tag{5}$$

The function $\mathrm{sgn}(x)$ is equal to 1, 0, or -1 if x is positive, zero, or negative, respectively. The constant C in Eq. (1) is the integral of Y from $-\infty$ to ∞.

Thus it can be seen that one needs to know the distribution of modal frequency spacings in order to be able to apply Eq. (1). Comparison of Eq. (1) with experimental results suggests that the GOE spacing distribution is the most appropriate distribution to use. This is because Eq. (1) tends to over-estimate experimental results and the GOE spacing distribution gives the lowest results. However, it is still desirable to make a direct determination of the distribution of modal frequency spacings. This is not easy to do because the

modes can only be separated if the modal overlap is less than 1. This explains why there have not been any previous experimental determinations of room modal frequency spacing statistics. Schroeder (1987a) used electromagnetic microwaves in metallic cavities, while Weaver (1989b) used ultrasonic vibrations in solid blocks. These two approaches enabled them both to obtain the necessary low values of modal overlap.

III. EXPERIMENTS

The measurements described in this paper were made in a 607 m^3 reverberation room between 14 and 90 Hz. The room has an irregular pentagonal floor plan and a sloping ceiling, but its walls are vertical. Its shell is constructed of 300 mm thick reinforced concrete. The measurements were made with the room in four different configurations. The first two configurations were the bare room (denoted as bare-empty) and the bare room with 22.5 m^2 of 50 mm thickness 100 kg/m^3 density mineral wool on the floor of the room (bare-sillan). The mineral wool was sold under the trade name of Sillan. The second two configurations were with 32 diffusing panels added to the room. The total area (all sides) of these panels and other diffusing surfaces in the room was 141 m^2. The measurements were again made without (diff-empty) and with the Sillan (diff-sillan).

A Marconi type TF2101 Wien bridge oscillator was used above 30 Hz because of its good frequency stability. Below 30 Hz an Exact type 250 function generator was used. The frequency was monitored by a Racal type SA520 frequency counter running in period mode. The signal was fed to Celestion type G18C 450 mm diameter loudspeaker via a Leak TL/25 Plus power amplifier. The loudspeaker was mounted in a small box and placed in one of the floor corners facing into the corner. The output voltage of the Leak amplifier was maintained at 10 V by monitoring it with a B&K type 2409 voltmeter. A B&K type 4131 1 in. condenser microphone and its associated microphone preamplifier were placed in the only right-angle corner of the room and powered from a B&K type 2107 frequency analyzer. The output of the analyzer was displayed on a B&K type 2301 or 2305 high speed level recorder and on a BWD type 502 oscilloscope. Because the frequency analyzer could only be tuned down to 20 Hz, a low pass filter with a 3 dB down point of 32 Hz was used between the analyzer and the display devices for frequencies less than 20 Hz. This was necessary to avoid detecting the excitation of higher frequency modes by higher harmonics of the test signal. The oscilloscope was also useful for this purpose. For the same reason, the analyzer was used in its maximum selectivity mode for measurements above 20 Hz. This gives a bandwidth of about 3%.

A thermohydrograph was used to monitor the temperature and humidity of the room and a fan was run between measurements to stir the air in the room in order to avoid stratification. The measured peaks in the frequency response of the room were assumed to correspond to the modal frequencies. Because of the irregular shape of the room, degenerate modal frequencies were considered to be unlikely. All the measurements were performed at temperatures close to 10 °C. Since the modal frequencies vary with temperature because of variation in the speed of sound, the modal frequencies were converted to wavelengths using the measured air temperature and then converted to equivalent frequencies for a temperature of 10 °C.

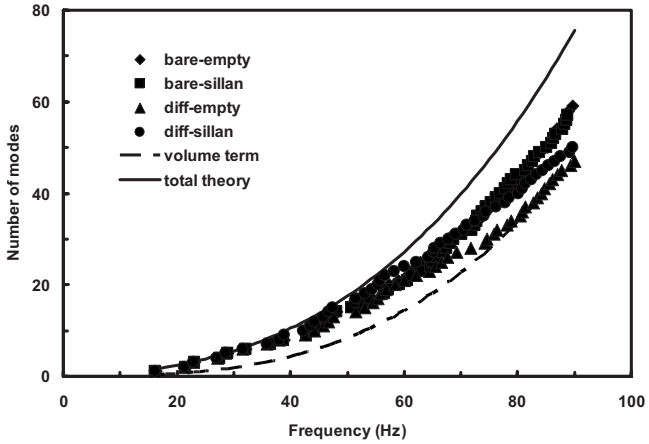


FIG. 1. The number of modes less than a given frequency.

IV. ANALYSIS OF EXPERIMENTAL RESULTS

The number of modes less than a given frequency is plotted in Fig. 1 for all four room configurations. The values for the two cases without diffusers agree fairly well with each other, but the two cases with diffusers differ from each other and the first two cases above 40 Hz. This was quite unexpected. It had originally been expected that any major differences between the measured values would be due to modal frequencies having been missed. It had been planned to insert missing modes by comparing the four cases. However, it now became apparent that this would lead to major subjective additions to the measured results and it was decided to make no additions to the measured frequencies. Also shown in Fig. 1 are the theoretical asymptotic equation for a rectangular parallelepiped room with rigid walls and the same equation if only the room volume term is included. Again it is surprising that all the measured values are less than the theoretical equation. However, they are all greater than the volume term except for the diff-empty case near 90 Hz. It is hard to ascribe these results to missed modal frequencies when the diff-sillan case, with the highest modal overlap, is greater than the other three cases around 60 Hz and greater than the diff-empty case above 40 Hz. It should be pointed out that the theoretical equation is only valid for rigid walls. Balian and Bloch (1970) showed that the surface term can vary between plus three times and minus the rigid wall term for a lossless wall as the imaginary part of the admittance of the wall varies between $-\infty$ and $+\infty$. The author is not aware of any theoretical treatment for lossy walls.

It had originally been intended to use the theoretical equation for normalizing the measured frequency spacings to unity. However, because of the differences between theory and experiment this was not possible. An approach similar to that used by Weaver (1989a, 1989b) was followed. Because the theoretical equation is a cubic in frequency, a cubic polynomial in frequency was fitted to each case using the method

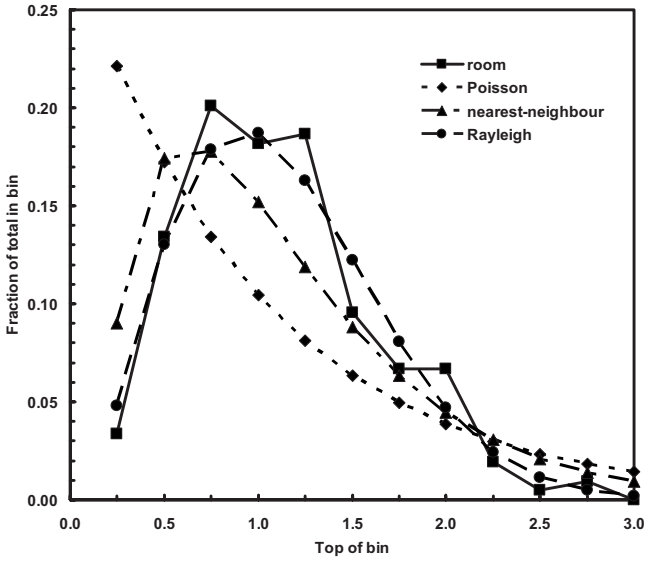


FIG. 2. Fraction of total number of normalized modal frequencies in each bin.

of least-squares. The respective cubic equation was then used to transform the measured frequencies for each particular case into a sequence of normalized frequencies. The differences between adjacent frequencies were calculated and these differences were grouped into 11 bins of 0.25 width from 0 to 2.75 and a bin containing those values greater than 2.75. The distribution of values in the bins for each case was similar so the numbers in each bin for the four cases were added together. The number in each bin was then divided by the total number of normalized frequencies to obtain the fraction of the total number in each bin. These experimentally determined fractions were then compared with the theoretical fractions for the Poisson, nearest neighbor, and Rayleigh distributions. The Rayleigh distribution was used as an approximation to the GOE spacing distribution.

The result is shown in Fig. 2. The horizontal axis shows the top of each bin of width 0.25 except for 3.0 which is the bin for all values greater than 2.75. It can be seen that the Rayleigh distribution agrees with the experimental results much better than the other two distributions. The goodness of fit was compared by calculating chi-squared values. This was done by calculating the square of the difference between the experimental and theoretical fractions and dividing by the theoretical fraction for each bin. The values so obtained were summed across each bin and the total multiplied by the total number of experimental values. The values so obtained were 109.7, 31.2, and 8.3 for the Poisson, nearest neighbor, and Rayleigh distributions. These values should be distributed as chi-squared with 11 degrees of freedom if the particular distribution is correct. This is Pearson's chi-squared goodness-of fit test [see Eq. (30.5) of Kendall and Stuart, 1967]. The 109.7 and 31.2 lie outside the 99% confidence limits while the 8.3 lies within the 50% confidence limits. Hence statistically the Rayleigh distribution is the only one that is likely to be correct. The root mean square (rms) deviation of the theory from the experiment was also calculated. The results were 0.074, 0.032, and 0.015 for the Poisson, nearest neighbor, and Rayleigh distributions, respectively, which again shows that the Rayleigh distribution agrees best with experiment. The total number of room modal frequency spacings from the four different room configurations used to generate the room curve in Fig. 2 was 209.

The data of other authors on modal frequency spacing were analyzed in the same manner. Weaver (1989b) had 313 ultrasonic modal frequency spacings from two different solids in 12 different bins. His chi-squared values were 142.2, 35.1, and 4.7, while the rms deviations were 0.068, 0.025, and 0.01. The 142.2 and 35.1 are outside the 99% confidence limits while the 4.7 is within the 90% confidence limits. Lyon (1969) quoted data gathered by Gurevich and Pevsner (1957) on 63 energy level spacings from several heavy atomic nuclei in 15 bins. Their chi-squared values were 31.9, 11.5, and 16.3, while the rms deviations were 0.059, 0.029, and 0.025. The 31.9 is outside the 99% confidence limits while the 11.5 and 16.3 both lie inside the 50% confidence limits. Thus in this we can reject the Poisson distribution but cannot choose between the nearest neighbor or Rayleigh distribution. This means that the data that Lyon (1969) used to justify the choice of the nearest neighbor distribution can equally well be used to justify the choice of the Rayleigh distribution and hence of the GOE spacing distribution. Schroeder (1987a) used 228 microwave modal frequency spacings grouped into 15 bins. His chi-squared values were 95.9, 29.4, and 32.7, while his rms deviations were 0.054, 0.020, and 0.020. The 95.9 and 32.7 are outside the 99% confidence limits and the 29.4 is outside the 98% confidence limits. Thus Schroeder's (1987a, 1987b) data do not agree with any of the distributions examined in this paper.

V. THE MULTIPLE SOURCE POSITION METHOD

Equation (3) of ANSI, 1980 is used to compute the number of source positions to be used in the multiple source position method for measurement of sound power in a reverberant room. This equation also appears as Eq. (4) of ISO, 1996. Baade asked for clarification of the statement in Davy, 1989 that "It was also shown that the value of the constant 0.79 in equation (3) of ANSI (1980) is wrong because of an error in Lyon's (1969) paper." It is shown in the following that the constant should be approximately 1.

Using the notation of Eq. (1) above, Eq. (3) of ANSI, 1980 can be reorganized to read

$$\frac{1}{B} \geq \frac{1}{LN} + \frac{1}{M}\frac{Ka}{N}, \tag{6}$$

where B is the constant K of Eq. (3) of ANSI, 1980, $K=27/8$, and $a=1/2$. In other words, the relative variance of the averaged transmission function of the reverberation room must be less than $1/B$. Comparison of the right hand side of Eq. (6) with the right hand side of Eq. (1) shows that Eq. (6) cannot be theoretically correct. This is because the term which multiplies $1/M$ does depend on the number of receiver positions L, and cannot be expressed as a constant divided by N, the number of source locations.

However, it will be assumed that L is large enough so that it can be set to infinity in the term which multiplies $1/M$. Setting θ to zero in Eq. (1) gives

$$\text{relvar} = \frac{1}{LN} + \frac{1}{M}\left(\frac{K-1}{N} + 1 - C\right). \tag{7}$$

This approximation is reasonable because the term $[(K-1)/L]+1$ tends to 1 as L increases and becomes almost independent of L for large values of L. On the other hand, the term $1/LN$ continues to decrease as L increases.

For the GOE distributions of modal frequency spacings, which is now believed to be correct, Eq. (7) becomes

$$\text{relvar} = \frac{1}{LN} + \frac{1}{M}\frac{K-1}{N} \tag{8}$$

since $C=1$ for this case. The right hand side of Eq. (6) agrees with the right hand side of Eq. (8) except for the fact that K should have one subtracted from it instead of being multiplied by $a=1/2$.

Averaging over all possible receiver positions enables a true estimate of the sound power actually injected into the room. Setting the number of receiver positions L to infinity, the number of source positions N to 1, and the angular frequency difference θ to 0 in Eq. (1) gives the relative variance of the real part of the input impedance of a reverberation room,

$$\text{relvar} = \frac{K-C}{M}. \tag{9}$$

Lyon (1969) obtained this equation with the correct value C equals zero in the Poisson case. In the nearest neighbor case, he obtained this equation with Ka instead of $K-C$, where C equals 1/2. For high modal overlap Lyon's (1969) a is equal to 1/2, and this is the value used in Eq. (3) of ANSI, 1980.

Assuming a rectangular parallelepiped room with rigid walls, and ignoring tangential and axial modes, K is equal to $(3/2)^3=27/8$. Lyon's (1969) value for Ka, as used in the standard, is then 27/16. In the nearest neighbor case $K-C=27/8-1/2=23/8$ and the constant needs to be multiplied by $(23/8)/(27/16)=46/27=1.70$. In the now accepted GOE case $K-C=27/8-1=19/8$ and the constant needs to be multiplied by $(19/8)/(27/16)=38/27=1.41$. Weaver (1989a) stated that "This author is inclined somewhat to $K=3.0$ which is appropriate for a Gaussian distribution of amplitudes and based on vague arguments invoking the central limit theorem." For $K=3$ and GOE case, the constant needs to be multiplied by $2/(27/16)=32/27=1.19$.

For the Poisson case, Lyon (1969) derived equations for the relative covariance of the real part of the input impedance and for the relative covariance of the transmission function. For the nearest neighbor case, he derived an incorrect equation for the relative covariance of the real part of the input impedance. Waterhouse (1978) published a paper giving theoretical equations which were very different from those derived by Lyon (1969).

The main purpose of Davy (1981b) was to reject Waterhouse's (1978) paper and to support Lyon's (1969) paper both theoretically and experimentally. While doing so, Davy found and corrected Lyon's (1969) error in the equation for the relative covariance of the real part of the input impedance in the nearest neighbor case. One of the puzzles of Lyon's (1969) paper was that it should have been possible to combine his equations for the covariance of the real part of the input impedance and the covariance of the transmission function by deriving the covariance of the transmission averaged over a number of source and receiver positions. It was not obvious from Lyon's (1969) paper how to do this. In fact, Eq. (3) of ANSI (ANSI, 1980) is based on a reasonable but incorrect guess of how to combine the equations.

Equation (3) of ANSI, 1980 is based on Eq. (6) of Maling, 1973. Section 2.2 of Lubman, 1974 attributes this guess to Andres. The correct way to combine the variances is shown by Eq. (1) with the angular frequency difference θ set equal to zero. However, because the number of receiver positions L is normally relatively large, Eq. (8) shows that the form of the incorrect guess is approximately correct in the GOE case for high modal overlap where C is equal to 1. Only the constant needs to be changed in Eq. (3) of ANSI, 1980. Note that the above assumptions make the room variance zero. The argument against the multiple source position method is that this room variance is not zero at low frequencies.

The main contribution of Davy (1981b) was to show how to combine these equations in the Poisson case. Like Lyon (1969), Davy (1981a, 1981b) was unable to derive a equation for the relative covariance of the transmission function in the nearest neighbor case. Davy (1981a, 1981b) guessed that the equation was obtained from the Poisson case by replacing K with $K-1/2$, which he had shown was true for the equation for the covariance of the real part of the input impedance.

Davy (1987) used the data from seven experiments based on the pure tone qualification procedure, to calculate the value of K which gave his equation, for the covariance of the averaged transmission function, the best fit to the experimental data. In these experiments, the angular frequency difference was 0, the number of source positions averaged over was 1, and the number of independent receiver positions increased linearly over the frequency range from 100 to 630 Hz because of the use of a circular microphone traverse. Davy (1987) obtained the value $K-C$ equals 2.16. If tangential and axial modes were ignored, Davy's (1987) theoretical estimates were $K=3.375$ for the Poisson case and $K-0.5=2.875$ for nearest neighbor case. If tangential and axial modes were included, Davy's (1987) theoretical estimates were $K=3.10$ for the Poisson case and $K-0.5=2.60$ for the nearest neighbor case.

Weaver (1989a) pointed out that the GOE distribution was more appropriate, and derived an equation for the covariance of the averaged transmission function in the GOE case. His method also applied to the nearest neighbor case, and showed that Davy's (1987) guess for the covariance of the average transmission function in this case was incorrect. Weaver's (1989a) equation alters the form of the equation from Davy's (1987) equation and not just the value of K. However, if number of receiving positions is large, Eq. (8)

shows that it replaces K with $K-1$. Thus the theoretical estimates in Davy, 1987 for the GOE case become $K-1=2.375$ and $K-1=2.10$, depending on whether tangential and axial modes are excluded or included. If Weaver's (1989a) estimate of K equals 3.0 is accepted, then $K-1$ equals 2.0. Hence the GOE values agree well with the experimental result of $K-1=2.16$ from the pure tone qualification procedure. It should be noted that there are still some problems predicting the results of other measurements (Davy, 1987).

The 2.10 theoretical value and the 2.16 experimental value depend on the percentage of tangential and axial modes. Thus they depend on room volume and frequency. It must be borne in mind that the above results are for a 607 m^3 reverberation room. Reverberation rooms will normally be smaller than this volume. Thus these values would be expected to be slightly smaller in smaller reverberation rooms. On the other hand, Eq. (8) gives results which are slightly too small because the number of receiver positions has been set equal to infinity in the second term. To avoid the need to calculate the percentage of axial and tangential modes, the use of the $2.375=19/8$ value for $K-1$ in Eq. (8) is suggested. As shown above this means that the 0.79 constant should be multiplied by 1.41 to give 1.11. It is further suggested that this value be rounded to 1. This makes $K-1$ equal to 2.16, which is equal to Davy's (1987) experimental value and close to the three theoretical GOE values of 2.375, 2.10, and 2.0 which were calculated above.

A more exact re-analysis of Davy's (1987) original data, taking account of Weaver's (1989a) change to the form of Eq. (1), has yielded a value of $K=3.10$ with 90% confidence limits of ± 0.34. This is in agreement with the three theoretical values of 3.375, 3.10, and 3.0. Note that the re-analysis has only reduced the experimental estimate of $K-1$ by 0.06 which is much less than the experimental uncertainty. Experimental measurements on a block by Lobkis *et al.* (2000) gave $K=2.65$. Measurements on a plate by Langley and Brown (2004b) produced $K=2.5$. Numerical calculations on two dimensional plates by Langley and Brown (2004b, 2004a) and Langley and Cotoni (2005) produced a range of values for K including 2.5, 2.52, 2.67, 2.74, 2.75, 2.86, 2.87, and 3.01.

VI. THE PURE TONE QUALIFICATION PROCEDURE

The room qualification procedure for the measurement of discrete frequency components was deleted from the draft international standard (ISO, 1996). The alternative multiple source position method was retained. Baade asked for clarification of the statement "It is now known that multiple source positions will not necessarily solve all the problems, and hence it is desirable that all reverberation rooms which are to be used for sound power measurements should pass the qualification procedure." This statement appears in Davy, 1981a, and Davy, 1981a is Appendix C of Davy (1989).

The author's analysis of the 125–1000 Hz experimental values of source position variance in Fig. 3 of Maling, 1973 produces an experimental value of $K-C$ in Eq. (5) equal to 0.68. This is much less than Davy's (1987) experimental $K-C$ estimate of 2.16 (and the re-analyzed 2.10 estimate) for the total variance case, which was obtained using the pure tone qualification procedure's frequency variation method. This shows experimentally that source position variation does not produce the total variance that exists in pure tone measurements. In turn, this suggests that the pure tone qualification procedure should have been included in ISO, 1996.

Bodlund (1977) and Jacobsen (1979) separated the total variance into a room variance and a source position variance. Using numerical procedures, Bodlund (1977) obtained $K-C$ equals 1.42 for the room variance and $K-C$ equals 2.84 for the source position variance. Using theoretical techniques, and the Poisson assumption for the room variance case, Jacobsen (1979) obtained $K-C$ equals 1 for the room variance and $K-C$ equals 2.375 for the source position variance.

Note that Jacobsen's (1979) results are summed to produce $K-C$ equals 3.375, which is the correct result for the total variance in the Poisson case, providing that tangential and axial modes are ignored. Also note that Jacobsen's (1979) equations do include the effects of tangential and axial modes, but these terms have been ignored in this analysis. Bodlund's (1977) results are summed to produce $K-C$ equals 4.26 for the total variance. Both Jacobsen's (1979) and Bodlund's (1977) results are much higher than Maling's (1973) experimental result for the source position variance. Nevertheless, they both show that the room variance is significant. Since this room variance cannot be reduced by source position averaging, these results suggest that the pure tone qualification procedure should be included in ISO, 1996. Bodlund's (1977) and Jacobsen's (1979) results have been reiterated by Tohyama *et al.* (1989) and pp. 198–199 of Tohyama *et al.*, 1995.

Setting the number of source positions N and the number of receiver positions L equal to infinity and the angular frequency difference θ to zero in Eq. (1) gives

$$\text{relvar} = \frac{1-C}{M} \tag{10}$$

for the room variance. This means that $K-C$ is equal to $1-C$ for the room variance. Thus for the Poisson distribution, $K-C$ is equal to 1 for the room variance. This agrees with Jacobsen's (1979) theoretical result. It is also the result obtained for Davy's (1981b) incorrect guess of the form of Eq. (1) for the nearest neighbor distribution. (Davy (1981b) effectively guessed that C was equal to zero and that K was replaced by $K-1/2$.) The correct result for the nearest neighbor distribution is $K-C$ equal to 1/2 for the room variance.

For the GOE distribution, $K-C$ is equal to zero for the room variance. This surprising result suggests that the multiple source position method is equivalent to the pure tone qualification procedure. However, it will soon be seen that this result is not valid at low frequencies.

If the room variance and source position variance are uncorrelated, subtracting Eq. (10) from Eq. (9) gives the source position variance of the real part of the input impedance,

$$\text{relvar} = \frac{K-1}{M}. \tag{11}$$

Ignoring tangential and axial modes, this agrees with Jacobsen's (1979) theoretical value of $K-C=K-1=23/8=2.375$ for the source position variance. It is interesting to note that this is independent of the modal frequency spacing distribution as Jacobsen (1979) showed.

Equation (1) is not valid for the nearest neighbor and GOE distributions of modal spacings at low values of the statistical modal overlap M. For low values of M, the relative covariance for these distributions tend to that for the Poisson distribution (see Fig. 1 of Weaver, 1989a, Fig. 13 of Lyon, 1969, Appendix B of Davy, 1981a, Lobkis *et al.*, 2000, and Langley and Cotoni, 2005). This trend does not have a great effect on the total variance because it is offset by the increasing percentages of tangential and axial modes as the frequency reduces and the increasing variance of decay rate at low frequencies.

However, Eqs. (10) and (11) show that the choice of distribution only affects the room variance (via C), while the percentages of tangential and axial modes only affect the source position variance (via K). Also the room variance is less than half the total variance. This means that all the increase due to low modal overlap occurs in the smaller room variance, which is not decreased by the increasing percentages of tangential and axial modes. Thus this effect is very significant for room variance. This means that the GOE distribution of modal spacings predicts significant room variance at low frequencies. This is the frequency region where the variances are most significant because they are largest. Again, since this room variance cannot be reduced by source position averaging, this result suggests that the pure tone qualification procedure should be included in ISO, 1996. The use of three or more source positions will make the source position variance less than the low frequency limit of the room variance.

VII. CONCLUSIONS

The modal frequency spacings of a reverberation room are distributed according to the Rayleigh or Wigner distribution. Since this distribution is a good approximation to the GOE spacing distribution, the use of the GOE distribution in reverberation room theories is justified. The modal frequency spacings are not distributed according to the Poisson or exponential distribution or the nearest neighbor distribution.

All the experimental, theoretical, and numerical research results suggest that the pure tone qualification procedure should be included in ISO, 1996. The value of the constant 0.79 should be increased to 1 in the equation used to calculate the number of source positions in the multiple source method in ISO, 1996. These recommended changes were made when ISO (1999) was released.

ANSI (**1980**). "ANSI S1.32-1980 Precision methods for the determination of sound power levels of discrete-frequency and narrow band noise sources in reverberation rooms," Acoustical Society of America, New York.

Balian, R., and Bloch, C. (**1970**). "Distribution of eigenfrequencies for wave equation in a finite domain: 1. 3-dimensional problem with smooth boundary surface," Ann. Phys. (N.Y.) **60**, 401–447.

Bodlund, K. (**1977**). "A normal mode analysis of sound power injection in reverberation chambers at low-frequencies and effects of some response averaging methods," J. Sound Vib. **55**, 563–590.

Burkhardt, J., and Weaver, R. L. (**1996**). "The effect of decay rate variability on statistical response predictions in acoustic systems," J. Sound Vib. **196**, 147–164.

Davy, J. L. (**1981a**). "The qualification of a reverberation room for pure-tone sound power measurements," in *Acoustics and Society, Proceedings of the 1981 Annual Conference of the Australian Acoustical Society*, edited by D. A. Gray (Australian Acoustical Society, Cowes, Victoria, Australia), pp. 3C5:1–3C5:5.

Davy, J. L. (**1981b**). "The relative variance of the transmission function of a reverberation room," J. Sound Vib. **77**, 455–479.

Davy, J. L. (**1987**). "Improvements to formulas for the ensemble relative variance of random noise in a reverberation room," J. Sound Vib. **115**, 145–161.

Davy, J. L. (**1989**). Research Proposal for ASHRAE Research Project No. 624-TRP, CSIRO Division of Building, Construction and Engineering, Melbourne.

Gurevich, I. I., and Pevsner, M. I. (**1957**). "Repulsion of nuclear levels," Nucl. Phys. **2**, 575–581.

ISO (**1988**). "ISO 3742:1988(E) Acoustics—Determination of sound power levels of noise sources—Precision methods for discrete frequency and narrow-band sources in reverberation rooms," International Organization for Standardization, Geneva, Switzerland.

ISO (**1996**). "ISO/DIS 3741:1996 Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for reverberation rooms," International Organization for Standardization, Geneva, Switzerland.

ISO (**1999**). "ISO 3741:1999(E) Acoustics—Determination of sound power levels of noise sources using sound pressure—Precision methods for reverberation rooms," International Organization for Standardization, Geneva Switzerland.

Jacobsen, F. (**1979**). *Sound Power Determination in Reverberation Rooms—A Normal Mode Analysis* (The Acoustics Laboratory, Technical University of Denmark, Copenhagen).

Kendall, M. G., and Stuart, A. (**1967**). *The Advanced Theory of Statistics* (Charles Griffin & Co. Ltd., London).

Langley, R. S., and Brown, A. W. M. (**2004a**). "The ensemble statistics of the band-averaged energy of a random system," J. Sound Vib. **275**, 847–857.

Langley, R. S., and Brown, A. W. M. (**2004b**). "The ensemble statistics of the energy of a random system subjected to harmonic excitation," J. Sound Vib. **275**, 823–846.

Langley, R. S., and Cotoni, V. (**2005**). "The ensemble statistics of the vibrational energy density of a random system subjected to single point harmonic excitation," J. Acoust. Soc. Am. **118**, 3064–3076.

Legrand, O., and Mortessagne, F. (**1996**). "On spectral correlations in chaotic reverberation rooms," Acust. Acta Acust. **82**, S150.

Legrand, O., Mortessagne, F., and Sornette, D. (**1995**). "Spectral rigidity in the large modal overlap regime—Beyond the Ericson–Schroeder hypothesis," J. Phys. I **5**, 1003–1010.

Lobkis, O. I., Weaver, R. L., and Rozhkov, I. (**2000**). "Power variances and decay curvature in a reverberant system," J. Sound Vib. **237**, 281–302.

Lubman, D. (**1974**). "Review of reverberant sound power measurement standard and recommendations for further research," National Bureau of Standards Technical Note 841.

Lyon, R. H. (**1969**). "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 545–565.

Maling, G. C. (**1973**). "Guidelines for determination of the average sound power radiated by discrete-frequency sources in a reverberation room," J. Acoust. Soc. Am. **53**, 1064–1069.

Schroeder, M. R. (**1987a**). "Normal frequency and excitation statistics in rooms—Model experiments with electric waves," J. Audio Eng. Soc. **35**, 307–316.

Schroeder, M. R. (**1987b**). "Statistical parameters of the frequency response curves of large rooms," J. Audio Eng. Soc. **35**, 299–306.

Tanner, G., and Sondergaard, N. (**2007**). "Wave chaos in acoustics and elasticity," J. Phys. A **40**, R443–R509.

Tohyama, M., Imai, A., and Tachibana, H. (**1989**). "The relative variance in sound power measurements using reverberation rooms," J. Sound Vib. **128**, 57–69.
Tohyama, M., Suzuki, H., and Ando, Y. (**1995**). *The Nature and Technology of Acoustic Space* (Academic, London).
Waterhouse, R. V. (**1978**). "Estimation of monopole power radiated in a reverberation chamber," J. Acoust. Soc. Am. **64**, 1443–1446.
Weaver, R. L. (**1989a**). "On the ensemble variance of reverberation room transmission functions, the effect of spectral rigidity," J. Sound Vib. **130**, 487–491.
Weaver, R. L. (**1989b**). "Spectral statistics in elastodynamics," J. Acoust. Soc. Am. **85**, 1005–1013.

Acoustic simulations of Mudejar-Gothic churches

Miguel Galindo,[a] Teófilo Zamarreño, and Sara Girón
Departamento de Física Aplicada II, Universidad de Sevilla, ETS de Arquitectura IUACC, Avenida Reina Mercedes 2, 41012-Sevilla, Spain

(Received 13 January 2009; revised 22 June 2009; accepted 24 June 2009)

In this paper, an iterative process is used in order to estimate the values of absorption coefficients of those materials of which little is known in the literature, so that an acoustic simulation can be carried out in Mudejar-Gothic churches. The estimation of the scattering coefficients, which is even less developed, is based on the size of the irregularities. This methodology implemented is applied to six Mudejar-Gothic churches of Seville (southern Spain). The simulated monophonic acoustic parameters, both in the frequency domain and as a function of source-receiver distance (spatial distribution), are analyzed and compared with the *in situ* measures. Good agreement has been found between these sets of values, whereby each parameter is discussed in terms of the just noticeable difference. This procedure for existing buildings, especially for those which are rich in heritage, enables a reliable evaluation of the effect on the maintenance, restoration, and conditioning for new uses, as well as the recreation of the acoustic environment of ancient times. Along these lines, the acoustic influence of the timber roof and the presence of the public in these churches have also been studied. [DOI: 10.1121/1.3180632]

PACS number(s): 43.55.Gx, 43.55.Ka [NX] Pages: 1207–1218

I. INTRODUCTION

Computational models[1] based on geometric acoustics first appeared around 1967, and since 1990 have been maturing toward attaining results closer to the true acoustic conditions of a closed enclosure. Thanks to the improvement in the calculation speed in parallel with the technological equipment features and to the appearance of new analysis methods, these techniques currently provide highly reliable results despite the limitations associated with geometric models. Among the latest improvements, the most recent algorithms can produce auralizations of high quality.[2]

The computational base is the ray-tracing technique. It is an appropriate method of analysis in the high frequency range (four times above Schröeder's frequency) but with limitations in the bass frequency region. However, the results hitherto obtained have not been completely satisfactory[2] due to this loss of the ondulatory character of sound. One way to recover some of the effects associated with this wave nature is the introduction of the scattering coefficient which is assigned to each surface of the enclosure. In this way the reflections that take place can be modified from purely specular behavior to relatively diffuse behavior.

It has been sufficiently demonstrated that all surfaces require the incorporation of an appropriate scattering coefficient[3,4] to achieve close simulation to the real acoustic conditions of an enclosure. The great difficulty arises when there is a lack of empirical data for the scattering coefficients of the different materials, although some research is along these lines,[2,5–8] whereby measurements are studied both in the laboratory and *in situ*. Moreover, an ISO (Ref. 9) normalized procedure for the measurement of scattering has also been established.

Likewise, it should be pointed out that a very detailed geometric model does not necessarily lead to greater precision in the results of the simulated acoustic parameters.[4] As a rule of thumb,[2] details must be greater than 0.5 m. It is much more important to describe the acoustic properties of the surfaces (absorption and scattering coefficients), for which data with a suitable approximation are not readily available.

Various algorithms have been developed to implement the propagation of the acoustic energy inside an enclosure by using ray-tracing techniques.

The algorithm of ray tracing was the first model used in auditorium[10] designs. It is based on the trace and pursuit of the sound rays from a point of the enclosure that acts as the source, as far as the reception point, by following the laws of geometric optics and which takes up to a certain order of reflections into account. This model enables the effects of the dispersion to be incorporated. The greatest drawback of this technique arises from the high number of rays necessary to cover a typical enclosure, which causes an increment in the calculation time when evaluating the path traveled by each ray for each octave band. The method has since been further developed[11] and the path of each ray has become a circular cone[12] or pyramidal with a triangular base.[13]

The algorithm of the image sources is used to generate an echogram by bearing in mind the intensity associated with each reflection and the time of arrival in relation to the direct sound. When arbitrary surfaces exist then the number of possible images increases exponentially with the reflection order, resulting in a complicated model, as occurs in concert halls.

The great advantage of this algorithm is its deterministic character when building the impulsive response. On the other hand, its great inconvenience is the enormous computational

[a]Author to whom correspondence should be addressed. Electronic mail: mgalindo@us.es

TABLE I. Absorption (α *above*) and scattering (δ *below*) coefficients of the main materials used in the simulations.

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Wooden framework (WF)	0.30	0.29	0.24	0.21	0.20	0.21
	0.50	0.60	0.70	0.70	0.80	0.80
Marble and ceramic tiles (MR)	0.01	0.01	0.01	0.02	0.02	0.01
	0.10	0.10	0.10	1.10	0.10	0.10
Pews (PW)	0.16	0.18	0.18	0.17	0.17	0.16
	0.30	0.40	0.50	0.60	0.70	0.80
Wood (WO)	0.30	0.25	0.20	0.17	0.15	0.10
	0.20	0.20	0.20	0.20	0.20	0.20
Altarpiece (AL)	0.32	0.28	0.27	0.27	0.20	0.17
	0.30	0.40	0.50	0.60	0.70	0.80
Visible brick (VB)	0.02	0.03	0.03	0.04	0.05	0.05
	0.10	0.10	0.10	0.10	0.10	0.10
Velvet (VE)	0.02	0.04	0.08	0.20	0.35	0.40
	0.10	0.10	0.10	0.10	0.10	0.10
Ceramic flooring block (CB)	0.09	0.09	0.08	0.07	0.07	0.08
	0.10	0.15	0.15	0.20	0.20	0.20
MN Walls (visible brick with thick mortar) (MNW)[a]	0.11	0.11	0.12	0.11	0.11	0.11
	0.10	0.15	0.15	0.20	0.20	0.20
JU Walls (plastered and painted brick with some oil paintings) (JUW)[a]	0.10	0.10	0.11	0.12	0.12	0.11
	0.10	0.15	0.15	0.20	0.20	0.20
PE Walls (plastered and painted brick with many oil paintings altars and wood sculptures) (PEW)[a]	0.18	0.16	0.14	0.13	0.15	0.16
	0.30	0.40	0.50	0.60	0.70	0.80
ES Walls (visible brick with thick mortar and many oil paintings) (ESW)[a]	0.14	0.12	0.13	0.15	0.19	0.21
	0.20	0.20	0.30	0.30	0.40	0.50
MC Walls (painted brick) (MCW)[a]	0.04	0.04	0.06	0.06	0.07	0.07
	0.10	0.10	0.10	0.10	0.10	0.10
CA Walls (plastered and painted brick with some oil paintings) (CAW)[a]	0.13	0.15	0.16	0.13	0.11	0.11
	0.10	0.15	0.15	0.20	0.20	0.20

[a]Absorption coefficients estimated by the iteration process. The other absorption coefficients from Ref. 2, 7, and 28 and data from internal reports. All scattering coefficients estimated by the size of the irregularities of the surfaces (Ref. 3).

cost. Hence, it is usually only used to build the first part of the impulse response (IR), thereby obtaining the first reflections accurately, which is of maximum importance to determine the main perceptive characteristics of an acoustic field.

In 1989 Vorländer[14] presented the first hybrid algorithm in an attempt to combine the advantages of previous algorithms and to limit the incidence of their drawbacks. It consists of finding images that have a high probability of being valid by means of the layout of rays from the source and of taking into account those surfaces which have valid scores.

The finite number of rays used provides an echogram with limited length and with energy deficiencies since the number of traced rays is finite and not all the image sources are considered. To this end, it is necessary to use other methods (usually statistical methods) to gather the rest of the contributions, thereby allowing the integrated IR tail of the enclosure to be generated.

This technique has been applied by some authors in concert halls,[15] theaters, some of them with a similar typology,[16] in mosques,[17] and also in the study of ancient Greek and Roman open-air theaters.[18]

II. MUDEJAR-GOTHIC CHURCHES

Twelve churches of the same typology but varying in volume, dimensions, interior finishing, and furnishing have been acoustically examined.[19] All these churches were built

in the Middle Ages whose architectural style is a result of a unique Spanish artistic movement since it was influenced by both Islamic and Christian Gothic elements.

The Mudejar-Gothic churches in Seville, all located in its historical center, are morphologically characterized by this stylistic dualism: a vaulted Gothic apse and a body of three naves with a timber roof (collar beam in the main nave) of Moorish origin. Their brick walls are complemented with portals and a stone apse. The supports are also clearly Islamic, with quadrangular or sometimes octagonal pillars and with raised brick mouldings as decoration. Pointed, round, or segmental arches rest on these supports. A complete description of the furnishings and other acoustic information on these temples has been previously published.[19]

Six churches have been chosen, from the total of 12 acoustically studied, for the acoustic simulations. In order of decreasing volume they correspond to Santa Marina (MN), San Julián (JU), San Pedro (PE), San Esteban (ES), San Marcos (MC), and Santa Catalina (CA) churches, respectively. Several reasons are put forward for the choice of these six churches. The first is that they belong to two different groups as regards the values of their reverberation times: Santa Marina church and San Marcos church are included in the first group, with the longest reverberation times of the sample and longer than optimal tonal curves, proposed by Knudsen *et al.*[20] The remaining churches, on the other hand, present short reverberation times which give the best conditions both for speech and for musical use.

The second reason is to cover the whole range of volumes. In this way the biggest volume of the group (10 708 m^3) corresponds to Santa Marina church, and intermediate volumes are for San Julián and San Pedro (6226 and 6180 m^3), respectively. Those of San Esteban, San Marcos, and Santa Catalina (4746, 4623 and 4362 m^3, respectively) are located among the smallest churches of the sample. Moreover, San Marcos church, due to the treatment of its walls and the lack of a wooden roof in its central nave, is found within the group of churches with a long reverberation time but with one of the smallest volumes.

The third reason has to do with the surface finishing. In this way, two churches present scarcely decorated walls, while the remaining churches are considerably more embellished. Santa Marina and San Marcos are the two with scarce decoration: practically limited to the wooden pews located in the congregational seating area (central nave). In the remaining churches, the altarpieces of the wooden main altar, other paintings, sculptures, and textile surfaces are present.

In Santa Marina and San Esteban churches, the walls are made of visible brick with thick mortar, while in the other four churches the walls of brick have been plastered and painted or directly painted. In San Julián and San Pedro churches, their perimetral walls and pillar bases are finished with a ceramic baseboard around 2 m high.

In San Marcos church, the wooden ceiling in the central nave was destroyed by a fire, and rough ceramics substitute the original wooden board. In San Pedro church the wooden boards of the lateral nave ceilings have been substituted by tiles. These details are presented in Table I and in Figs. 1 and 2 and are described in Sec. IV A.

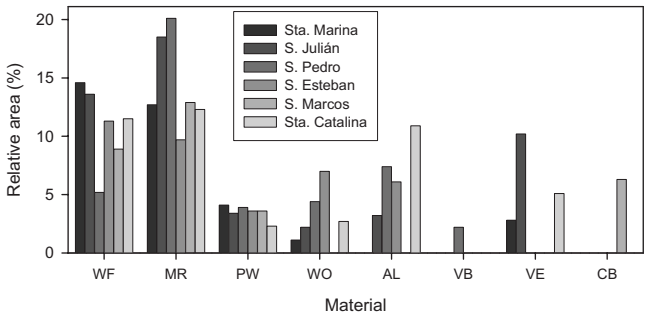


FIG. 1. Relative area, as a percentage of total inner area, for the main materials used in the acoustic models of the churches (except surfaces adjusted by the iteration process). The categories of the horizontal axis correspond to the acronym of the materials specified in Table I.

III. MEASUREMENT TECHNIQUE

The procedures employed were those established in ISO 3382 (Ref. 21) and IEC 60268-16 (Ref. 22) standards, and all measures were carried out in unoccupied churches. Air temperature, relative humidity, and atmospheric pressure were monitored during the measurements. The range of variation in all churches was 22.6–27.4 °C for the temperature, 35.7%–65.7% for the relative humidity, and 101.7–102.5 kPa for the atmospheric pressure.

Monophonic IRs and other room responses to stationary signals were measured to determine the following parameters, among others, for each frequency band between 125 and 4000 Hz in all receiver positions: reverberation time (T); sound strength (G); center time (T_s); clarity (C_{80}) and definition (D_{50}), as energy-based parameters related to the early-to-late or early-to-total sound energy ratio; early lateral energy fraction (LF) to study the spatial impression phenomena in these places; and finally, the RASTI index to evaluate the intelligibility from the degradation of the modulation transfer function.

The IR has been obtained using maximum length sequence (MLS) signals[23] generated and analyzed by the analyzer MLSSA. The omnidirectional source (B&K 4296) was placed at the most usual point of location of the natural source: the altar at a height of 1.70 m from the floor. The microphone (omnidirectional B&K 4190 $\frac{1}{2}$ in. or multipat-

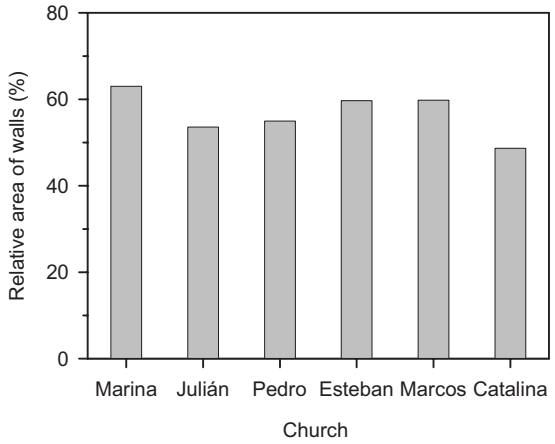


FIG. 2. Relative area, as a percentage of total inner area, for the surfaces adjusted by the iteration process in each church.

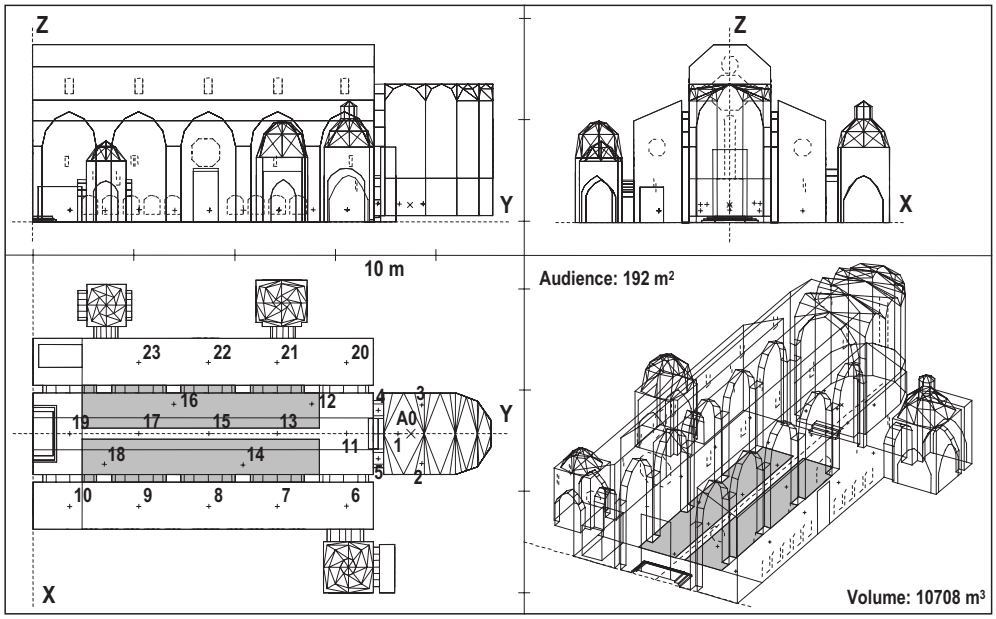


FIG. 3. Longitudinal and cross sections (above) and ground plan and 3D model (below) created for the computer simulation of Santa Marina church.

tern Audio-Technica AT4050/CM5) was located at the approximate height of the head of a seated person ~1.20 m, in a predetermined number of positions distributed in the central nave and the lateral naves, ranging from 12 reception points for Santa Catalina church to 23 for Santa Marina (see Fig. 3).

The calculation of sound strength uses the emission of a stationary signal, related to the level produced by the source, at 10 m distance under free-field conditions. The equivalence of these results with those obtained by MLS signals has also been published elsewhere.[19]

The background noise measurements are carried out by following the same procedure as for the sound strength but by substituting the calibrated source with the background noise in each case. As for measurements in the frequency domain, an average of more than 400 spectra are taken using MLSSA, over a period of approximately 6 min. These measures are carried out at one/several representative reception points (avoiding the vicinity of the entrance or other positions near potential noise disturbance sources).

The experimental RASTI values used in this work have been obtained using the B&K 3361 system, which uses the modulated stationary noise technique. The equivalence of these results with those obtained through the MLS spectrum has also been proved in previously published work.[19]

To characterize the spatial distribution of the aforementioned acoustic parameters, each parameter has been averaged spectrally as follows.

- Clarity, as a direct average of the values at frequencies of 500, 1000, and 2000 Hz.
- Definition, as a weighted average as Marshall[24] proposed for C_{50}:

$$D_{50\mathrm{av}} = 0.15D_{50}(500\ \mathrm{Hz}) + 0.25D_{50}(1\ \mathrm{kHz}) + 0.35D_{50}(2\ \mathrm{kHz}) + 0.25D_{50}(4\ \mathrm{kHz}). \quad (1)$$

- Reverberation time, center time, and sound strength, as a mean of the corresponding values at mid-frequency 500 and 1000 Hz bands.[25]
- Lateral energy fraction, as a direct average of 125, 250, 500, and 1000 Hz octave bands.[26]

IV. RESULTS AND DISCUSSION

A. General computational aspects

The software used is CATT-ACOUSTIC. The full detailed calculation makes use of the randomized tail-corrected cone-tracing (RTC) algorithm with statistical corrections of the tail which combines features of specular cone tracing, standard ray tracing, and the image source algorithm. This prediction method enables the calculation of numeric values for room acoustic parameters and the echogram production which can be used in the auralization processes. To mitigate the inconveniences of the method, the direct sound, the first order specular and diffuse reflections, and the specular reflections of second order are handled in a deterministic way by the image source method.

The simulations undertaken use a calibration process that is not possible to implement for new buildings since it is based on an adjustment of the values of absorption coefficients of those materials that do not appear in the literature and whose measurement in the laboratory or *in situ* is excessively complex. This process is carried out by means of an iterative procedure whose final objective is that the spatially averaged simulated reverberation times do not differ more than 5% from those measured *in situ* in each church. The limit for these differences is based on the subjective limen, the just noticeable difference (JND), for which a value of 5% is widely accepted.[4,26] Since CATT-ACOUSTIC offers the possibility of evaluating the values of these reverberation times in an interactive way starting from classic formulas, the adjustment procedure does not necessarily require a complete

TABLE II. Absolute difference between measured and simulated distance (m) values, and the standard deviation.

Church	Absolute difference	Standard deviation
Santa Marina	0.13	0.16
San Julián	0.08	0.07
San Pedro	0.09	0.07
San Esteban	0.10	0.08
San Marcos	0.11	0.08
Santa Catalina	0.21	0.14

TABLE III. Parameters used in the simulations. In bold those calculated with the new algorithm.

Church	Number of rays	Truncation time t_r (s)
Santa Marina	48 412	2.93
San Julián	**72 122**	2.68
San Pedro	**56 702**	2.10
San Esteban	30 590	1.95
San Marcos	37 664	3.71
Santa Catalina	**51 440**	1.63

simulation in the first stages. Nevertheless, due to the peculiarity of this type of building, the classic formulas (Eyring for these spaces with coupled rooms and non-uniformly distributed absorption) cannot achieve accurate predictions, and hence, for the final adjustment of the coefficients, a full simulation must be launched.

Although this calibration process has the drawback that a previous measure is necessary, in the case of a new building the acoustic test for the materials implemented may be known before their installation. On the other hand, in the case of an existing enclosure, especially if listed as cultural heritage, this process allows the performance in the processes of maintenance, restoration, or conditioning to new uses, to be evaluated in a very reliable way, and also facilitates the recreation of the acoustic environment of past eras.

In all cases, atmospheric pressure conditions, air temperature, and relative humidity were monitored *in situ* during the experimental measurements and were used in the simulation process. These physical variables exercise influence on the determination of the speed of sound (and therefore, on the relative delays of the different reflections when building echograms) and on the estimation of the sound absorption by the interior air of the churches. The background noise spectrum was also adjusted to make it coincide with the values measured *in situ*: this incidence is especially important when estimating the values of the STI index. The sound source has been chosen by adjusting its emission levels to make it coincide with that used in the experimental measures, with the main purpose of comparing the measured and simulated values of pressure level when the source emits in a stationary way. No special modifications have been carried out to simulate the deviations from the omnidirectional pattern of the real acoustic source, and hence an omnidirectional-pattern directivity source has been used in the simulations.

Since one of the possible error sources when comparing experimental and simulated data is the variation in the location of receivers between the real building and the implemented pattern, meticulous care has been exercised to prevent this variation. To this end, Table II presents the absolute differences between measured (obtained from the flying time of direct sound from the source to the receiver and the sound speed) and simulated distance values and their standard deviations for all churches. These differences are in the range from 0.08 m for San Julian church to 0.21 m for Santa Catalina church and, therefore, in all cases, these differences are less than the distance that separates contiguous seats.

In order not to force the simulation parameters for the different models, the estimations of the program, for both the number of rays and for the truncation time, were accepted by means of selecting the corresponding options ("autonumber" and "autotime") when forming software for the full calculation process, on the condition that the value of truncation time fulfilled the requirement of being of the order of the reverberation time. In Table III, the values of these parameters for the six simulated churches are shown. In CATT-ACOUSTIC v8h a new fully detailed calculation algorithm option for the late part of the IR is implemented.[27] The new option differs only in the late part of the echogram where randomized ray tracing is used instead of RTC. The benefit is that no tail correction is necessary, and hence no assumptions are required about reflection density growth, and therefore coupled rooms and other unusual shapes can be better predicted. The penalty is a higher random error in the late part and hence more rays have to be used (the automatically generated ray number is set twice as high as for the RTC). However, the new algorithm variant is slightly faster, resulting in a total calculation time of approximately the same as for RTC. This option was used in the three churches marked in bold in Table III.

All the surfaces of the churches for the simulations have been characterized with their corresponding scattering coefficient, at the different frequencies, estimated from the size of their irregularities (see Table I). The default value of the program has been adjusted to 0.1.

B. Results of the simulations

As a sort of sample, Fig. 3 exhibits the longitudinal and cross sections, the ground plan at the audience level and a view in perspective of the three-dimensional (3D) model produced for Santa Marina church. In the ground plan, the location of receivers, the source position next to the major altar, which is in accordance with the usual conditions of the liturgy, and the area for the congregational listeners (in gray), are shown. The area for the audience, total volume, and a graphic scale are given to give an idea of the dimensions of the church.

In the same way, the main materials with their respective absorption and scattering coefficients used in the computer models at the different octave bands are presented in Table I. In Figs. 1 and 2, the percentage area of each material in each church is presented. The majority of the interior materials (shown in Fig. 1) have absorption coefficients that are well known in the literature [2,7,28] [wooden framework (WF) and altarpiece from internal reports], and hence the estimation

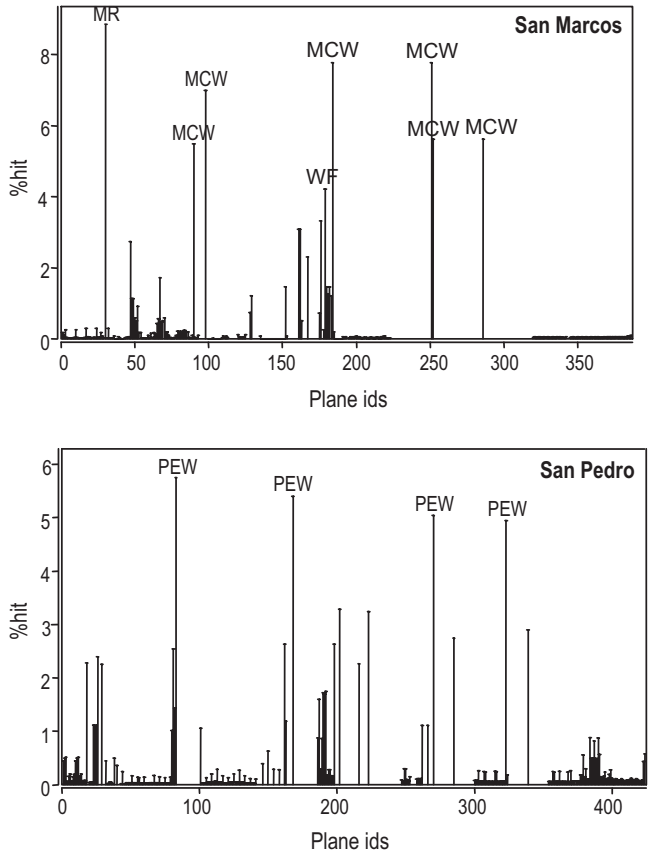


FIG. 4. Histograms of reflections produced on the surfaces (identified by their number in the model): San Marcos church (above) and San Pedro church (below). The main surfaces (those with a score higher than 4%) are identified in each case by the acronyms assigned in Table I.

from the iterative process is made for the vertical walls (shown in Fig. 2) of the three naves of the churches (excluding the ceramic baseboard present throughout the whole church in San Julián and San Pedro churches). The studied churches present different wall finishings, number of paintings, altars, and decorations, thereby establishing the differences in the acoustic coefficients and their behavior versus frequency. The scattering coefficient of the surfaces, which is even less developed, is estimated by the size of the irregularities. The surface shape is described by the average structural depth and the average structural length.[2] For a better understanding of the obtained coefficients, Table I gives a rough description of these walls.

The program enables the relative importance of the different surfaces to be evaluated by means of an interactive histogram where the impacts on each surface are presented. Figure 4 shows the hit score for the planes of San Marcos and San Pedro churches in terms of their identifiers in the model. The planes with a score higher than 4% are identified by the acronyms assigned in Table I. The most important contributions in all cases (adding the scores from the different planes to those of the same acoustic finishing) correspond to the wooden ceiling, the floor, and the longitudinal lateral walls of the three naves.

In relation to the comparison of the results of the simulated acoustic descriptors and those measured *in situ*, Fig. 5 presents the reverberation times as a function of frequency in

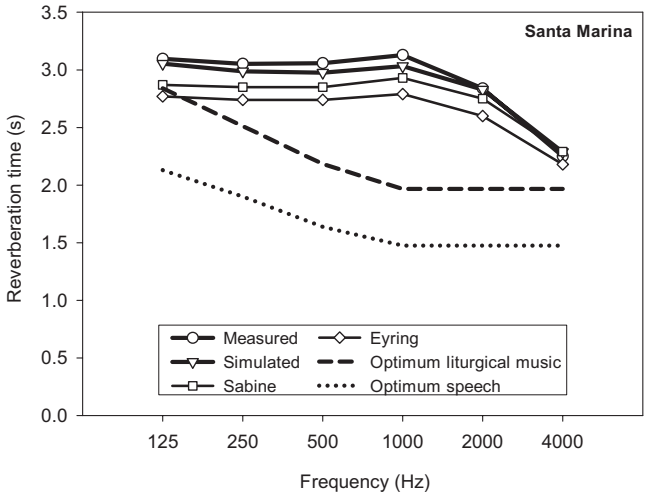


FIG. 5. Reverberation times in octave bands: measured, simulated, and calculated for Santa Marina church. Optimum values are shown for comparison purposes.

octave bands for Santa Marina church. The Sabine and Eyring reverberation times have been included, as well as the optimal tonal curves for music and for speech uses according to Knudsen *et al.*[20] The Sabine reverberation time has been estimated from surface data with their corresponding absorption coefficients. The Eyring reverberation time is estimated from the mean free path calculated from all ray segments, and the mean absorption coefficient from the arithmetic mean of all absorption values encountered by the rays.

In all cases, the simulated values of reverberation times differ by less than 5% to those measured experimentally, due to the adjustment of the absorption coefficients of those materials that offer greater uncertainty: fundamentally the vertical walls of the three naves of each church.

The values of the objective parameters related to speech intelligibility are deduced from the echogram obtained during the simulation process by previously deriving the modulation transfer function to later obtain the values of the RASTI index. A great similarity of the simulated values to the experimental data is observed, in spite of the fact that these simulated values are obtained from the modulation transfer functions deduced from the echograms, while the

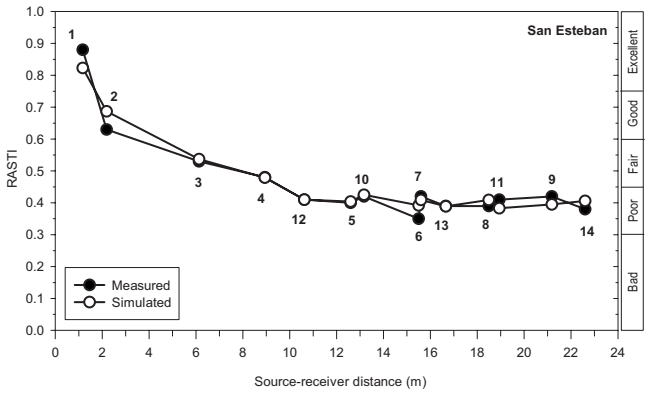


FIG. 6. Comparison of measured (black dots) and simulated (white dots) of RASTI values for San Esteban church at various locations. Receivers 1 and 2 in the presbytery, 3–9 in the central nave, and 10–14 in the lateral naves.

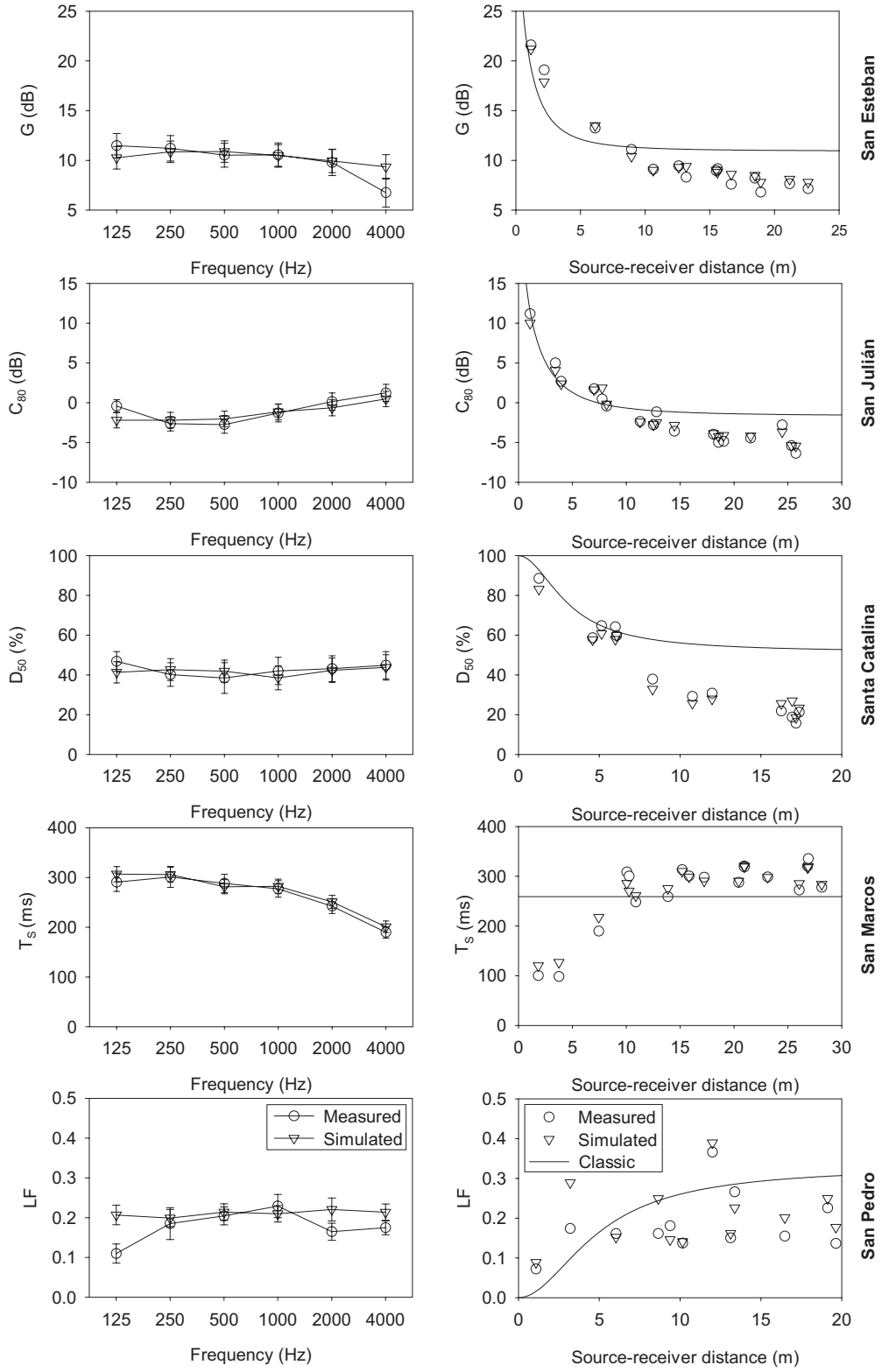


FIG. 7. Frequency (left column) and source-receiver distance dependence (right column) for sound strength, clarity, definition, center time, and lateral energy fraction. Comparison of measured (○) and simulated (▽) values in five churches.

experimental measures have been obtained using stationary signals of modulated noise. In these worship spaces, the background noise level is insignificant and dominates the reverberant sound field, and therefore a good adjustment of the absorption coefficients has played a crucial role in obtaining simulated values similar to the measured values.

In this way, Fig. 6 compares the simulated results with the experimental measures in San Esteban church for the RASTI index taken at the different reception points in the church. The experimental results correspond to the measurements carried out with the B&K equipment, without electroacoustic support and with an adjusted emission level ref

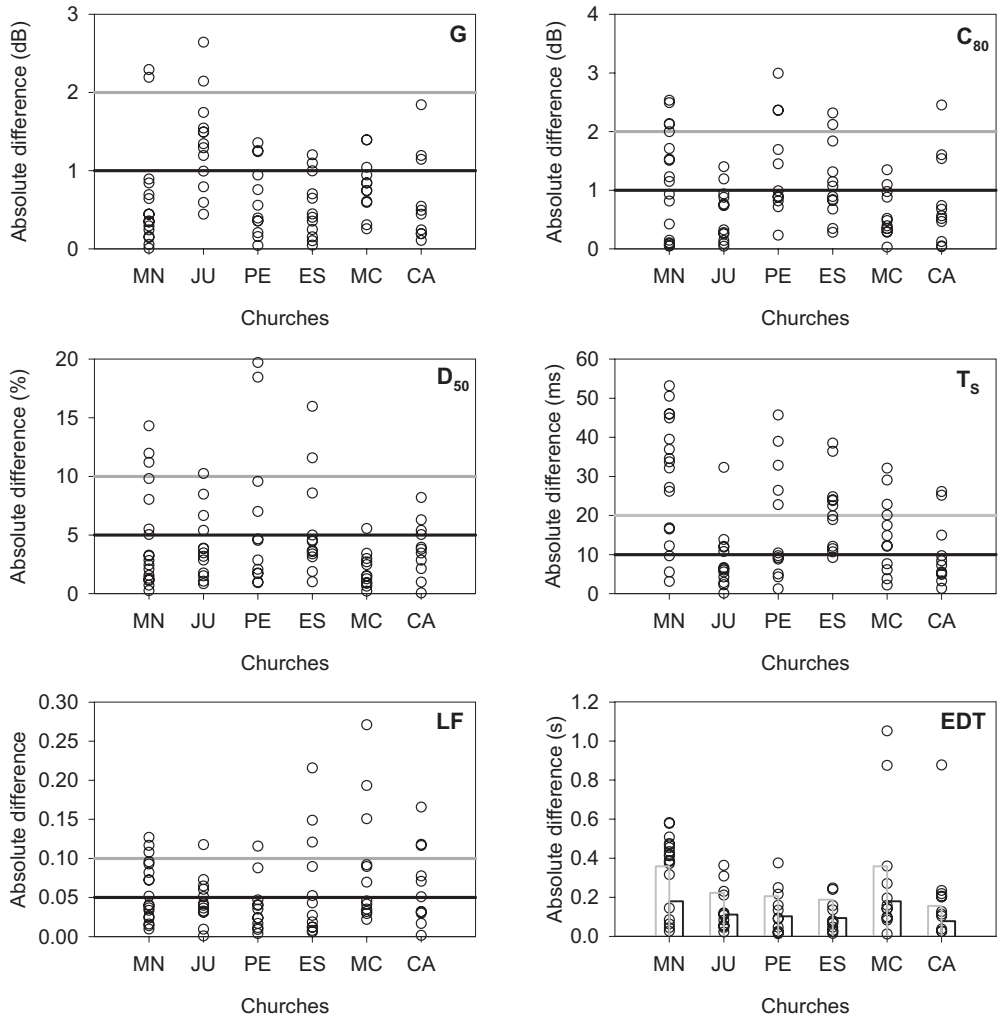


FIG. 8. Absolute difference between spectrally averaged measured and simulated values for all reception points in each church and for each acoustic parameter (circles). Two limits as a function of JND are indicated: black line for JND and gray line for 2 JNDs.

+10 dB, which supposes 69 dB in the 500 Hz band and 60 dB in the 2 kHz band at 1 m from the loudspeaker. On the right-hand-side vertical axis of the figure, the intervals of qualitative qualification for the intelligibility of speech appear. The coincidence is clearly more than acceptable in all the areas of the ecclesiastic building.

Continuing with the comparison between simulated and measured results, Fig. 7 exhibits the results of the following monophonic acoustic parameters: sound strength, clarity, definition, center time, and lateral energy fraction for five selected churches. In all cases, the behavior versus frequency and the spatial distribution of the spectrally averaged values as a function of source-receiver distance are studied. As regards the behavior versus frequency, the spatial dispersion is shown for each band by a vertical bar whose length is the size of its standard error (or mean quadratic error). In Fig. 7, all the plotted graphs use a common nomenclature. It is worth pointing out the acceptable coincidence, both for the spectral dependence and the spatial dependence, between the simulated results and the results from measurements *in situ*.

As a reference, in the dependence of these parameters as a function of source-receiver distance, the curve predicted by the classic theory has been superimposed, whereby perfectly diffuse behavior of the acoustic field of the churches is supposed and the church volumes and measured reverberation times are considered in the calculations. The churches clearly deviate from a pure exponential decay, and hence all measured and simulated parameters move away from classic predictions. It is a common observation in churches that the extinction decay traces show complex behavior circumscribed to the very early reflection pattern. For the type of churches under study, a suitable analytical model of energy distribution[29] was proposed which is capable of predicting the measured energy-based acoustic parameter versus source-receiver distance. The model proposed earlier by Cirillo *et al.*[30] for Italian churches was also discussed for the Mudejar-Gothic churches in that work, and recent theoretical research in multiple-rate decay using powerful estimation methods[31] has inspired Martellotta[32] to a refinement of the previous linear model[30] for churches by avoiding complexity in calculations. A more detailed explanation of sound propagation in these complex spaces requires in-depth analysis and more experimental data on other types of worship places, and constitutes part of some research in progress.

As for the inspection of simulation results, for all energy-based acoustic parameters, the lateral energy fraction is the most deficiently simulated acoustic parameter when considering its dependence on frequencies, since this param-

TABLE IV. Subjective difference limen JND for each parameter.

Parameter	JND
G	1 dB
C_{80}	1 dB
D_{50}	5%
T_s	10 ms
LF	0.05
EDT	5%

eter is the most sensitive to the relative source-receiver position. The mean value for each frequency differs by around 0.05 (1 JND) and occasionally for the 125 Hz octave band does not go beyond 0.1 (2 JNDs). In all churches, simulated mean values are almost constant for all frequencies and the measured values increase between 125 Hz and 1 kHz and decrease between 1 and 4 kHz, as Fig. 7 shows for San Pedro (PE) church. According to the softened directional echograms that the program presents along the three perpendicular directions (upwards-downwards, forwards-backwards, and left-right), and focusing on this last direction, the biggest discrepancies appear in around the first 80 ms.[33] In all cases the behavior for low frequencies and especially that in the 125 Hz octave band presents average values that widely differ from the experimental values. Better results are obtained starting from 500 to 4000 Hz, especially in the case of San Marcos church.

In the analysis of the distance dependence, the experimental LF spectrally averaged values are higher at the reception points of the lateral naves than those situated in the central naves.[33] These results are not confirmed by the simulations, perhaps due to limitations of the ray-tracing technique which creates acoustic shadows for the columns and pillars. This creation is possible to check through the audience area mapping calculation option of the software.

To complete this information, Fig. 8 shows the absolute difference between spectrally averaged (see Sec. III) measured and simulated values for all the reception points in each church and for each acoustic parameter. Since MLSSA software calculates the source-receiver distance from the arrival of the direct sound, the receptors that prevent this arrival are omitted. Bork[4] proposed a standard way for the comparison of the measured and simulated values through the subjective difference limen of the acoustical parameters, which are shown in Table IV. Accuracy between the two sets of values can be in the range of 1–2 times the JNDs.[2,4]

It can be seen from Fig. 8 that the accuracy range includes the majority of the reception points for all the aforementioned acoustic descriptors, including the EDT parameter, thereby supporting the goodness of the simulations. To quantify the results, the percentage of these differences in the interval of 1 and 2 JNDs is shown in Table V. There are isolated exceptions in some churches, such as in Santa Marina (MN) church, which has about 40% of its points out of the range of 2 JNDs for the two highly correlated parameters, T_S and EDT.

TABLE V. Percentage of the absolute difference between spectrally measured and simulated values for all the reception points in each church and for each acoustic parameter within one and two JND.

Range	G	C_{80}	D_{50}	T_s	LF	EDT
Within JND (%)	75.5	64.1	73.1	39.7	62.8	47.4
Within 2 JNDs (%)	94.9	87.2	89.7	62.8	83.3	64.1

Special attention must be paid to the lateral energy fraction. Due to its definition and the systematic error associated with the measurement (orientation and sensitivity of the multipattern microphone),[34] one would expect LF to have the greatest discrepancies. This is not true for the spectrally averaged LF in all these churches, especially in the case of San Pedro (PE) church, whose differences are all, except at one reception point, within the range of the JND. Since the vertical walls of the three naves of San Pedro (PE) church display profuse and uniformly distributed decoration (see Table I), these peculiarities were not modeled and were included as wall characteristics in its absorption and scattering coefficients. This seems to be a weak point but the results demonstrate otherwise. Obviously this argument fails when the analysis is made using the octave frequency bands at low frequencies, particularly at 125 Hz.

In order to give point-by-point precision of the simulation for each church, the first column for each parameter in Table VI displays the spatially averaged value of the absolute differences, calculated at each receiver between the spectrally averaged measured results and simulated results. The column on the right-hand side presents the standard error of these differences for each parameter and for each church. The spectral average of the different parameters is as specified in Sec. III. For the calculation of these mean values, reception points at shorter than the minimum distance[21] $d_{\min}=2\sqrt{V/(cT)}$ have been omitted. This table shows the good agreement between measured and simulated values for these worship spaces. For all parameters the mean values of the differences are about 1 JND for all churches. The values

TABLE VI. Spatially averaged values of the absolute differences between spectrally averaged measured and simulated values in each receiver (left column) and their standard error (right column) for each parameter and church.

Parameter	MN		JU		PE		ES		MC		CA	
G (dB)	0.51	0.09	1.33	0.16	0.75	0.13	0.51	0.11	0.95	0.08	0.77	0.11
C_{80} (dB)	1.23	0.20	0.55	0.13	1.78	0.45	1.12	0.19	0.89	0.16	0.71	0.20
D_{50} (%)	4.17	0.82	4.22	1.01	9.08	3.08	4.79	1.30	1.94	0.44	3.85	0.80
T_s (ms)	26.30	3.09	26.37	2.91	32.77	7.40	13.08	2.45	10.45	2.44	11.44	3.58
LF	0.08	0.02	0.04	0.01	0.03	0.007	0.06	0.02	0.09	0.03	0.06	0.02
EDT (s)	0.42	0.03	0.09	0.02	0.11	0.03	0.08	0.02	0.15	0.02	0.12	0.03

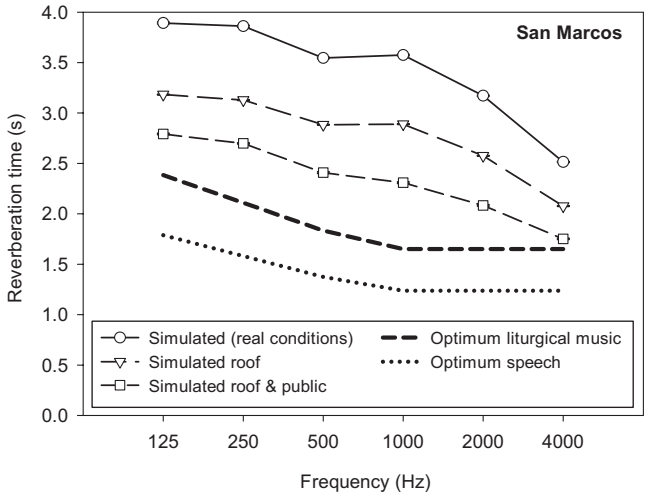


FIG. 9. Simulated reverberation time in octave bands for San Marcos church are compared with optimum reverberation times for liturgical music and speech.

of C_{80} in San Pedro (PE) church, the LF values in Santa Marina (MN) and San Marcos (MC) churches, and the EDT values in Santa Marina church are all approximately 2 JNDs. The worst case appears for T_s in San Pedro, San Julián (JU), and Santa Marina churches.

C. Remodeling in San Marcos and San Pedro churches

In order to study the recreation of the acoustic environment of ancient times and the importance of a timber roof in these places of worship, a new simulation has been carried out in San Marcos (MC) church which takes into account the (now non-existent) Moorish ceiling in its central nave. The surface of the ceiling is of 175 m^2. The absorption and scattering coefficients of the current ceiling correspond to ceramic flooring block (CB) and the coefficients of the simulated ceiling correspond to WF (see Table I). The acoustical software enables a reconstruction of the situation and an evaluation of the importance of the wooden ceiling. It is evident that the substitution of the rough ceramics by a wooden ceiling would increase the amount of absorption and decrease the reverberation time. This drop in reverberation time from *in situ* values alone would not be enough to provide better acoustical conditions (see Fig. 9) and would still locate it with Santa Marina church (with very similar reverberation time versus frequency) as one of the worst cases. The fall in reverberation time is more pronounced for low and midfrequencies, with drops ranging between 0.3 and 0.7 s.

The remaining acoustical parameters would also be influenced by this substitution. The parameters related to the intelligibility of speech and the quality for music give better conditions for the new simulation in comparison with the CB simulation. For a comparison of the change in the acoustic conditions, Fig. 10(a) presents clarity dependence on frequency. It shows an increment of around 1 dB (about 1 JND) at all octave bands from the actual situation to the modeled situation with a WF in the ceiling. The spatial dispersion, in terms of the standard error, is approximately constant. Table VII gives the mean simulated values and their standard errors for the other parameters and for each frequency octave band. From this table it is possible to see that the spatially averaged G values diminish by around 1 dB (1 JND) for each octave band from the simulated situation of the current ceramic ceiling to the situation modeled with the wooden ceiling. The values of the standard errors are almost constant for the two situations. In the case of D_{50}, the spatially averaged values increment around 3% with very similar spatial dispersion. Especially important is the reduction in the spatially averaged center time values that suppose variations between 5 and 7 JNDs. In this case the spatial dispersion parameter shows a slight reduction for all octave bands (around 5 ms). No significant variations are observed for the LF values. Finally, for the EDT parameter, variations are observed from 0.22 s (at 4000 Hz) to 0.76 s (at 125 Hz) which supposes variations between 2 and 3 JNDs.

Figure 10(b) presents the dependency of the C_{80} acoustic parameter on source-receiver distance. Simulated values follow similar trends, with a logical increase of values for the remodeling simulation. Analog behavior is observed for the remaining acoustic parameters.

The effect of the presence of the public in this place of worship may also be estimated. Here the proposal is 100% of the audience seated on the wooden pews. The absorption and

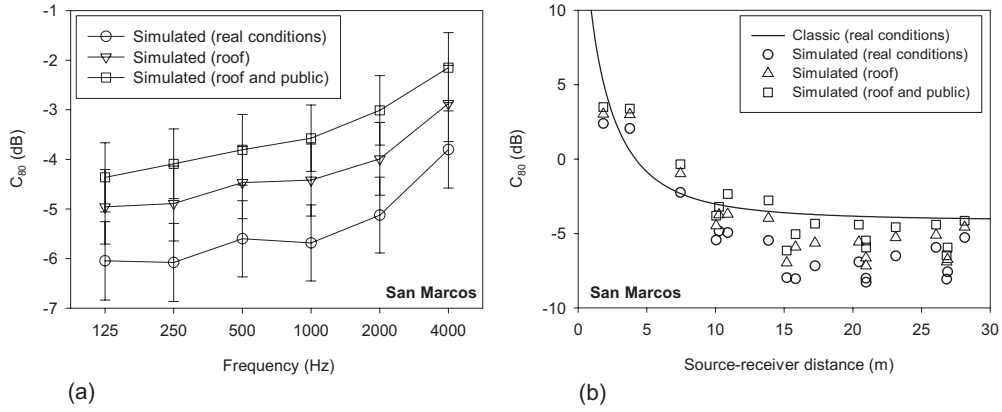


FIG. 10. Comparison of the three simulations carried out in San Marcos church for the clarity parameter (a) versus frequency and (b) versus source-receiver distance.

TABLE VII. Spatially averaged simulated values and their standard errors for each frequency octave band in San Marcos church for real conditions (above) with an imagined wooden ceiling in its central nave (center) and with an imagined wooden ceiling in its central nave and 100% occupancy (below).

Frequency	G (dB)		C_{80} (dB)		D_{50} (%)		T_s (ms)		LF		EDT (s)	
125 Hz	13.4	0.6	−6.0	0.8	14.7	3.3	306.5	15.4	0.23	0.02	4.14	0.08
	12.3	0.6	−5.0	0.8	16.9	3.3	239.8	12.1	0.23	0.02	3.38	0.11
	11.6	0.7	−4.4	0.7	18.4	3.3	211.8	10.5	0.22	0.02	2.69	0.07
250 Hz	13.4	0.6	−6.1	0.8	14.3	3.2	305.9	15.3	0.23	0.02	4.08	0.08
	12.3	0.7	−4.9	0.8	17.2	3.4	233.3	11.8	0.23	0.02	3.37	0.11
	11.5	0.7	−4.1	0.7	19.0	3.2	206.6	10.5	0.24	0.02	2.63	0.08
500 Hz	13.0	0.6	−5.6	0.8	15.1	3.3	281.1	14.4	0.25	0.02	3.73	0.09
	11.9	0.7	−4.5	0.7	18.6	3.3	219.9	11.1	0.24	0.02	3.23	0.10
	11.0	0.8	−3.8	0.7	19.9	3.4	189.4	9.91	0.23	0.02	2.38	0.07
1 kHz	13.0	0.6	−5.7	0.8	15.1	3.2	283.5	14.5	0.25	0.02	3.77	0.09
	11.9	0.7	−4.4	0.7	18.4	3.4	219.5	11.5	0.23	0.02	3.31	0.11
	10.7	0.8	−3.6	0.7	20.2	3.3	182.2	9.39	0.23	0.02	2.29	0.07
2 kHz	12.3	0.6	−5.1	0.8	16.6	3.3	250.8	13.2	0.25	0.02	3.31	0.10
	11.3	0.7	−4.0	0.7	19.9	3.4	198.7	10.7	0.23	0.02	2.96	0.10
	10.2	0.8	−3.0	0.7	22.3	3.5	165.8	9.14	0.23	0.02	2.07	0.06
4 kHz	11.1	0.7	−3.7	0.7	20.3	3.7	199.0	11.3	0.26	0.02	2.59	0.09
	10.2	0.8	−2.9	0.8	23.1	3.7	163.8	9.8	0.24	0.02	2.37	0.09
	9.2	0.9	−2.2	0.7	25.0	3.7	143.6	8.35	0.24	0.02	1.81	0.06

scattering coefficients from 125 Hz up to 4 kHz correspond to (0.57, 0.61, 0.75, 0.86, 0.91, 0.86) (Ref. 7) and (0.3, 0.4, 0.5, 0.6, 0.7, 0.8), respectively. This configuration is very frequent in cultural uses. For the congregation, the percentage is variable according to the day of the week and the hour of the day. The same observations can be made as in the previous case [see Figs. 10(a) and 10(b) and Table VII] for all the acoustical parameters, with an increase in all frequencies. The lateral energy fraction is the only parameter that does not produce a significant change in the subjective perception. The new tonal contour with the addition of an audience appears in Fig. 9.

Although it is not shown, the increase in the RASTI index in the two remodeling simulations still maintains the poor qualification of the intelligibility of speech in San Marcos church.

To further study the importance of the timber roof, another simulation was made in San Pedro church. The ceramic material (MR) which currently makes up the lateral ceilings (169 m^2) was replaced with the original wooden roof. The computer results give no variations in frequency nor in source-receiver distance for the reverberation time nor for the rest of the acoustical parameters. Isolated variations are observed in some of the receptors placed in the lateral naves.

From the detailed information of Fig. 4, the histograms of San Marcos and San Pedro churches show that the percentage of impacts on the ceiling (by adding together all the contributions from the different planes that constitute the ceiling) in the lateral naves is greater than or equal to that in the central nave (7.9% in the lateral naves and 6.4% in the central nave for San Marcos church, and 5.4% in the lateral naves and 5.4% in the central nave for San Pedro church). The remaining simulated churches present a histogram similar to that of San Marcos church. It can be seen that only a few surfaces have an important number of impacts (lateral walls, floor, and wooden roof) and hence any change of the corresponding material would produce a significant modification of the acoustical characteristics, as happened in San Marcos church. On the other hand, the number of impacts is more widely distributed for San Pedro church and the substitution of the ceramic material by the wooden roof in its lateral naves would not produce significant changes in the measurements.

V. CONCLUSIONS

Knowledge of the experimental data for reverberation times and the background noise spectrum in the churches under study allows adjustment of the values of the absorption and scattering coefficients that characterize the acoustics of their interior surfaces. This procedure, which could suppose a limitation, is at the same time an advantage since it enables the effects on the acoustic behavior of the spaces to be evaluated in an efficient way when approaching future reformation projects, restoration, maintenance, or conditioning for temporary specific uses.

In fact, this initial starting point ensures an appropriate simulation from the IR and, consequently, great precision in the values of the most significant acoustic parameters deduced from this simulated IR. It should be highlighted that this accuracy appears both in the spatially averaged values for the different octave bands and their spatial dispersion, and in the spatial distribution of the spectrally averaged values of the parameters commonly used to qualify the acoustics of the enclosure, mainly in reference to their dependence on source-receiver distance. Hence, in those parameters

which are very sensitive to position, such as the early decay time, clarity, and definition, which includes the lateral energy fraction, the results obtained in the simulation are in excellent agreement with the values measured experimentally. This agreement has been analyzed in terms of the JND for each parameter. The RASTI results have also shown a very remarkable agreement in all zones of the church in spite of the concerns regarding the various methods employed to obtain the modulation transfer functions.

The iterative process undertaken and the estimation of the irregularities appear capable of adequately describing the acoustical characteristics of materials by means of their absorption and scattering coefficients. This procedure allows a non-uniform surface with various materials to be characterized.

By means of computer simulations, it is possible to recreate the acoustic behavior of Mudejar-Gothic churches in the configurations and liturgical celebrations of past and current times. These recreations have been successfully carried out in two Mudejar-Gothic churches.

ACKNOWLEDGMENTS

The authors would like to show their appreciation for the valuable suggestions and the constructive criticism from the reviewers and the editor of the review, and also to thank the parish priests and church management for allowing the measurements to be carried out. This work has been partially supported by the Spanish MCYT Project No. BIA2003-09306-CO4-02.

[1]J. H. Rindel, "The use of computer modelling in room acoustics," Journal of Vibroengineering **3**, 219–224 (2000).
[2]M. Vorländer, *Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality* (Springer-Verlag, Berlin, 2008).
[3]I. Bork, "A comparison of room simulation software—The 2nd round robin on room acoustical computer simulation," Acust. Acta Acust. **86**, 943–956 (2000).
[4]I. Bork, "Report on the 3rd round robin on room acoustical computer simulation—Part II: Calculations," Acta. Acust. Acust. **91**, 753–763 (2005).
[5]Y. W. Lam, "A comparison of three diffuse reflection modeling methods used in room acoustics computer models," J. Acoust. Soc. Am. **100**, 2181–2192 (1996).
[6]M. Vorländer and E. Mommertz, "Definition and measurement of random-incidence scattering coefficients," Appl. Acoust. **60**, 187–199 (2000).
[7]T. J. Cox and P. D'Antonio, *Acoustic Absorbers and Diffusers. Theory, Design and Application* (Spon, London, 2004).
[8]T. J. Cox, B.-I. L. Dalenbäck, P. D'Antonio, J. J. Embrechts, J. Y. Jeon, E. Mommertz, and M. Vorländer, "A tutorial on scattering and diffusion coefficients for room acoustic surfaces," Acta. Acust. Acust. **92**, 1–15 (2006).
[9]ISO 17497-1, "Acoustics—Sound-scattering properties of surfaces. Part 1: Measurement of the random-incidence scattering coefficient in a reverberation room," International Organisation for Standardisation, Geneva, Switzerland (2004).
[10]A. Krockstadt, S. Ström, and S. Sörsdal, "Calculating the acoustical room response by the use of a ray tracing technique," J. Sound Vib. **8**, 118–125 (1968).
[11]A. Kulowski, "Algorithmic representation of the ray tracing technique," Appl. Acoust. **18**, 449–469 (1985).
[12]I. A. Drumm and Y. W. Lam, "The adaptive beam-tracing algorithm," J. Acoust. Soc. Am. **107**, 1405–1412 (2000).
[13]T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics," Appl. Acoust. **38**, 161–178 (1993).
[14]M. Vorländer, "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 (1989).
[15]R. San Martín and M. Arana, "Predicted and experimental results of acoustic parameters in the new Symphony Hall in Pamplona, Spain," Appl. Acoust. **67**, 1–14 (2006).
[16]G. Cammarata, A. Fichera, A. Pagano, and G. Rizzo, "Acoustical prediction in some Italian theatres," ARLO **2**, 61–66 (2001).
[17]A. ElKhateeb and M. Refat, "Sounds from the past: the acoustics of Sultan Hassan Mosque and Madrasa," Build. Acoust. **14**, 109–132 (2007).
[18]S. L. Vassilantonopoulos and J. N. Mourjopoulus, "A study of ancient Greek and Roman theater acoustics," Acta. Acust. Acust. **89**, 123–136 (2003).
[19]M. Galindo, T. Zamarreño, and S. Girón, "Acoustic analysis in Mudejar-Gothic churches: Experimental results," J. Acoust. Soc. Am. **117**, 2873–2888 (2005).
[20]V. O. Knudsen and C. M. Harris, *Acoustical Design in Architecture*, 5th ed. (Acoustical Society of America, New York, 1988).
[21]ISO 3382, "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters," International Organisation for Standardisation, Geneva, Switzerland (1997).
[22]IEC 60268-16, "Sound system equipment, Part 16: Objective rating of speech intelligibility by speech transmission index," International Electrotechnical Commission, Geneva, Switzerland (2003).
[23]N. Xiang and M. R. Schroeder, "Reciprocal maximum-length sequence pairs for acoustical dual source measurements," J. Acoust. Soc. Am. **113**, 2754–2761 (2003).
[24]L. G. Marshall, "An acoustic measurement program for evaluating auditoriums based on the early/late sound energy ratio," J. Acoust. Soc. Am. **96**, 2251–2261 (1994).
[25]W. Ahnert and H. P. Tennhardt, "Acoustics for Auditoriums and Concert Halls," in *Handbook for Sound Engineers*, edited by G. M. Ballou, (Elsevier, New York, 2005), pp. 109–155.
[26]ISO/DIS 3382-1, "Acoustics—Measurement of room acoustic parameters. Part I: Performance rooms," International Organisation for Standardisation, Geneva, Switzerland (2006).
[27]J. E. Summers, R. R. Torres, Y. Shimizu, and B.-I. L. Dalenbäck, "Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**, 1491–1502 (2005).
[28]Physikalisch-Technische Bundesanstalt. http://www.ptb.de/en/org/1/17/172/datenbank.htm (Last viewed April, 2009).
[29]T. Zamarreño, S. Girón, and M. Galindo, "Acoustic energy relations in Mudejar-Gothic churches," J. Acoust. Soc. Am. **121**, 234–250 (2007).
[30]E. Cirillo and F. Martellotta, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**, 232–248 (2005).
[31]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**, 3744–3749 (2006).
[32]F. Martellota, "A multi-rate decay model to predict energy-based acoustic parameters in churches (L)," J. Acoust. Soc. Am. **125**, 1281–1284 (2009).
[33]S. Girón, M. Galindo, and T. Zamarreño, "Distribution of lateral acoustic energy in Mudejar-Gothic churches," J. Sound Vib. **315**, 1125–1142 (2008).
[34]I. Bork, "Report on the 3rd round robin on room acoustical computer simulation—Part I: Measurements," Acta. Acust. Acust. **91**, 740–752 (2005).

Evaluating signal-to-noise ratios, loudness, and related measures as indicators of airborne sound insulation

H. K. Park and J. S. Bradley[a)]
National Research Council, Montreal Road. Ottawa K1A 0R6, Canada

(Received 1 June 2009; revised 3 July 2009; accepted 6 July 2009)

Subjective ratings of the audibility, annoyance, and loudness of music and speech sounds transmitted through 20 different simulated walls were used to identify better single number ratings of airborne sound insulation. The first part of this research considered standard measures such as the sound transmission class the weighted sound reduction index (R_w) and variations of these measures [H. K. Park and J. S. Bradley, J. Acoust. Soc. Am. **126**, 208-219 (2009)]. This paper considers a number of other measures including signal-to-noise ratios related to the intelligibility of speech and measures related to the loudness of sounds. An exploration of the importance of the included frequencies showed that the optimum ranges of included frequencies were different for speech and music sounds. Measures related to speech intelligibility were useful indicators of responses to speech sounds but were not as successful for music sounds. *A*-weighted level differences, signal-to-noise ratios and an *A*-weighted sound transmission loss measure were good predictors of responses when the included frequencies were optimized for each type of sound. The addition of new spectrum adaptation terms to R_w values were found to be the most practical approach for achieving more accurate predictions of subjective ratings of transmitted speech and music sounds. [DOI: 10.1121/1.3192347]

PACS number(s): 43.55.Hy [NX] Pages: 1219–1230

I. INTRODUCTION

Airborne sound transmission between spaces is measured using standardized tests[1–3] and the results of these tests, when reduced to single number ratings,[4,5] are intended to be indicative of the perceived disturbance caused by various types of transmitted sounds. A previous part of this research[6] reviewed earlier evaluations of these measures and presented new results to evaluate their effectiveness as predictors of subjective ratings of the annoyance and loudness of transmitted speech and music sounds. Although the standard sound transmission class,[4] (STC) and R_w (weighted sound reduction index[5]) ratings were not the best predictors of responses, some variations of them were found to be very successful. In particular, combinations of the R_w measure with new spectrum adaptation terms were found to be good predictors of annoyance and loudness ratings. However, for the most accurate results, different spectrum adaptation terms were needed for each type of sound (i.e., music or speech).

An earlier investigation evaluated speech intelligibility scores of transmitted speech sounds as ratings of the airborne sound insulation of walls.[7,8] Measures intended to relate to speech intelligibility scores were found to be quite successful predictors of sound insulation, when rated in terms of the intelligibility of the transmitted speech. These measures included the articulation index (AI),[9] the speech intelligibility index (SII),[10] and the articulation class (AC),[11] as well as a number of related quantities. Measures to indicate the speech privacy of enclosed meeting rooms[12] were also found to be good indicators of the intelligibility of transmitted speech sounds.

Values of the standard sound transmission measures (STC and R_w) are determined from comparisons of plots of measured sound transmission loss (TL) values versus frequency with standard rating contours. In the case of R_w, they may also include spectrum adaptation terms.[5] However, most of the speech intelligibility rating measures are based on the basic concepts developed by French and Steinberg,[13] which are incorporated in the AI and SII measures. These include frequency weighted arithmetic summations of the signal-to-noise ratios in decibels, a process that has produced successful indicators of speech intelligibility scores for speech in noise. This same concept is included in the uniformly weighted signal-to-noise ratio measure, SNR_{UNI32}, found to be a good indicator of both the audibility and intelligibility of speech transmitted through walls.[12]

The current paper extends the search for better single number ratings for accurately predicting annoyance, audibility, and loudness ratings of music and speech sounds transmitted through walls. The physical quantities considered were mostly based on existing measures of the intelligibility of speech or the loudness of sounds. Because of their known characteristics relative to intelligibility and loudness, they were expected to be at least reasonably good predictors of ratings of speech and music sounds transmitted through walls. It was hoped that the new analyses would reveal more about the important characteristics of better predictors of subjective ratings of transmitted speech and music sounds.

[a)]Author to whom correspondence should be addressed. Electronic mail: john.bradley@nrc-cnrc.gc.ca

II. EXPERIMENTAL PROCEDURE

The experimental procedure used in this study was exactly the same as described in the first part of this paper[4] and is mostly not repeated here. Subjects listened to speech and music sounds in combination with a constant level of noise representative of typical indoor ambient noise. The speech and music sounds were modified to represent transmission through walls with the TL characteristics of 20 different real walls having STC values evenly distributed from STC 34 to STC 58. Subjects heard the test sounds in a sound isolated acoustically dead room. Speech sounds were reproduced by loudspeakers 2 m in front of the listener and simulated ambient noise from a second set of loudspeakers located above the listener.

Subjects rated the annoyance of the sounds on a seven-point scale with labels from *Not at all annoying* to *Extremely annoying*. Similarly loudness ratings were obtained using a scale with labels from *Not at all loud* to *Extremely loud*. The loudness rating scale also included a possible 0 response, which could be selected to indicate that the sounds were inaudible.

Listeners heard three different Harvard sentences[14] and three different music samples through each of the 20 simulated walls for a total of 60 sentences and 60 music samples. The order of the speech and music samples and of the walls was randomized so that subjects heard conditions in one of three different randomized orders. The effective source levels of the music and speech sounds were kept constant throughout the experiment, as was the level of the simulated ambient noise at the position of the listener. The speech and music sounds, heard by the listeners, only varied due to the different sound TL characteristics of the 20 walls. This made it possible to compare signal-to-noise type measures with measures of only the sound attenuation of transmission through the walls.

Results were analyzed by plotting the mean ratings of the 20 walls versus measures of the sound insulation of the walls. Boltzmann equations were fitted to these plots and the associated R^2 values were used as indicators of the goodness of fit. Since there were always 20 data points and the same format of regression equation, the significance is simply related to the R^2 value (i.e., the coefficient of determination). Any R^2 value ≥ 0.193 is statistically significant at $p < 0.05$ and an R^2 value ≥ 0.317 at $p < 0.01$. The Boltzmann equation describes a sigmoidal-shaped relationship and was used because it fits well the form of the responses, which approach asymptotically to the maximum and minimum values of the response scales. There were a number of highly significant results and the calculated R^2 values were used to rank order the better relationships to determine the best predictors of each rating. It was the goal of this work to identify those single number measures that best predict the responses to the transmitted speech and music sounds and to determine the important components of such measures.

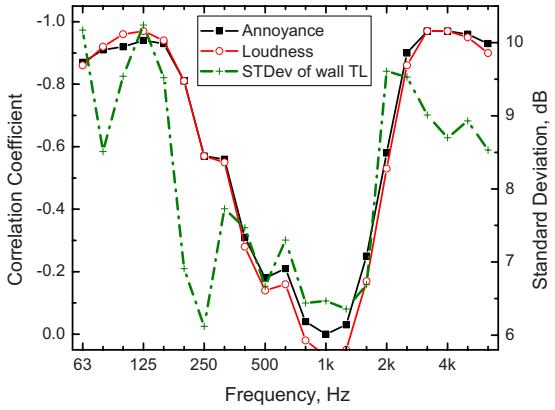


FIG. 1. (Color online) Correlation coefficients versus frequency from correlations of mean annoyance and mean loudness ratings of music sounds with 1/3-octave band TL values. The standard deviations (STDev) of the wall TL values are also plotted for comparison and have values indicated by the right hand scale.

III. EFFECTS OF INCLUDED FREQUENCIES

A. More important frequencies for TL data

The first analyses looked at the basic question of the importance of the range of frequencies included in sound insulation measures. Correlations were performed between mean subjective ratings from the annoyance and loudness tests and TL values at each 1/3-octave band frequency for the 20 walls. Figure 1 shows the correlation coefficients for linear relationships between the loudness and annoyance ratings of music sounds and the TL values at each frequency. Similar correlation coefficients for the speech sounds are given in Fig. 2. Although the results for annoyance and loudness ratings are very similar in each graph, there are large differences between the two graphs. That is, the frequencies most important for responses to music sounds are quite different than those most important for speech sounds. The dash-dotted line with cross symbols indicates the standard deviations of the wall TL values, plotted in terms of the right hand axis scale so that they appear to be comparable in magnitude with the correlation coefficients in each graph.

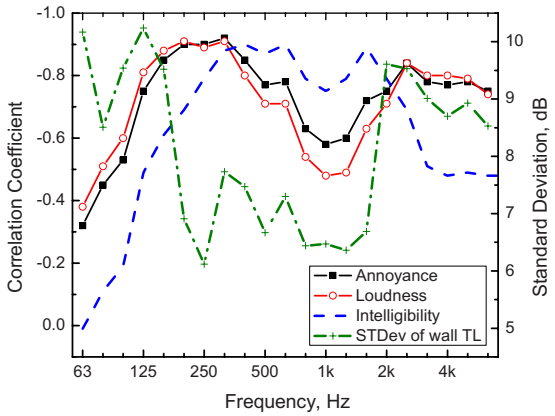


FIG. 2. (Color online) Correlation coefficients versus frequency from correlations of mean annoyance and mean loudness ratings of speech sounds as well as mean speech intelligibility scores (Refs. 7 and 8) with 1/3-octave band TL values. The standard deviations (STDev) of the wall TL values are also plotted for comparison and have values indicated by the right hand scale.

For the responses to music sounds in Fig. 1, there is a strong similarity between the variations in the correlation coefficients and the variations in the standard deviations of TL values with frequency. Since the source music had high levels at all frequencies included in Fig. 1,[15] the correlations are stronger at frequencies where there was more variation in the TL values and hence in the levels of transmitted music sounds. Thus, for music, the low and high frequency TL values best predict the annoyance and loudness ratings of the music sounds.

Figure 2 compares correlation coefficients versus frequency for mean annoyance and loudness ratings of transmitted speech sounds as well as those for speech intelligibility scores from the previous study.[7,8] As in Fig. 1, the standard deviations of the TL values are also plotted to make possible further comparisons. Compared to Fig. 1, Fig. 2 shows stronger correlation coefficients at mid-frequencies for the annoyance and loudness ratings of speech sounds. The correlation coefficients for these ratings tend to be quite large in magnitude at frequencies from 125 Hz and higher, but with a dip in the values around 1000 Hz. For speech sounds, the more limited bandwidth of the source speech signals limited the range of frequencies at which strong correlations with TL values were obtained.

The correlations with the speech intelligibility scores,[7,8] from the previous study, have a narrower range of maximum values, which extends from about 250 to 2000 Hz. These differences indicate that annoyance and loudness ratings of speech sounds are influenced by a wider range of frequencies than were speech intelligibility scores.

B. Included frequencies for arithmetic averages of TL values

In the previous study of speech intelligibility ratings of transmitted speech,[7,8] arithmetic averages of TL values over various frequency ranges were found to be good correlates of speech intelligibility scores. Arithmetic average TL values, AA(f_1–f_2), were calculated over frequency ranges from some lower frequency f_1 to an upper frequency f_2. The lower frequency, f_1, was varied from 63 to 2000 Hz and the upper frequency f_2 was varied from 200 to 6300 Hz. All of these arithmetic average TL values were correlated with annoyance ratings of speech and music sounds.

Strong correlations with annoyance ratings of music sounds only occurred when either low or high frequency TL values were included and the strongest correlations occurred when only a limited range of lower frequency bands (63–250 Hz) were included.[15] AA(63–250) values were most strongly correlated with annoyance ratings of music sounds and annoyance ratings are plotted versus this measure in Fig. 3. Although annoyance ratings of music sounds were very strongly related to AA(63–250) values (R^2=0.959), annoyance ratings of speech sounds were not so well related to this measure (R^2=0.531). (All regression coefficients from fits of Boltzmann equations are summarized in Table VII, for ratings of music sounds, and Table VIII, for ratings of speech sounds. All R^2 values are also included in the related figure titles.)

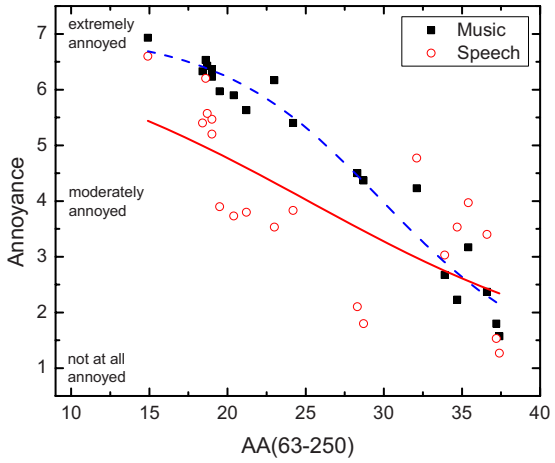


FIG. 3. (Color online) Mean annoyance ratings of speech and music sounds versus AA(63–250) and best-fit Boltzmann regression lines (music: R^2=0.959, speech: R^2=0.531).

When the range of included frequencies was extended from 63 to 6300 Hz, AA(63-6300) values were less well related to annoyance ratings of music sounds (R^2=0.788) but better related to speech sounds (R^2=0.896). The R^2 values associated with fitting Boltzmann equations to various arithmetic average TL values are compared on Table I.

Correlations of annoyance responses to speech sounds with arithmetic average TL values [AA(f_1–f_2)] for the same wide range of combinations of lower frequency, f_1, and upper included frequency, f_2, led to better correlations when a broad range of mid- and high frequency TL values was included. A number of combinations yielded quite strong correlation coefficients, but as illustrated in Fig. 4, simply

TABLE I. Summary of R^2 values from Boltzmann equation fits to mean annoyance and mean loudness ratings versus several arithmetic average TL measures [AA(f_1–f_2)], calculated over the frequency ranges from f_1 to f_2. R^2 values equal to or greater than 0.95 are in bold font.

Measure	Annoyance speech	Annoyance music	Loudness speech	Loudness music	Speech intelligibility[a]
AA(100–5000)	**0.952**	0.625	0.944	0.593	0.751
AA(200–2500)	0.839	0.209	⋯	⋯	**0.959**
AA(200–3150)	0.892	0.290	0.832	0.258	0.948
AA(160–3150)	0.924	0.345	0.871	0.309	0.931
AA(63–250)	0.531	**0.959**	0.648	**0.982**	⋯
AA(63–6300)	0.896	0.788	0.919	0.745	⋯

[a]References 7 and 8.

including all 1/3-octave bands from 100 to 5000 Hz [AA(100–5000)] was a successful combination and included all frequencies assessed in standard sound transmission tests. Although this measure predicted annoyance ratings of speech sounds very well, it was not as successful for predicting annoyance ratings of music sounds. The regression line for speech intelligibility scores from the previous study[7,8] is also included in Fig. 4 for comparison with the annoyance ratings of the speech sounds. The two speech related regression lines have different slopes near their mid-points and speech intelligibility scores vary more rapidly with AA(100–5000) values than do the ratings of annoyance to speech sounds. As a result, there is still reported annoyance when the intelligibility scores are close to 0 because listeners can hear the speech sounds even when they are not intelligible.

Several other arithmetic average TL measures were considered to find a range of frequencies that worked well for ratings of both speech and music sounds. The resulting R^2 values are summarized in Table I. In the previous study,[7,8] AA(200–2500) was best related to the speech intelligibility scores (R^2=0.959). When tested as a predictor of annoyance responses in the current work, this same measure was correlated reasonably well with annoyance to speech responses (R^2=0.839) but only weakly related to annoyance to music ratings (R^2=0.209). AA(200–3150) values were well related to annoyance to speech sounds (R^2=0.892) but were not well related to annoyance ratings of music sounds (R^2=0.290). Speech intelligibility scores (from Refs. 7 and 8) were predicted very well by this measure (R^2=0.948).

Another arithmetic average measure with a further modified frequency range, AA(160–3150), was found to be a little better compromise for both speech intelligibility scores and annoyance to speech sounds. When Boltzmann equations were fitted to plots of speech intelligibility scores versus AA(160–3150) values, the related R^2 was 0.931 and when annoyance ratings of speech sounds were considered, the associated R^2 was 0.924. However, again annoyance ratings of music sounds were not well related to this measure (R^2=0.345).

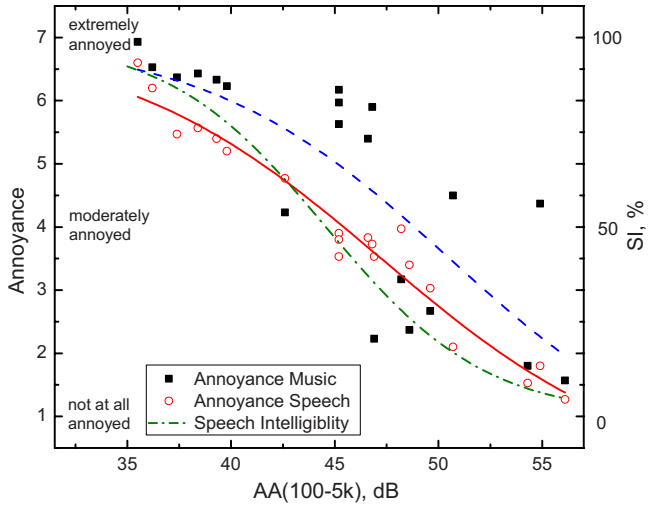


FIG. 4. (Color online) Mean annoyance ratings of speech and music sounds as well as mean speech intelligibility scores versus AA(100–5000) and best-ft Boltzmann regression lines (music: R^2=0.625, speech: R^2=0.952, speech intelligibility: R^2=0.751).

TABLE II. Summary of R^2 values from Boltzmann equation fits to mean annoyance and mean loudness ratings versus several energy average TL measures [EA(f_1–f_2)], calculated over the frequency ranges from f_1 to f_2. R^2 values equal to or greater than 0.95 are in bold font.

Measure	Annoyance speech	Annoyance music	Loudness speech	Loudness music
EA(100–5000)	0.376	0.915	0.587	**0.970**
EA(80–5000)	0.382	0.922	0.501	**0.983**
EA(500–5000)	0.931	0.417	0.879	0.365
EA(200–3150)	0.902	0.509	0.910	0.524
EA(500–3150)	0.896	0.343	0.836	0.294

Table I also includes the R^2 values that resulted from fitting Boltzmann equations to mean loudness ratings of speech and music sounds plotted versus arithmetic average TL values. As expected, arithmetic average TL values using frequency ranges that were successful for annoyance ratings were similarly successful for loudness ratings of the same type of sound.

The summary of these results in Table I shows that the optimum frequency range for speech intelligibility scores is different than that for annoyance ratings of speech sounds. However, none of the frequency ranges considered led to very accurate predictions of the ratings of both speech and music sounds.

C. Effects of included frequencies for energy average TL measures

Broadband TL values were also created by energy averaging the measured TL values over various frequency ranges rather than using arithmetic averages of the decibel values. Energy average TL measures, EA(f_1–f_2), calculated from a lower frequency, f_1 (varied from 63 to 2000 Hz) to an upper frequency, f_2 (between 200 and 6300 Hz) were correlated with annoyance ratings of speech and music sounds. The pattern of higher correlation coefficients was different than those that resulted for arithmetic average TL values.[15] Energy average TL values were strongly related to annoyance ratings of music sounds when a wide range of frequencies was included. An energy average over the frequencies from 80 to 5000 Hz, EA(80–5000), was one of the most successful measures for predicting annoyance and loudness ratings of music sounds (annoyance R^2=0.922, loudness R^2=0.983).

However, the 80–5000 Hz range was not optimum for predicting annoyance ratings of speech sounds. When energy average TL values were correlated with annoyance ratings of speech sounds, measures that included a broad range of mid- and high frequency TL values were more successful. The EA(500–5000) measure and the EA(200–3150) measure were both good predictors of ratings of speech sounds. However, these measures were not good predictors of ratings of music sounds. None of the energy average TL measures included in the results of Table II were good predictors of ratings of both speech and music sounds.

For both arithmetic average and energy average TL values, the optimum frequency ranges for speech and music sounds were different and more or less mutually exclusive. Whatever worked well for speech sounds was not successful

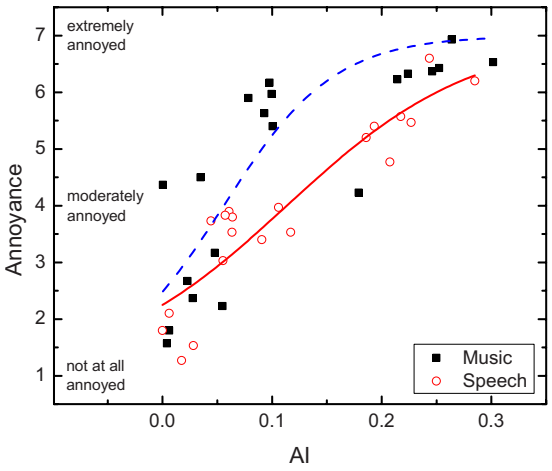


FIG. 5. (Color online) Mean annoyance ratings of speech and music sounds versus AI values and best-fit Boltzmann regression lines (music: R^2=0.671, speech: R^2=0.843).

TABLE III. Summary of R^2 values from Boltzmann equation fits to mean annoyance and mean loudness ratings versus speech intelligibility and loudness related measures. R^2 values equal to or greater than 0.95 are in bold font.

Measure	Annoyance speech	Annoyance music	Loudness speech	Loudness music
AI	0.843	0.671	0.808	0.618
SII	0.830	0.626	0.790	0.624
AC/10	0.911	0.469	0.857	0.416
SNR_{AI}	0.879	0.680	0.827	0.646
SNR_{SII22}	0.879	0.474	0.798	0.439
SNR_{UNI32}	0.898	0.425	0.913	0.456
SNR(A)	0.800	0.914	0.854	**0.963**
LR	0.693	0.908	0.793	**0.962**
LLD	0.690	0.905	0.749	0.947
STA	0.747	0.916	0.826	0.898

for music sounds and vice versa. This makes it difficult to obtain a single measure that accurately predicts ratings of both types of sounds.

The optimum range of included frequencies for arithmetic average TL measures was more restricted than those for energy average TL values. For example, the AA(63–250) measure that best predicted annoyance ratings of annoyance to music sounds was based on the seven 1/3-octave bands from 63 to 250 Hz. In comparison the EA(80–5000) measure includes 19 1/3-octave bands to predict annoyance ratings of music sounds very accurately. It is probably better to base a successful sound insulation measure on a broader range of frequencies to more completely represent the TL characteristic of walls. Because of this, energy average TL values could be better predictors of ratings of music sounds than are the best arithmetic average values.

IV. EVALUATION OF SPEECH INTELLIGIBILITY AND LOUDNESS RELATED MEASURES

Various existing measures were evaluated as predictors of annoyance and loudness ratings of the transmitted sounds. These included measures that were developed as indicators of speech intelligibility or speech privacy. Others, considered in this section, are usually intended to rate the loudness of sounds.

A. Speech intelligibility related measures

Figure 5 plots mean annoyance ratings versus AI values.[9] Although AI values are a signal-to-noise ratio measure, the source levels of the test sounds and the simulated ambient noises were held constant in these experiments and subjects heard only the changes that resulted from the varied transmission characteristics of the walls. Thus signal-to-noise ratio measures such as AI were evaluated to look for characteristics of better sound insulation measures. As might be expected, the R^2 values associated with the regression lines in Fig. 5 indicate that AI values were well correlated with annoyance ratings of speech sounds but were not so well related to ratings of music sounds. Similar R^2 values were obtained from plots of loudness ratings versus AI values and are included in Table III.

The SII is a revised version of the AI measure.[10] Regression fits of loudness and annoyance ratings of speech and music sounds with SII values resulted in very similar R^2 values as were found for AI values and are included in Table III. The R^2 values show that SII values were well correlated with responses to speech sounds but not so well related to responses to music sounds.

The AC (Ref. 11) is a frequency-weighted average attenuation measure that uses the frequency weightings from the AI standard.[9] Because AC values have the same frequency weightings as AI values, one would expect that the R^2 values associated with plots of annoyance and loudness ratings versus AC values would again show better relationships with responses to speech sounds than to those for music sounds, as shown in Table III.

Studies of ratings of the speech privacy of meeting rooms[12] found frequency-weighted arithmetic summations of signal-to-noise ratios (in decibels) to be good indicators of ratings of transmitted speech sounds. SNR_{UNI32} values were previously found to be a good indicator of both the intelligibility and audibility of transmitted speech sounds.[12] This is a uniformly weighted, arithmetic summation of signal-to-noise ratios over the 1/3-octave band frequencies from 160 to 5000 Hz. In calculating SNR_{UNI32} values, the signal-to-noise ratios in each band are clipped to never fall below −32 dB where they would have no effect on the audibility of the sounds. Annoyance and loudness ratings of speech sounds were strongly related to this measure (see Table III) and the results for annoyance ratings are shown in Fig. 6. Responses to music sounds were again only moderately well related to SNR_{UNI32} values as found for the other speech intelligibility related measures. This is most likely due to the lack of low frequency information in this measure.

Table III also includes R^2 values from plots of SNR_{AI} and SNR_{SII22} values versus annoyance and loudness ratings. SNR_{AI} values are a frequency-weighted arithmetic summation of signal-to-noise ratios over the same frequencies (i.e., 200–5000 Hz) and using the same frequency weightings as

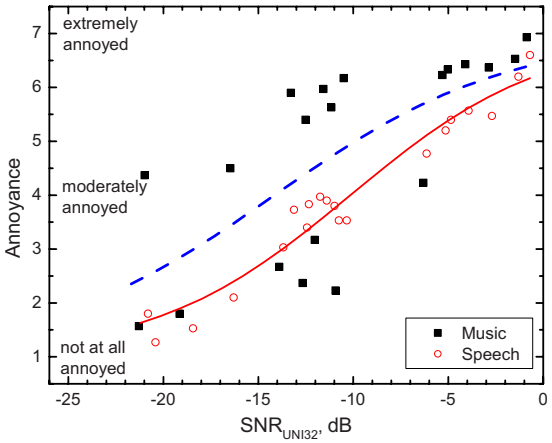


FIG. 6. (Color online) Mean annoyance ratings of speech and music sounds versus SNR_{UNI32} values and best-ft Boltzmann regression lines, [Music: R^2=0.425, Speech: R^2=0.898].

the AI measure. SNR_{SII22} is a weighted arithmetic summation of signal-to-noise ratios over the 1/3-octave band frequencies from 160 to 5000 Hz using the frequency weightings from the SII measure. In calculating SNR_{SII22} values, the signal-to-noise ratios in each band are clipped to never fall below −22 dB, below which they would have no effect on intelligibility. SNR_{SII22} values were previously found to be a good indicator of the intelligibility of transmitted speech.[12] The results for both SNR_{AI} and SNR_{SII22} values in Table III indicate relatively strong relationships with annoyance and loudness ratings of speech sounds but less strong relationships with ratings of music sounds.

B. Loudness related measures

While there are complex measures for assessing the loudness of sounds, simple *A*-weighted sound levels are often used as an approximate indicator of the loudness of sounds. Similarly, the signal-to-noise ratios of the *A*-weighted speech and noise levels [i.e., the difference in the *A*-weighted levels of the transmitted speech or music and the ambient noise, SNR(*A*)] can be used to rate sound insulation. Previous studies have found SNR(*A*) values to be a poor indicator of speech intelligibility.[7,8] However, the results in Fig. 7, indicate that SNR(*A*) values are well related to annoyance ratings of speech sounds and even better related to annoyance ratings of music sounds. Plots of loudness ratings of speech and music sounds versus SNR(*A*) values led to even higher R^2 values (speech: R^2=0.854, music: R^2=0.963). This suggests that *A*-weighted ratings may be quite successful indicators of the sound insulation provided against music sounds. This would be in agreement with much earlier results of Vian *et al.*[16] They found *A*-weighted level differences, limited to the 1/3-octave band frequencies from 125 to 4000 Hz, to be good predictors of the annoyance ratings of transmitted music sounds.

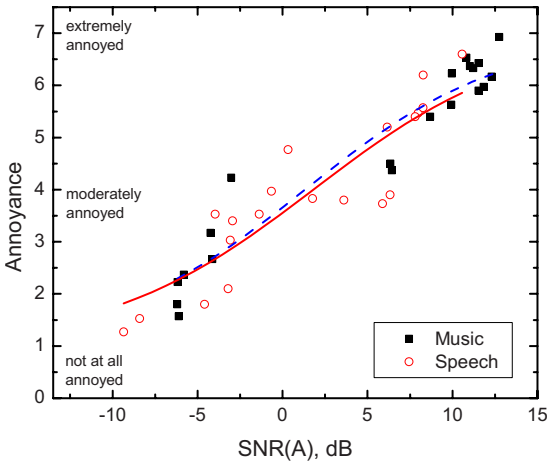


FIG. 7. (Color online) Mean annoyance ratings of speech and music sounds versus SNR(*A*) values and best-ft Boltzmann regression lines (music: R^2=0.914, speech: R^2=0.800).

If SNR(*A*) values are considered to be an approximate measure of loudness, then true loudness measures would be expected to be successful predictors of the subjective ratings. Loudness ratios (LRs) were created from loudness values in sones[17] for the transmitted speech (or music) and noise signals. The R^2 values from fitting Boltzmann regression equations to plots of annoyance and loudness ratings of speech and music sounds versus LR values are included in Table III. Like the SNR(*A*) values, LR values were moderately good predictors of responses to speech sounds and were very good predictors of responses to music sounds. Although the LR values are less well related to speech sounds than to those of music sounds, they are better related to loudness ratings of the speech and music sounds than to annoyance ratings of these sounds.

As the R^2 values in Table III indicate, plots of annoyance and loudness ratings versus loudness level differences (LLDs) led to slightly lower R^2 values than those for LR values. However, the LLD values were again better predictors of responses to music sounds than those to speech sounds and predicted loudness ratings better than annoyance ratings for both speech and music sounds.

Because *A*-weighted signal-to-noise ratios were quite successful predictors of subjective ratings, it was of interest to test an *A*-weighted sound TL measure, STA, as a predictor of annoyance and loudness ratings of the transmitted sounds. STA was defined as follows:[18]

$$\mathrm{STA} = -10\log\left\{\frac{1}{17}\sum_{b=100}^{4k} 10^{(-TL_b+Awt_b)/10}\right\}\ \mathrm{dB}, \qquad (1)$$

where b is the centre frequency of each 1/3-octave band, TL_b is the sound TL in the 1/3-octave band with center frequency b, and Awt_b is the attenuation of the *A*-weighting curve at frequency b.

STA values were found to predict annoyance and loudness ratings of speech and music sounds quite well. Although the R^2 values are not quite as high as for some measures included in Table III, STA values are seen to be a relatively good predictor of all responses.

The results of the various regression analyses in this section are summarized in Table III. They generally show a consistent pattern of measures that are either better related to responses to speech sounds or to those for music sounds but are not equally well related to responses to both types of sounds. Measures intended to relate to the intelligibility of speech or other closely related measures (AI, SII, AC,

TABLE IV. Summary of R^2 values from Boltzmann regression equations fitted to plots of audibility ratings versus various sound insulation measures. R^2 values equal to or greater than 0.95 are in bold font. Results for STC, R_w, R_w+C, and $R_w+C_{tr(100–3150)}$ from Ref. 4.

Measure	Audibility speech	Audibility music
STC	**0.968**	0.452
R_w	**0.971**	0.526
R_w+C	0.903	0.757
R_w+C_{tr}	0.853	0.920
SNR_{SII22}	0.480	0.356
SNR_{UNI32}	0.650	0.242
SNR(A)	0.932	0.946
LR	0.269	0.770
LLD	0.895	0.911
STA	0.874	0.827

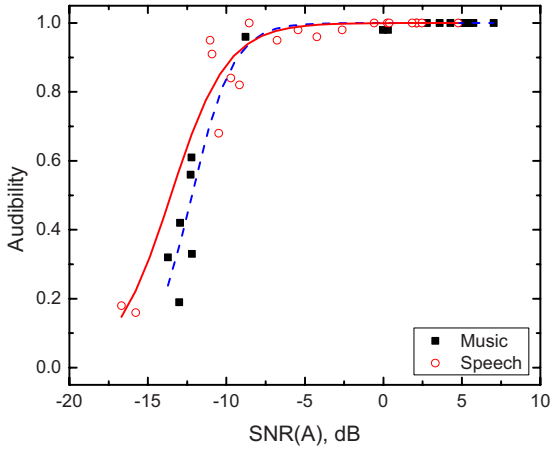


FIG. 8. (Color online) Fraction of subjects finding sounds just audible versus SNR(A) values and best-fit Boltzmann regression lines (music: $R^2=0.946$, speech: $R^2=0.932$).

SNR_{AI}, SNR_{SII22}, and SNR_{UNI32}) were most strongly related to responses to speech sounds. Loudness related measures [LR values from loudness ratings in sones, and LLD values from loudness level ratings in phons, SNR(A), and STA values] were most strongly related to responses to music sounds. Almost all of this latter group were a little better related to subjective ratings of loudness than to ratings of annoyance.

V. AUDIBILITY OF TRANSMITTED SOUNDS

During the loudness-rating test, subjects could give a score of zero to indicate they could not hear any music or speech sounds. The fraction of subjects scoring greater than 0 was used as a measure of the audibility of the sounds. The first part of this paper[4] showed that both STC and R_w values were very good predictors of the audibility of speech sounds but were not as successful for the audibility of music sounds. When spectrum adaptation terms [C and $C_{tr(100–3150)}$] were added to R_w values,[5] the measures became better predictors of the audibility of music sounds but less successful as predictors of the audibility of speech sounds. The R^2 values from the Boltzmann equation best-fit regression lines to plots of audibility ratings versus STC, R_w, R_w+C and $R_w+C_{tr(100–3150)}$ values (from Ref. 4) are repeated in Table IV.

Previous speech privacy studies[12] found the SNR_{UNI32} measure to be successfully related to both audibility and intelligibility ratings of transmitted speech. In the new results, neither SNR_{UNI32} nor SNR_{SII22} values were very good predictors of audibility ratings. However, the R^2 values in Table IV show that SNR_{UNI32} values were better for predicting the audibility of speech sounds than the audibility of music sounds.

Although LLD values were good predictors of the audibility of both speech and music sounds, LR values were less successful (see R^2 summary in Table IV). Surprisingly SNR(A) values were very good predictors of the audibility of both speech and music sounds, as shown in Fig. 8. Because SNR(A) values were found to be very good predictors of audibility ratings, the A-weighted TL measure, STA, was also tested as a predictor of audibility ratings. The R^2 values in Table IV show that STA values predicted the audibility of speech and music sounds a little less well than did the SNR(A) values.

The pattern of relationships between the fraction of subjects finding sounds audible and the various sound insulation ratings summarized in Table IV was different than for the relationships for annoyance and loudness ratings. Because of these differences, the importance of the included frequencies in sound insulation measures for predicting the audibility of sounds was also investigated. As before this was done by calculating arithmetic average TL values, AA(f_1–f_2), and energy average TL values, EA(f_1–f_2), for a range of combinations of lower, f_1, and upper, f_2, included frequencies.

The correlations with arithmetic average TL values only showed higher magnitude correlation coefficients for quite limited frequency ranges and were still quite modest in value. For correlations with ratings of the audibility of music sounds, the strongest correlations resulted when only a limited range of low frequency 1/3-octave bands (63–200 Hz) were included and the highest magnitude correlation coefficient was only −0.87 with the AA(63–200) values. For correlations with the audibility ratings of speech sounds, the highest magnitude correlation coefficients occurred when only the highest frequency 1/3-octave bands were included. For the audibility of speech sounds, the highest magnitude correlation coefficient was found for the AA(1000–6300) measure with a value of −0.75.

When energy average TL values were considered, correlation coefficients with slightly higher magnitude values were obtained. For ratings of the audibility of music sounds, correlations with the EA(63–6300) measure led to a correlation coefficient of −0.89. For correlations with audibility ratings of speech sounds the EA(1000–6300) measure led to a correlation coefficient of −0.77.

The correlation coefficients with both types of frequency averaged TL values were disappointing and suggested that these average TL values are not as successful as other approaches considered for predicting the audibility ratings of sounds.

Table IV summarizes the results of the regression analyses of audibility ratings versus sound insulation measures.

TABLE V. Description of the *A*-weighted level difference measures calculated.

Name	Source spectrum	Included frequencies
$LDA_P(63–6300)$	Pink	63–6300 Hz
$LDA_P(200–6300)$	Pink	200–6300 Hz
$LDA_P(100–4000)$	Pink	100–4000 Hz
$LDA_E(63–6300)$	E597	63–6300 Hz
$LDA_E(200–6300)$	E597	200–6300 Hz
$LDA_E(100–4000)$	E597	100–4000 Hz

While some measures were better predictors of the audibility of speech sounds (STC, R_w, R_w+C, and STA), others were better predictors of the audibility of music sounds [$R_w+C_{tr(100–3150)}$, LLD, and SNR(A)]. Only $R_w+C_{tr(100–3150)}$, LLD, and SNR(A) values were reasonably good predictors of the audibility of both speech and music sounds and SNR(A) values best predicted the audibility of both types of sounds.

VI. A-WEIGHTED LEVEL DIFFERENCES FOR RATING SOUND INSULATION

Earlier research by Vian *et al.*,[16] found that annoyance ratings of transmitted music sounds were significantly related to the *A*-weighted levels of the sounds transmitted through simulated walls if the included frequencies were limited to the range from 125 to 4000 Hz. If a wider frequency range of transmitted sounds from 40 to 10 000 Hz was included, the *A*-weighted levels were not significantly related to annoyance responses.

For a number of years ASTM had a standard measurement procedure [ASTM E597 (Ref. 19)] for rating the *A*-weighted level reductions between two spaces. The standard (now withdrawn) was intended to be a simple method for obtaining approximate STC values by measuring *A*-weighted level differences between room-average levels in adjacent rooms.

Previous analyses in this paper have indicated that some *A*-weighted ratings can be successful indicators of the annoyance, loudness and audibility of transmitted speech and music sounds. The use of the *C*-type spectrum adaptation term with R_w values is equivalent to measuring an *A*-weighted level difference with a pink source spectrum over the frequencies from 100 to 3150 Hz. *A*-weighted signal-to-noise ratios, SNR(A), were found to be best related to ratings of the loudness and audibility of transmitted speech and music sounds. An *A*-weighted sound TL measure, STA, was also reasonably well related to annoyance and loudness ratings of transmitted speech and music sounds. However, our previous studies of the intelligibility of transmitted speech[7,8,12] showed the STA and SNR(A) measures were not very successful as indicators of the intelligibility of transmitted speech sounds.

Because of the previous interest in *A*-weighted sound insulation ratings, and because of the success of some *A*-weighted measures reported earlier in this paper, further considerations of this type of measure are included in this section. The new measures consisted of *A*-weighted level differences between the *A*-weighted source and transmitted levels for the simulated walls. These were calculated for two source spectra: (a) an ideal flat pink spectrum and (b) the mean of the range of spectra acceptable in the previous ASTM E597 standard. The ASTM E597 spectrum peaked at 500 Hz and dropped off gradually either side of this band with a more rapid decrease below 125 Hz. For each source spectrum, *A*-weighted transmitted levels were calculated for three different ranges of included frequencies. The ranges of included frequencies were established from initial tests to determine the most appropriate ranges for responses: (a) to speech sounds (200–6300 Hz), (b) to music sounds (63–6300 Hz), and (c) a range similar to existing standards (100–4000 Hz). Table V lists the symbols used to indicate each of these *A*-weighted level differences along with the source spectrum and included frequency range used.

The mean values of all available responses were plotted versus each of these measures including annoyance, loudness, and audibility ratings for both speech and music sounds as well as the speech intelligibility scores from the previous study.[7,8] As in previous analyses, Boltzmann equations were fitted to the data and the associated R^2 values determined. These are included in Table VI along with results for STA and SNR(A) values, which are repeated from earlier sections of this paper to facilitate comparisons.

The results in Table VI show that annoyance and loudness ratings of music sounds were most strongly related to *A*-weighted noise level differences using a pink noise source with included frequencies from 63 to 6300 Hz. However, for loudness ratings of music sounds, the R^2 value for the rela-

TABLE VI. Summary of R^2 values for Boltzmann regression equations fitted to plots of mean responses versus various *A*-weighted level differences. R^2 values equal to or greater than 0.95 are in bold font. As described in previous sections, the STA and SNR(A) measures include frequencies from 63–6300 Hz.

Source spectrum	Pink	Pink	Pink	E597	E597	E597	Speech /music	Speech /music
Included frequencies	63–6300	200–6300	100–4000	63–6300	200–6300	100–4000	STA	SNR(A)
Annoyance music	**0.978**	0.530	0.916	0.801	0.408	0.777	0.915	0.911
Loudness music	**0.966**	0.511	0.898	0.792	0.398	0.773	0.897	**0.968**
Audibility music	0.823	0.244	0.753	0.539	0.149	0.507	0.827	0.946
Annoyance speech	0.668	**0.977**	0.865	0.862	**0.954**	0.868	0.746	0.756
Loudness speech	0.763	**0.960**	0.826	0.913	0.930	0.917	0.825	0.859
Audibility speech	0.891	0.753	0.912	**0.978**	0.669	**0.978**	0.874	0.932
Intelligibility speech[a]	0.246	0.778	0.294	0.527	0.819	0.515	0.361	0.259

[a]Speech intelligibility results from previous study (Refs. 7 and 8).

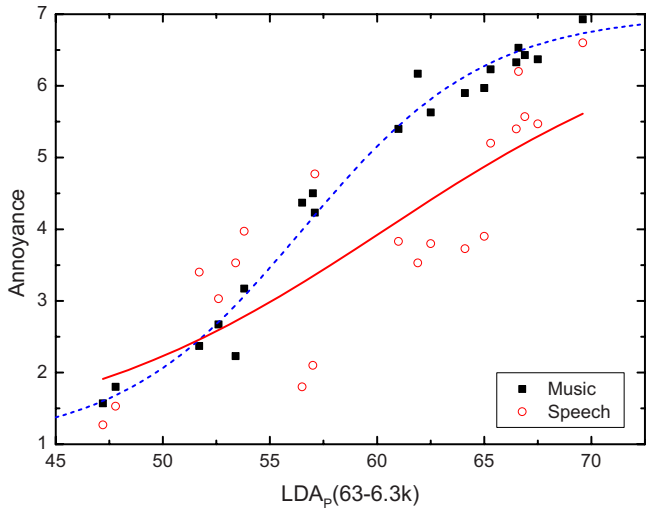


FIG. 9. (Color online) Mean annoyance ratings of speech and music sounds versus *A*-weighted level differences [LDA$_P$(63–6300)] using a pink source spectrum and best-ft Boltzmann regression lines (music: R^2=0.978, speech R^2=0.668).

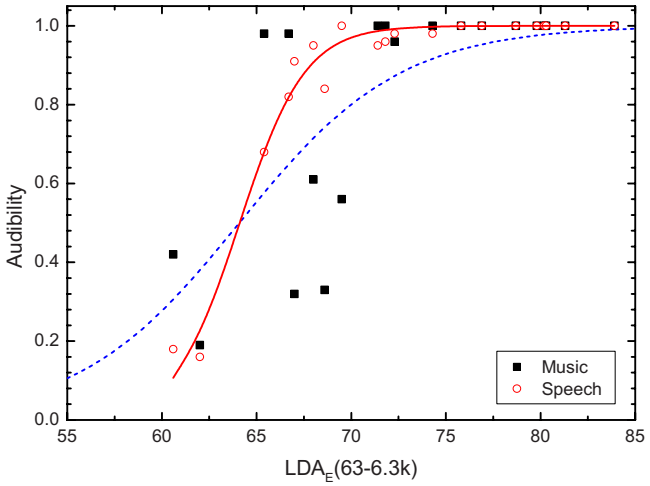


FIG. 10. (Color online) Fraction of subjects finding sounds just audible versus *A*-weighted level differences, LDA$_E$(63 to 6300 Hz), using an E597 source spectrum and best-ft Boltzmann regression lines (music: R^2=0.539, speech: R^2=0.978).

tionship with SNR(*A*) values was slightly higher. SNR(*A*) values were also very strongly related to ratings of the audibility of music sounds.

For annoyance and loudness ratings of speech sounds, R^2 values were highest for *A*-weighted level differences over the 200–6300 Hz range with a pink source spectrum. The audibility of speech was better related to *A*-weighted level differences using an E597 source spectrum for either the 63–6300 Hz range or the 100–4000 Hz range. This may imply that some lower frequency sound helps to make faint speech sounds more audible, even though the low frequency sound does not significantly improve the intelligibility of the speech. The 200–6300 Hz range, which most limited low frequency content, was most helpful for predicting speech intelligibility scores when the E597 source spectrum was used.

Figure 9 plots the relationship between annoyance ratings of speech and music sounds and LDA$_P$(63–6300) values. As indicated in the title of Fig. 9 and Table VI (which gives a summary of the R^2 values), annoyance to music responses were very well related to LDA$_P$(63–6300) values. However, annoyance ratings of speech sounds were less well related to this measure.

A plot of annoyance to speech and music sounds versus LDA$_P$(200–6300) values showed a very good fit for annoyance ratings of speech sounds. However, annoyance ratings of music sounds were less well related to this measure.

The audibility scores for speech sounds were best related to LDA$_E$(100–4000) and LDA$_E$(63–6300) values. The latter relationship is shown in Fig. 10. These two relationships had the same R^2 value, which was the highest of any included in Table VI. However, these two measures predicted the audibility of music sounds less well. The audibility of music sounds is best predicted from SNR(*A*) values as shown previously in Fig. 8 and included in the Table VI summary.

The results in Table VI show that level difference measures that used a pink source spectrum more frequently lead to the highest R^2 values than did those using the E597 spectrum. As will be seen in Sec. VII of this paper, some of the R^2 values in Table VI are among the highest found in this study.

VII. SUMMARY AND DISCUSSION

In this paper, subjective ratings of the audibility, annoyance, and loudness of speech and music sounds transmitted through 20 different walls were related to a number of potentially useful single number measures of the sound transmission characteristics of the walls. For responses to music sounds, Table VII summarizes the most important regression results including R^2 values and the regression coefficients for the Boltzmann equations[4] fitted to each sound insulation measure. Table VIII provides a similar summary for relationships with responses to speech sounds. These tables also include the key results of the first part of this research[4] to facilitate comparisons of the effectiveness of the various measures.

Although a large number of apparently different measures were considered, they can be grouped into a small number of categories according to how they were calculated in this study. This is done in Table IX which shows that there are two main categories of measures: (1) those which were derived only from the 1/3-octave band TL values and (2) those that were derived from the TL values and were also influenced by the source spectrum. Of these two main categories there are five sub-categories describing the calculation method. For measures derived from only TL values, some were obtained using rating contours, some from arithmetic summations of TL values, and others from energy summations of TL values. For the measures including the effect of both TL values and a source spectrum, calculations were either in the form of a signal-to-noise ratio at the receiving point or a level difference between incident and transmitted sounds. Finally, all of the measures were

TABLE VII. Summary of R^2 values and Boltzmann equation regression coefficients (Ref. 6) of the main cases for music sounds in this study. R^2 values equal to or greater than 0.95 are in bold font. An R^2 value ≥ 0.193 is statistically significant at $p<0.05$ and R^2 values ≥ 0.317 at $p<0.01$.

			Annoyance			Loudness			Audibility		
Symbol	f_1	f_2	R^2	X_0	dx	R^2	X_0	dx	R^2	X_0	dx
STC	125	4000	0.728	47.916	5.667	0.734	43.417	5.275	0.452	54.742	5.593
R_w	100	3150	0.798	47.507	4.602	0.779	43.744	4.527	0.526	52.979	4.331
STC_{no8}	125	4000	0.670	48.884	5.334	0.654	44.839	5.074	0.378	55.820	5.761
R_w+C	100	3150	0.918	44.624	4.371	0.900	40.591	4.638	0.757	48.789	2.910
R_w+C_{mod}	50	5000	0.580	37.605	5.313	0.556	33.735	5.106	0.297	44.594	5.949
$R_w+C_{200-3150}$	200	3150	0.508	33.101	6.110	0.521	28.800	5.378	0.228	41.378	7.177
$R_w+C_{tr100-3150}$	100	3150	**0.950**	39.140	5.035	**0.960**	34.528	5.074	0.920	43.946	2.369
$R_w+C_{tr50-5000}$	50	5000	0.943	36.812	5.753	**0.980**	31.574	4.543	0.948	43.143	1.878
$R_w+C_{tr\_mod}$	50	5000	0.983	34.966	4.683	0.991	30.778	4.099	0.889	39.989	2.944
$R_{w(63-5000)}$	63	5000	0.940	46.060	3.864	0.907	42.633	3.863	0.750	50.226	2.979
$R_{w(160-5000)}$	160	5000	0.562	49.977	5.808	…	…	…	…	…	…
AA(100–5000)	100	5000	0.625	48.522	4.974	0.593	44.769	5.149	0.330	55.063	5.591
AA(200–3150)	200	3150	0.290	53.277	8.507	0.258	47.710	8.790	…	…	…
AA(63–250)	63	250	**0.959**	29.904	5.182	**0.982**	25.284	4.937	0.926	35.121	2.062
AA(63–6300)	63	6300	0.788	46.733	4.157	0.745	43.363	4.298	0.508	51.844	4.159
EA(100–5000)	100	5000	0.915	33.424	5.885	**0.970**	28.051	5.649	0.866	−38.239	1.714
EA(80–5000)	80	5000	0.922	31.551	6.218	0.983	26.084	5.165	0.933	−39.514	1.636
EA(500–5000)	500	5000	0.417	53.806	6.666	0.365	49.220	6.845	0.174_{ns}	−63.048	8.134
AI	200	5000	0.671	0.056	0.050	0.618	0.040	0.039	…	…	…
SII	160	8000	0.626	0.081	0.076	0.624	0.057	0.052	…	…	…
AC/10	200	5000	0.469	56.039	6.360	0.416	51.542	6.686	…	…	…
SNR_{AI}	200	5000	0.680	−25.501	4.722	0.646	−27.910	4.711	…	…	…
SNR_{SII22}	160	5000	0.474	−15.474	5.436	0.439	−16.254	4.039	0.356	−21.725	2.511
SNR_{UNI32}	160	5000	0.425	−14.144	5.154	0.456	−15.792	1.636	0.242	−27.335	6.425
SNR(A)	63	6300	0.914	1.323	5.814	**0.963**	0.416	4.613	**0.946**	−12.192	1.312
LR	25	12500	0.908	1.190	0.464	**0.962**	1.027	0.317	0.770	0.2737	0.0521
LLD	25	12500	0.905	−0.4395	7.065	0.947	−1.432	6.062	0.911	−18.843	2.689
STA	100	4000	0.916	46.905	4.355	0.898	42.889	4.622	0.755	51.115	2.935
LDA_P(63–6300)	63	6300	**0.978**	56.538	4.261	**0.966**	60.400	4.174	0.823	51.319	2.683
LDA_P(200–6300)	63	6300	0.530	50.642	5.667	0.511	54.668	5.426	0.244	42.573	6.924
LDA_P(100–4000)	63	6300	0.916	55.403	4.359	0.898	59.417	4.623	0.753	51.222	2.942
LDA_E(63–6300)	63	6300	0.801	69.442	4.618	0.792	73.277	4.584	0.539	64.690	4.261
LDA_E(200–6300)	63	6300	0.408	65.035	6.899	0.398	69.701	6.575	0.149_{ns}	54.287	9.273
LDA_E(100–4000)	63	6300	0.777	69.244	4.808	0.733	73.182	4.730	0.507	63.681	4.476

influenced by the range of 1/3-octave band frequencies included in their calculation.

Some measures that might appear to be quite different actually are very similar. For example, the R_w+C measure can be exactly equivalent to an A-weighted level difference with a pink source spectrum, $LDA_P(f_1-f_2)$, if the range of included frequencies, f_1-f_2, is exactly the same. An example of this is seen in the results in Table VII, where the R^2 for the fit of annoyance ratings of speech versus R_w+C values is 0.918 (100–3150 Hz) and for LDA_P(100–4000) is 0.916. The slightly different range of included frequencies leads to only a very small difference in R^2 values.

From examining the results in Tables VII and VIII, it was not possible to detect a consistent trend for better results to be obtained for one of the two main categories of sound insulation measures. There are some very high R^2 values for each major type of measure as well as for most subcategories of measures. Other differences such as the range of included frequencies and the combination of the range of included frequencies with the type of calculation seem to determine which measure is a better predictor of particular responses.

VIII. CONCLUSIONS

In general, loudness and annoyance responses yielded similar information. Although there was a small trend for loudness ratings of music sounds to be more accurately predicted than annoyance ratings, usually little was gained by asking subjects to rate sounds in terms of both concepts. However, audibility ratings usually led to a different pattern of results than those for the annoyance and loudness ratings.

The various measures are strongly influenced by the range of included frequencies and the optimum range of included frequencies is different for predicting ratings of music sounds than for ratings of speech sounds. Because of this some measures are strongly related to ratings of speech sounds and others to ratings of music sounds, but none were highly successful for both types of sounds.

TABLE VIII. Summary of R^2 values and Boltzmann equation regression coefficients (Ref. 6) of the main cases for speech sounds in this study. R^2 values equal to or greater than 0.95 are in bold font. An R^2 value ≥ 0.193 is statistically significant at $p<0.05$ and R^2 values ≥ 0.317 at $p<0.01$.

			Annoyance			Loudness			Audibility		
Symbol	f_1	f_2	R^2	X_0	dx	R^2	X_0	Dx	R^2	X_0	dx
STC	125	4000	0.856	43.955	7.018	0.886	39.322	6.906	**0.968**	54.106	1.731
R_w	100	3150	0.890	43.946	6.176	0.933	39.891	6.013	**0.971**	53.045	1.551
STC_{no8}	125	4000	**0.950**	45.271	5.804	**0.970**	41.299	5.842	0.919	54.915	1.830
R_w+C	100	3150	0.741	40.552	7.459	0.821	36.050	6.818	0.903	50.763	1.777
R_w+C_{mod}	50	5000	**0.975**	34.198	4.994	**0.973**	30.716	5.142	0.863	42.877	1.720
$R_w+C_{200–3150}$	200	3150	0.901	29.393	5.551	0.912	25.719	5.557	0.701	38.349	2.211
$R_w+C_{tr100–3150}$	100	3150	0.566	34.370	9.529	0.676	29.197	8.063	0.853	46.252	0.465
$R_w+C_{tr50–5000}$	50	5000	0.388	32.276	11.912	0.482	26.447	9.560	0.856	44.888	0.090
$R_w+C_{tr\_mod}$	50	5000	0.541	31.084	8.840	0.634	26.422	7.167	0.866	42.593	0.945
$R_{w(63–5000)}$	63	5000	0.769	42.623	6.361	0.848	38.816	5.701	0.871	52.352	1.493
$R_{w(160–5000)}$	160	5000	**0.972**	46.301	5.359	…	…	…	…	…	…
AA(100–5000)	100	5000	**0.952**	45.154	5.310	0.944	41.358	5.602	0.831	53.764	2.062
AA(200–3150)	200	3150	0.892	48.380	5.471	0.832	44.555	5.793	…	…	…
AA(63–250)	63	250	0.531	25.161	9.811	0.648	19.993	8.016	0.852	36.969	0.180
AA(63–6300)	63	6300	0.896	43.468	4.157	0.919	39.744	5.520	0.937	52.075	1.327
EA(100–5000)	100	5000	0.376	28.085	12.311	0.587	21.976	9.433	0.236	−46.473	8.754
EA(80–5000)	80	5000	0.382	26.489	12.754	0.501	20.393	9.809	0.375	41.223	1.440
EA(500–5000)	500	5000	0.931	49.660	5.008	0.879	46.198	5.156	0.676	−58.787	2.804
AI	200	5000	0.843	0.113	0.085	0.808	0.074	0.049	…	…	…
SII	160	8000	0.830	0.147	0.102	0.790	0.106	0.064	…	…	…
AC/10	200	5000	0.911	52.047	5.469	0.857	48.171	5.832	…	…	…
SNR_{AI}	200	5000	0.879	−21.300	5.731	0.827	−23.309	6.126	…	…	…
SNR_{SII22}	160	5000	0.879	−11.809	4.716	0.798	−13.658	4.046	0.480	−20.994	1.111
SNR_{UNI32}	160	5000	0.898	−10.149	5.154	0.913	−12.307	5.350	0.650	−23.804	2.336
SNR(A)	63	6300	0.800	1.814	6.056	0.854	−0.413	6.079	0.932	−13.538	1.788
LR	25	12500	0.693	0.9739	0.474	0.796	0.770	0.2923	0.269	0.214	0.0153
LLD	25	12500	0.690	−1.784	7.805	0.749	−4.308	8.043	0.895	−22.359	−1.045
STA	100	4000	0.747	42.863	7.400	0.826	38.384	6.769	0.912	52.985	1.757
LDA_P(63−6300)	63	6300	0.668	64.407	7.669	0.763	64.747	6.536	0.891	49.481	1.277
LDA_P(200−6300)	63	6300	**0.977**	54.139	4.980	**0.960**	57.674	5.257	0.753	45.355	2.578
LDA_P(100−4000)	63	6300	0.865	59.447	7.393	0.826	63.922	6.770	0.912	49.341	1.792
LDA_E(63−6300)	63	6300	0.862	73.058	6.423	0.913	77.243	6.231	**0.978**	64.112	1.654
LDA_E(200−6300)	63	6300	0.954	69.058	5.074	0.930	72.616	5.351	0.669	64.112	1.654
LDA_E(100−4000)	63	6300	0.868	72.910	6.467	0.917	77.150	6.306	**0.978**	63.819	1.621

A compromise approach for achieving better predictions of both speech and music sounds would be to use a measure that is reasonably well related to responses to both speech and music sounds such as $R_{w(63–5000)}$, AA(63–6300) or SNR(A) values. However, these compromise approaches would lead to less accurate predictions of some responses with lower R^2 values than the best possible relationships that had R^2 values of 0.97 and higher.

TABLE IX. Lists of sound insulation measures arranged by category and sub-category.

Categories	Depend only on wall TL			Depend on source spectrum and wall TL	
Sub-categories	Rating contour	Arithmetic summation	Energy summation	Signal-to-noise ratios	Level differences
Measures	STC STC_{no8} $R_w(f_1-f_2)$	AA(f_1-f_2) AC	EA(f_1-f_2) STA	SNR_{AI} SNR_{SII22} SNR_{UNI32} AI SII LR LLD SNR(A)	$LDA_P(f_1-f_2)$ $LDA_E(f_1-f_2)$ R_w+C R_w+C_{tr}

More accurate predictions can be obtained, in a manner that is practical to implement, by using R_w with different spectrum adaptation terms for each type of sound. For example, one could use the C_{mod} spectrum adaptation term for speech sounds and the $C_{\mathrm{tr\_mod}}$ spectrum adaptation term[6] for music sounds. This led to very high R^2 values (>0.97) for annoyance and loudness ratings of both speech and music sounds and could be achieved by adding new spectrum weighting terms to the existing ISO procedure.

Various *A*-weighted measures can provide reasonably accurate predictions of some responses using an easy to implement measure. Choosing an *A*-weighted level difference measure summed over the frequency range best suited to each type of sound [i.e., $\mathrm{LDA}_P(63-6300)$ for music sounds and $\mathrm{LDA}_P(200-6300)$ for speech sounds] would lead to predictions essentially as accurate as the best of the other options. Of course, this would not have the convenience of building on an existing standard procedure. Alternatively, measures such as $\mathrm{LDA}_P(100-4000)$ and $\mathrm{SNR}(A)$ would be simple but accurate compromise measures for both speech and music sounds.

These results may be limited by the characteristics of the 20 walls that were simulated. They were chosen to include a wide range of construction types and included a wide range of overall sound insulation values. There may be other constructions that might not fit the same pattern of results. In particular, the simulated TL characteristics were based on the results of standard laboratory tests of walls. It is now well established that sound TL in real buildings is typically limited by various flanking paths.[20,21] These flanking paths often significantly reduce the actual sound TL in a manner that would vary with frequency. Carrying out a similar study based on field measurements of the apparent sound TL characteristics of a wide range of walls might lead to some differences.

ACKNOWLEDGMENTS

The authors would like to acknowledge that a Korea Research Foundation Grant, funded by the Korean Government (MOEHRD) (KRF-2006-352-D00200) to Dr. H.K.P. supported his contribution to this work. They would also like to thank Dr. Brad Gover for helpful discussions during this project.

[1]ASTM E90-92, "Standard test method for laboratory measurement of airborne sound transmission loss of building partitions and elements," ASTM International, West Conshohocken, PA.
[2]ASTM E336-97, "Standard test method for measurement of airborne sound insulation in buildings," ASTM International, West Conshohocken, PA.
[3]ISO 140 "Acoustics—Measurement of sound insulation in buildings and of building elements"—Part 3, "Laboratory measurement of airborne sound insulation of building elements" (2004), Part 4, "Field measurements of airborne sound insulation between rooms" (1998).
[4]ASTM E413, "Classification for rating sound insulation," ASTM International, West Conshohocken, PA.
[5]ISO-717-1, "Acoustics—Rating of sound insulation in buildings and of building elements—Part 1: Airborne sound insulation," International Organization for Standardization (1996).
[6]H. K. Park and J. S. Bradley, "Evaluating standard airborne sound insulation measures in terms of annoyance, loudness, and audibility ratings," J. Acoust. Soc. Am. **126**, 208–219 (2009).
[7]H. K. Park, J. S. Bradley, and B. N. Gover, "Evaluation of airborne sound insulation in terms of speech intelligibility," Institute for Research in Construction, National Research Council, Ottawa, Canada, Research Report No. RR-228, March 2007 (Revised June 2007).
[8]H. K. Park, J. S. Bradley, and B. N. Gover, "Evaluating airborne sound insulation in terms of speech intelligibility," J. Acoust. Soc. Am. **123**, 1458–1471 (2008).
[9]ANSI S3.5-1969, American national standard methods for the calculation of the articulation index, Standards Secretariat, Acoustical Society of America, New York.
[10]ANSI S3.5-1997, "Methods for calculation of the speech intelligibility index," American National Standard, Standards Secretariat, Acoustical Society of America, New York.
[11]ASTM E1110-01, "Standard classification for determination of articulation class," ASTM International, West Conshohocken, PA.
[12]B. N. Gover and J. S. Bradley, "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms," J. Acoust. Soc. Am. **116**, 3480–3490 (2004).
[13]N. R. French and J. C. Steinberg, "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119 (1947).
[14]"IEE Recommended Practice for Speech quality Measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246 (1969).
[15]H. K. Park, J. S. Bradley, and B. N. Gover, "Rating airborne sound insulation in terms of annoyance and loudness of transmitted speech and music sounds," Institute for Research in Construction, National Research Council, Ottawa, Canada, Research Report No. RR-242, November 2008.
[16]J.-P. Vian, W. F. Danner, and J. W. Bauer, "Assessment of significant acoustical parameters for rating sound insulation of party walls," J. Acoust. Soc. Am. **73**, 1236–1243 (1983).
[17]E. Zwicker, H. Fastl, and C. Dallmayr, "BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B," Acustica **55**, 63–67 (1984).
[18]J. S. Bradley, "Subjective rating of party walls," Can. Acoust. **11**, 37–45 (1983).
[19]ASTM E597-1999, "Standard Practice for Determining a Single-Number Rating of Airborne Sound Isolation for Use in Multiunit Building Specifications," ASTM International, West Conshohocken, PA.
[20]R. J. M. Craik, T. R. T. Nightingale, and J. A. Steel, "Sound transmission through a double leaf partition wall with edge flanking," J. Acoust. Soc. Am. **101**, 964–969 (1997).
[21]T. R. T. Nightingale, J. D. Quirt, F. King, and R. E. Halliwell, "Flanking transmission in multi-family dwellings: Phase IV," Institute for Research in Construction, National Research Council, Ottawa, Canada, Research Report No. RR-218 , March 2006.

A *k*-space method for acoustic propagation using coupled first-order equations in three dimensions

Jason C. Tillett
Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627

Mohammad I. Daoud
Department of Electrical and Computer Engineering and Robarts Research Institute, University of Western Ontario, London, Ontario N6A 5B9, Canada

James C. Lacefield
Department of Electrical and Computer Engineering, Department of Medical Biophysics, and Robarts Research Institute, University of Western Ontario, London, Ontario N6A 5B9, Canada

Robert C. Waag
Department of Electrical and Computer Engineering and Department of Imaging Sciences, University of Rochester, Rochester, New York 14627

(Received 31 December 2008; revised 26 May 2009; accepted 28 May 2009)

A previously described two-dimensional *k*-space method for large-scale calculation of acoustic wave propagation in tissues is extended to three dimensions. The three-dimensional method contains all of the two-dimensional method features that allow accurate and stable calculation of propagation. These features are spectral calculation of spatial derivatives, temporal correction that produces exact propagation in a homogeneous medium, staggered spatial and temporal grids, and a perfectly matched boundary layer. Spectral evaluation of spatial derivatives is accomplished using a fast Fourier transform in three dimensions. This computational bottleneck requires all-to-all communication; execution time in a parallel implementation is therefore sensitive to node interconnect latency and bandwidth. Accuracy of the three-dimensional method is evaluated through comparisons with exact solutions for media having spherical inhomogeneities. Large-scale calculations in three dimensions were performed by distributing the nearly 50 variables per voxel that are used to implement the method over a cluster of computers. Two computer clusters used to evaluate method accuracy are compared. Comparisons of *k*-space calculations with exact methods including absorption highlight the need to model accurately the medium dispersion relationships, especially in large-scale media. Accurately modeled media allow the *k*-space method to calculate acoustic propagation in tissues over hundreds of wavelengths.
 [DOI: 10.1121/1.3158857]

PACS number(s): 43.58.Ta, 43.20.Fn, 43.80.Qf, 43.35.Fj [TDM] Pages: 1231–1244

I. INTRODUCTION

The calculation of acoustic propagation through tissue over a large number of wavelengths is a challenging problem because the resources required for computations over such distances can exceed commonly available desktop facilities when the computations are performed in three dimensions. The solution of this problem is an integral part of inversion algorithms and a valuable tool for studying aberration estimation and correction techniques in ultrasound imaging. Furthermore, large-scale calculations can model research and clinical experimental configurations at full scale. The problem, sometimes called the forward problem, is solved in this paper by extending a computationally efficient solution to three dimensions from two dimensions and by using a cluster of computers with a message-passing interface (MPI). The efficient solution is an extension of the so-called *k*-space method described in the work of Tabei *et al.*[1] In that work, which extended second-order formulations[2–5] of this method, coupled first-order equations were used, a perfectly matched layer (PML) (Ref. 6) was incorporated on the boundary of the computational domain, and loss was described by relaxation absorption,[7] which is the dominant cause of loss in tissue. The *k*-space method was originally described by Bojarski.[8,9]

Although *k*-space code that calculates propagation in three dimensions can be executed on a typical desktop computer, the computational scale is limited by the physical memory that is available because the number of cells or voxels must be less than 9×10^7 for a desktop computer with 16 Gbytes of memory. This translates for the current method into a $416\times416\times416$ grid for a representative 45 single-precision floating-point variables per grid cell. The large number of variables is comprised of, but not limited to, a decomposed pressure, velocity field components, medium parameters (one of which is required on four spatial grids), relaxation times and generalized compressibilities (one for each absorption process), and parameters of the PML. Using a minimum of four points-per-wavelength at the highest frequency contained in a typical pulse propagating through the medium, desktop computing resources fall short of applica-

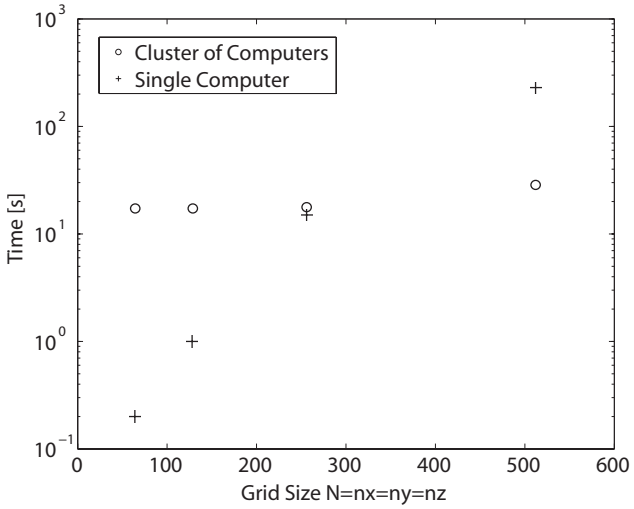


FIG. 1. Time required to execute a three-dimensional FFT on a single computer compared to the time to execute a three-dimensional FFT on a cluster of 32 computers for a grid size of $N=64$, 128, 256, or 512. Each point is the average of ten trials. The single-computer architecture is a 64-bit AMD Opteron having 16 Gbytes of random access memory (RAM) running Windows XP x64 and a 64-bit version of MATLAB. The cluster of computers consists of Pentium-III 1.4 GHz processors each having 0.5 Gbyte of RAM, running Linux, and connected via a 1-Gbit/s Ethernet switched network.

tions such as full-scale calculation of scattering in the breast, evaluation of aberration correction techniques, and simulation of *b*-scan imaging.[10] The three-dimensional *k*-space code described here for large-scale calculations is designed to run on relatively low-cost Beowulf clusters[11] using MPI, where the amount of physical memory is the aggregate of all of the individual memories, and uses a slab-decomposition of the medium and associated variables, where each compute node handles a range of indices on one of the three-dimensions. Memory capacity is not the only limiting factor. Execution time as shown in Fig. 1 increases as a power of the size of the three-dimensional fast Fourier transform (FFT), the numerical method of the three-dimensional *k*-space code containing the bulk of the necessary arithmetic operations required to implement the algorithm. However, if the three-dimensional FFT is distributed over many nodes, as can be accomplished using the fastest-Fourier-transform-in-the-west (FFTW) software,[12] the exponent of the power-law scaling is effectively decreased.

Second-order formulations of the *k*-space method have been extended to three dimensions.[2,13] These formulations can have absorbing boundary conditions and absorption. However, the PML incorporated in the first-order implementation can suppress reverberations more efficiently for thinner boundary layers. Additionally, relaxation absorption included in the first-order implementation allows for a frequency dependence like that observed in tissues. Other *k*-space methods for wave propagation in three dimensions are available[14,15] but constrain the incident field and are not amenable to incorporation of relaxation absorption that is dominant in tissues.

Analysis of errors and performance is of particular interest in the three-dimensional *k*-space forward-problem solution because high-cost specialized computing platforms, such as clusters of computers, must be used for the large-scale problems of interest. Specifically, any increase in the number of points-per-wavelength required to achieve a specified accuracy translates directly into larger grid sizes.

The *k*-space method is ideally suited for calculating propagation in media where fluctuations in compressibility and density are small as they are in soft tissue. In the calculation, the scattering medium is sampled but previous experience has shown that sudden changes in medium parameters introduce artifacts. Such artifacts have been addressed by using half-band filtering[16] that has been shown to achieve a good balance between suppressing artifacts and maintaining high spatial frequencies. This balance results in scattered wave amplitudes that closely match results of exact calculations. Accurate Fourier decomposition of the medium produces better half-band filtering but typically requires oversampling the medium and, thus, increased *k*-space grid size. Since *k*-space grid sizes for large-scale computations are already large, oversampling can become a problem even for a large cluster. Therefore, attention is given to characterize the impact of varying amounts of oversampling.

The treatment in this paper begins with a mathematical formulation that explicitly extends the two-dimensional *k*-space method to three dimensions and compares the method to developments in related publications. The methods used to evaluate performance are described next. Results of numerical experiments are then presented. This is followed by a discussion of the results. Finally, conclusions are drawn from the reported studies.

II. MATHEMATICAL FORMULATION

A. Physical acoustic foundations

The development of equations for application of the *k*-space method to calculate acoustic wave propagation in liquids over a scale of many wavelengths in three dimensions is summarized here and related to results in other publications. Throughout the summary, the notation and conventions are consistent. The use of different notations and conventions from publication to publication is noted.

The summary begins with the equation for conservation of mass

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{1}$$

the equation for momentum time rate of change

$$\rho\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla\right)\mathbf{v} = -\nabla\rho, \tag{2}$$

and the equation for conservation of energy neglecting thermal conduction

$$\rho\frac{D}{Dt}\left(\frac{1}{2}v^2 + u\right) = -\nabla \cdot (p\mathbf{v}), \tag{3}$$

which are commonly developed in acoustic textbooks, e.g., Ref. 17 or Ref. 18. In these equations, ρ is medium density, $\mathbf{v}$ is velocity with $\|\mathbf{v}\|=v$, p is total pressure, u is internal energy per unit volume, and the operator D/Dt is the time rate of change evaluated in a coordinate system that is sta-

tionary with respect to a moving infinitesimal volume element and is equivalent to the operator in parentheses on the left side of Eq. (2). Relaxation absorption is included using the phenomenological theory of irreversible thermodynamics[19,20] in which time rate of change in extensive parameters is related to deviations of the extensive parameters from equilibrium values and flows of energy and particles are driven by generalized forces (i.e., affinities) to obtain a Gibbs–Duhem relation[21]

$$T\frac{Ds}{Dt} = \frac{Du}{Dt} + p\frac{D(\rho^{-1})}{Dt} + \sum_{\nu} A_{\nu}\frac{Dn_{\nu}}{Dt}. \tag{4}$$

In this equation, T is temperature, s is entropy per unit volume, n_{ν} is particle number density, A_{ν} is particle affinity, and the subscript ν denotes the particle species.

Assuming the acoustic perturbations are small and neglecting effects of gravity, viscosity, and heat conduction lead Eqs. (1)–(4) as shown in Ref. 7 to a linear set of equations that are

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v}) = 0, \tag{5}$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\nabla p, \tag{6}$$

$$\frac{1}{c^2}\frac{\partial p}{\partial t} - \left(\frac{\partial \rho}{\partial t} + \mathbf{v} \cdot \nabla \rho_0\right) + \sum_{\nu} \rho_0 \kappa_{\nu} \frac{\partial (\Delta \xi_{\nu})}{\partial t} = 0, \tag{7}$$

and

$$\left(\frac{\partial}{\partial t} + \frac{1}{\tau_{\nu}}\right)(\Delta \xi_{\nu}) = -\frac{\partial p}{\partial t}. \tag{8}$$

In these equations, ρ_0 is the equilibrium density, c is acoustic wave speed, κ_{ν} and τ_{ν} are the isothermal compressibility and relaxation time constant, respectively, of species ν, and $\Delta \xi_{\nu}$ represents a pressure change produced by the deviation of n_{ν} from its equilibrium value $n_{\nu,0}$. This pressure change is given by

$$\Delta \xi_{\nu} = (n_{\nu} - n_{\nu,0})/(\partial n_{\nu,0}/\partial p)_T. \tag{9}$$

Substitution of Eq. (9) in the above equations to eliminate $\Delta \xi_{\nu}$ yields equations for acoustic pressure and particle velocity that are the starting point in parallel developments of time-domain calculation algorithms found in Refs. 1 and 22. These equations, without terms incorporating a PML that will be included later, are

$$\rho_0(\mathbf{x})\frac{\partial}{\partial t}\mathbf{v}(\mathbf{x},t) = -\nabla p(\mathbf{x},t) \tag{10}$$

and

$$\kappa(\mathbf{x},t) \otimes \frac{\partial}{\partial t} p(\mathbf{x},t) = -\nabla \cdot \mathbf{v}(\mathbf{x},t). \tag{11}$$

Equations (10) and (11) are three-dimensional extensions of Eqs. (13)–(16) in Ref. 1 or, equivalently, Eqs. (7)–(10) in Ref. 22 that are written in each reference as components in two-dimensional space.

Expansion of Eqs. (10) and (11) in Cartesian coordinates yields

$$\frac{\partial}{\partial x} p(\mathbf{x},t) = -\rho_0(\mathbf{x})\frac{\partial}{\partial t} v_x(\mathbf{x},t),$$

$$\frac{\partial}{\partial y} p(\mathbf{x},t) = -\rho_0(\mathbf{x})\frac{\partial}{\partial t} v_y(\mathbf{x},t),$$

$$\frac{\partial}{\partial z} p(\mathbf{x},t) = -\rho_0(\mathbf{x})\frac{\partial}{\partial t} v_z(\mathbf{x},t), \tag{12}$$

and

$$\kappa(\mathbf{x},t) \otimes \frac{\partial}{\partial t} p(\mathbf{x},t) + \frac{\partial}{\partial x} v_x(\mathbf{x},t) + \frac{\partial}{\partial y} v_y(\mathbf{x},t) + \frac{\partial}{\partial z} v_z(\mathbf{x},t) = 0. \tag{13}$$

In these equations, $\otimes$ denotes temporal convolution and $\kappa(\mathbf{x},t)$ is the generalized compressibility given by[7]

$$\kappa(\mathbf{x},t) = \kappa_{\infty}(\mathbf{x})\,\delta(t) + \sum_{\nu} \frac{\kappa_{\nu}(\mathbf{x})}{\tau_{\nu}(\mathbf{x})} e^{-t/\tau_{\nu}(\mathbf{x})} H(t), \tag{14}$$

where $H(t)$ is the Heaviside step function and

$$\kappa_{\infty}(\mathbf{x}) = \kappa_0(\mathbf{x}) - \sum_{\nu} \kappa_{\nu}(\mathbf{x}) \tag{15}$$

in which

$$\kappa_0(\mathbf{x}) = 1/[\rho(\mathbf{x})c^2(\mathbf{x})]. \tag{16}$$

B. Inclusion of a PML

A PML on the boundary of the computation is integrated into the development by assuming an $e^{j\omega t}$ time dependence, which is the same convention assumed in Ref. 23, and transforming Eqs. (12) and (13) into the temporal-frequency domain. The resulting equations are[a]

$$\frac{\partial}{\partial x} p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x}) v_x(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial y} p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x}) v_y(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial z} p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x}) v_z(\mathbf{x},\omega), \tag{17}$$

and

$$\nabla \cdot \mathbf{v}(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega) p(\mathbf{x},\omega). \tag{18}$$

Next, attenuation terms that are zero within the medium and nonzero in the PML are introduced. These terms are adjusted to attenuate waves in the boundary regions and eliminate reverberations of outgoing waves. An attenuation term $\alpha^*(\mathbf{x},\omega)$ associated with generalized compressibility is assumed to have dispersion like that of the generalized compressibility. An attenuation term $\alpha(\mathbf{x})$ associated with the density is assumed to be dispersionless because the density is dispersionless. The asterisk notation is the same as the nota-

tion in Refs. 1 and 22 but is opposite the convention in Ref. 23. The modified equations are

$$\frac{\partial}{\partial x}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_x(\mathbf{x},\omega) - \alpha(\mathbf{x})v_x(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial y}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_y(\mathbf{x},\omega) - \alpha(\mathbf{x})v_y(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial z}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_z(\mathbf{x},\omega) - \alpha(\mathbf{x})v_z(\mathbf{x},\omega), \tag{19}$$

and

$$\nabla\cdot\mathbf{v}(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p(\mathbf{x},\omega) - \alpha^*(\mathbf{x},\omega)p(\mathbf{x},\omega). \tag{20}$$

Two steps are now taken. The first is to decompose pressure artificially into three Cartesian components by writing

$$p(\mathbf{x},\omega) = p_x(\mathbf{x},\omega) + p_y(\mathbf{x},\omega) + p_z(\mathbf{x},\omega). \tag{21}$$

This decomposition enables inclusion of the PML boundary conditions. The second is to assume the PML attenuation coefficients are anisotropic. Decomposing the pressure and separating Eq. (20) into three equations yield

$$\frac{\partial}{\partial x}v_x(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_x(\mathbf{x},\omega) - \alpha^*(\mathbf{x},\omega)p_x(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial y}v_y(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_y(\mathbf{x},\omega) - \alpha^*(\mathbf{x},\omega)p_y(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial z}v_z(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_z(\mathbf{x},\omega) - \alpha^*(\mathbf{x},\omega)p_z(\mathbf{x},\omega). \tag{22}$$

If α^* is anisotropic, then Eq. (22) becomes

$$\frac{\partial}{\partial x}v_x(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_x(\mathbf{x},\omega) - \alpha_x^*(\mathbf{x},\omega)p_x(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial y}v_y(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_y(\mathbf{x},\omega) - \alpha_y^*(\mathbf{x},\omega)p_y(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial z}v_z(\mathbf{x},\omega) = -j\omega\kappa(\mathbf{x},\omega)p_z(\mathbf{x},\omega) - \alpha_z^*(\mathbf{x},\omega)p_z(\mathbf{x},\omega). \tag{23}$$

If α is anisotropic, then Eq. (19) becomes

$$\frac{\partial}{\partial x}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_x(\mathbf{x},\omega) - \alpha_x(\mathbf{x})v_x(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial y}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_y(\mathbf{x},\omega) - \alpha_y(\mathbf{x})v_y(\mathbf{x},\omega),$$

$$\frac{\partial}{\partial z}p(\mathbf{x},\omega) = -j\omega\rho_0(\mathbf{x})v_z(\mathbf{x},\omega) - \alpha_z(\mathbf{x})v_z(\mathbf{x},\omega). \tag{24}$$

Equations (23) and (24) can be further simplified using a relation between $\alpha^*(\mathbf{x},\omega)$ and $\alpha(\mathbf{x})$ that can be derived by enforcing the constraint that normally incident plane waves at the boundaries of the medium are not reflected. This condition follows from assuming a plane-wave solution with the forms

$$\tilde{\mathbf{p}} = \tilde{\mathbf{p}}_0 e^{-j\boldsymbol{B}\cdot\mathbf{x}} \tag{25}$$

and

$$\tilde{\mathbf{v}} = \tilde{\mathbf{v}}_0 e^{-j\boldsymbol{B}\cdot\mathbf{x}}, \tag{26}$$

substituting Eq. (25) into Eq. (23) and Eq. (26) into Eq. (24), assuming κ and ρ are independent of $\mathbf{x}$, solving for the unknowns $\boldsymbol{B}$ and $\tilde{\mathbf{v}}$, and using the techniques detailed in Ref. 6. The condition is

$$\alpha_{(\cdot)}^*(\mathbf{x},t) = \kappa(\mathbf{x},t)\frac{\alpha_{(\cdot)}(\mathbf{x})}{\rho(\mathbf{x})}, \tag{27}$$

in which $(\cdot)$ denotes x, y, or z. When Eqs. (23) and (24) are Fourier transformed back to the time domain and reflections are eliminated using Eq. (27) the set of time-dependent first-order k-space equations that describe acoustic propagation in three dimensions in a medium that has relaxation absorption and a PML on the boundary becomes

$$\rho(\mathbf{x})\left[\frac{\partial v_{(\cdot)}(\mathbf{x},t)}{\partial t} + \alpha_{(\cdot)}(\mathbf{x})v_{(\cdot)}(\mathbf{x},t)\right] = -\frac{\partial[p_x(\mathbf{x},t) + p_y(\mathbf{x},t) + p_z(\mathbf{x},t)]}{\partial(\cdot)} \tag{28}$$

and

$$\kappa(\mathbf{x},t)\otimes\left[\frac{\partial p_{(\cdot)}(\mathbf{x},t)}{\partial t} + \alpha_{(\cdot)}(\mathbf{x})p_{(\cdot)}(\mathbf{x},t)\right] = -\frac{\partial v_{(\cdot)}(\mathbf{x},t)}{\partial(\cdot)}. \tag{29}$$

The above set of equations is an extension of Eqs. (13)–(16) of Ref. 1 from two dimensions to three dimensions. If κ and α^* are assumed to be independent of temporal frequency and Eqs. (23) and (24) are Fourier transformed back to the time domain, then, using

$$\alpha = \alpha^*\rho/\kappa, \tag{30}$$

a symmetric set of equations in three dimensions results. These equations are the three-dimensional analogs of the two-dimensional equations (11)–(14) in Ref. 23. If Eqs. (23) and (24) are Fourier transformed back to the time domain, Eqs. (7)–(10) in Ref. 22 result.

C. Solution for scattered pressure

Equations (28) and (29) describe the propagation of total pressure. Corresponding equations that describe the propagation of only the scattered pressure can be derived by substituting $p=p_s+p_i$ into Eqs. (28) and (29), noting the incident pressure satisfies the same set of equations in a homogeneous background medium, and eliminating the total pressure in the resulting set of equations. To present results with fewer subscripts in what follows, the subscript i is used to denote incident quantities and the absence of a subscript denotes scattered quantities. A compact set of equations, which is the solution for scattered pressure in continuous time and continuous space, is written by defining a state variable $S_\nu(\mathbf{x},t)$ for each relaxation process. The state variables are filtered versions of the pressure field and are given by

$$S_\nu(\mathbf{x},t) \equiv p(\mathbf{x},t) \otimes \frac{\kappa_\nu(\mathbf{x})}{\tau_\nu(\mathbf{x})} e^{-t/\tau_\nu(\mathbf{x})} H(t). \tag{31}$$

These state variables combined with identities[22] for convolutions including time derivatives allow the equations for scattered pressure to be written as

$$\frac{\partial v_{(\cdot)}(\mathbf{x},t)}{\partial t} + \alpha_{(\cdot)}(\mathbf{x}) v_{(\cdot)}(\mathbf{x},t) = -\frac{1}{\rho(\mathbf{x})} \frac{\partial p(\mathbf{x},t)}{\partial(\cdot)} - \left[\frac{1}{\rho(\mathbf{x})} + \frac{1}{\rho_0}\right] \frac{\partial p_i(\mathbf{x},t)}{\partial(\cdot)} \tag{32}$$

and

$$\begin{aligned}
&[\kappa_\infty(\mathbf{x})]\left[\frac{\partial p_{(\cdot)}(\mathbf{x},t)}{\partial t} + \alpha_{(\cdot)}(\mathbf{x}) p_{(\cdot)}(\mathbf{x},t)\right] + \sum_\nu \frac{\kappa_\nu(\mathbf{x})}{\tau_\nu(\mathbf{x})} p_{(\cdot)}(\mathbf{x},t) \\
&\quad + \sum_\nu \left[-\frac{1}{\tau_\nu(\mathbf{x})} + \alpha_{(\cdot)}(\mathbf{x})\right] S_\nu^{(\cdot)}(\mathbf{x},t) \kappa_\nu(\mathbf{x}) \\
&\quad = -\frac{\partial v_{(\cdot)}(\mathbf{x},t)}{\partial(\cdot)} - [\kappa_\infty(\mathbf{x}) - \kappa_0]\left[\frac{\partial p_{(\cdot),i}(\mathbf{x},t)}{\partial t} + \alpha_{(\cdot)}(\mathbf{x}) p_{(\cdot),i}(\mathbf{x},t)\right] - \sum_\nu \frac{\kappa_\nu(\mathbf{x})}{\tau_\nu(\mathbf{x})} p_{(\cdot),i}(\mathbf{x},t) \\
&\quad - \sum_\nu \left[-\frac{1}{\tau_\nu(\mathbf{x})} + \alpha_{(\cdot)}(\mathbf{x})\right] S_\nu^{(\cdot),i}(\mathbf{x},t) \kappa_\nu(\mathbf{x}).
\end{aligned} \tag{33}$$

To facilitate a compact but more detailed presentation of these equations on a discrete spatial grid and at discrete temporal points, the following quantities are defined:

$$\mu_{(\cdot)}(\mathbf{x}) \equiv \frac{1}{\kappa_\infty(\mathbf{x})} \sum_\nu \frac{\kappa_\nu(\mathbf{x})}{\tau_\nu(\mathbf{x})} + \alpha_{(\cdot)}(\mathbf{x}) \tag{34}$$

and

$$\eta_\nu^{(\cdot)}(\mathbf{x}) \equiv \frac{\kappa_\nu(\mathbf{x})}{\tau_\nu(\mathbf{x})} - \kappa_\nu(\mathbf{x}) \alpha_{(\cdot)}(\mathbf{x}). \tag{35}$$

Also, analogous to the decomposition of the total pressure, the incident pressure is written as

$$p_i = p_{x,i} + p_{y,i} + p_{z,i}. \tag{36}$$

In principle, this decomposition of incident pressure introduces three more state variables than a corresponding solution for total pressure. In practice, however, only one additional state variable is required because the decomposed incident pressure is recombined at the beginning of each temporal iteration.

Finally, spatial and temporal staggering analogous to that described in Ref. 1 are used to improve computational accuracy. The three-dimensional extension of the spatially staggered two-dimensional grid defined in Ref. 1 is shown in Fig. 2. The spatial staggering and spatial derivatives are implemented using the so-called *k*-space operators[1] in a three-dimensional form that can be expressed as

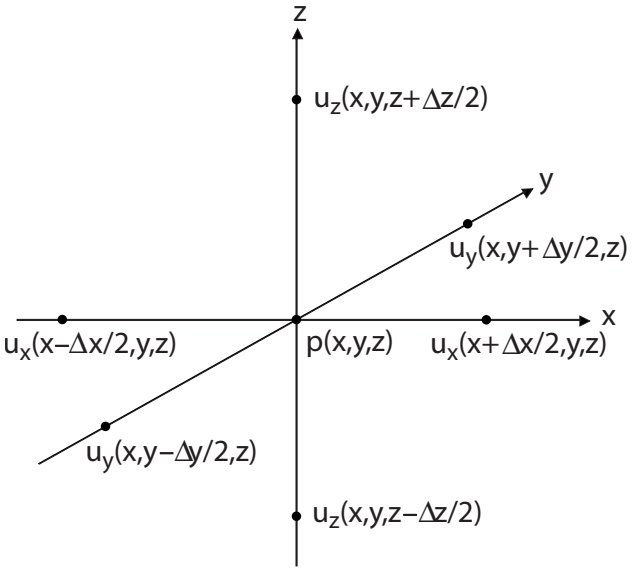


FIG. 2. Spatially staggered grid.

$$\frac{\partial p(\mathbf{x},t)}{\partial^{(c_0\Delta t)^\pm}(\cdot)} \equiv \mathbb{F}^{-1}[jk_{(\cdot)} e^{\pm jk_{(\cdot)}\Delta^{(\cdot)}/2} \operatorname{sinc}(c_0 \Delta t k/2) \mathbb{F}[p(\mathbf{x},t)]], \tag{37}$$

where $\mathbb{F}$ denotes a spatial Fourier transform, k is the magnitude of the wavenumber with components (k_x, k_y, k_z), and the superscripted partial derivatives denote the dependence of the derivatives on step size and equilibrium wave speed. The *k*-space operators replace the spatial derivatives in Eqs. (32) and (33). As a result, the staggered variables are evaluated at points on three distinct spatial grids, one for each component of the pressure and velocity. The points are $\mathbf{x}_1 = \mathbf{x} + \mathbf{e}_x \Delta x/2$, $\mathbf{x}_2 = \mathbf{x} + \mathbf{e}_y \Delta y/2$, and $\mathbf{x}_3 = \mathbf{x} + \mathbf{e}_z \Delta z/2$ in which $\mathbf{e}_{(\cdot)}$ denotes a unit vector in the direction defined by the subscript. During a single time-step Δt, pressure is updated at time t, velocity is updated at times $t^\pm = t \pm \Delta t/2$, and the sign of the phase term in the *k*-space operator that implements spatial shifts is changed between velocity and pressure updates to maintain the same spatial grid. These *k*-space operators result in exact time stepping in a homogeneous medium and distinguish the parallel treatments in Refs. 1 and 22.

The resulting compact first-order spatially and temporally discrete equations for pressure and particle velocity of a scattered wave propagating in a medium that has relaxation absorption and is surrounded by a perfectly matched boundary layer for the total fields are

$$v_x(\mathbf{x}_1,t^+) = e^{-\alpha_x(\mathbf{x}_1)\Delta t/2}\left\{e^{-\alpha_x(\mathbf{x}_1)\Delta t/2} v_x(\mathbf{x}_1,t^-) - \frac{\Delta t}{\rho(\mathbf{x}_1)} \frac{\partial p(\mathbf{x},t)}{\partial^{(c_0\Delta t)^+} x} - \left[\frac{\Delta t}{\rho(\mathbf{x}_1)} - \frac{\Delta t}{\rho_0}\right] \frac{\partial p_i(\mathbf{x}_1,t)}{\partial x}\right\},$$

$$v_y(\mathbf{x}_2,t^+) = e^{-\alpha_y(\mathbf{x}_2)\Delta t/2}\left\{e^{-\alpha_y(\mathbf{x}_2)\Delta t/2} v_y(\mathbf{x}_2,t^-) - \frac{\Delta t}{\rho(\mathbf{x}_2)} \frac{\partial p(\mathbf{x},t)}{\partial^{(c_0\Delta t)^+} y} - \left[\frac{\Delta t}{\rho(\mathbf{x}_2)} - \frac{\Delta t}{\rho_0}\right] \frac{\partial p_i(\mathbf{x}_2,t)}{\partial y}\right\},$$

$$v_z(\mathbf{x}_3,t^+) = e^{-\alpha_z(\mathbf{x}_3)\Delta t/2}\left\{ e^{-\alpha_z(\mathbf{x}_3)\Delta t/2} v_z(\mathbf{x}_3,t^-) - \frac{\Delta t}{\rho(\mathbf{x}_3)} \frac{\partial p(\mathbf{x},t)}{\partial^{(c_0\Delta t)^+} z} - \left[\frac{\Delta t}{\rho(\mathbf{x}_3)} - \frac{\Delta t}{\rho_0} \right] \frac{\partial p_i(\mathbf{x}_3,t)}{\partial z} \right\}, \tag{38}$$

$$p_x(\mathbf{x},t+\Delta t) = e^{-\mu_x(\mathbf{x})\Delta t} p_x(\mathbf{x},t) - \frac{\Delta t}{\kappa_\infty(\mathbf{x})} e^{-\mu_x(\mathbf{x})\Delta t/2} \left[F_x(\mathbf{x},t^+) + \frac{\partial v_x(\mathbf{x}_1,t^+)}{\partial^{(c_0\Delta t)^-} x} - \sum_\nu \eta_\nu^x(\mathbf{x}) S_\nu^x(\mathbf{x},t^+) \right],$$

$$p_y(\mathbf{x},t+\Delta t) = e^{-\mu_y(\mathbf{x})\Delta t} p_y(\mathbf{x},t) - \frac{\Delta t}{\kappa_\infty(\mathbf{x})} e^{-\mu_y(\mathbf{x})\Delta t/2} \left[F_y(\mathbf{x},t^+) + \frac{\partial v_y(\mathbf{x}_2,t^+)}{\partial^{(c_0\Delta t)^-} y} - \sum_\nu \eta_\nu^y(\mathbf{x}) S_\nu^y(\mathbf{x},t^+) \right],$$

$$p_z(\mathbf{x},t+\Delta t) = e^{-\mu_z(\mathbf{x})\Delta t} p_z(\mathbf{x},t) - \frac{\Delta t}{\kappa_\infty(\mathbf{x})} e^{-\mu_z(\mathbf{x})\Delta t/2} \left[F_z(\mathbf{x},t^+) + \frac{\partial v_z(\mathbf{x}_3,t^+)}{\partial^{(c_0\Delta t)^-} z} - \sum_\nu \eta_\nu^z(\mathbf{x}) S_\nu^z(\mathbf{x},t^+) \right], \tag{39}$$

where

$$F_x(\mathbf{x},t^+) = [e^{\mu_x(\mathbf{x})\Delta t/2}\kappa_\infty(\mathbf{x}) - \kappa_0] \frac{\partial p_{x,i}(\mathbf{x},t^+)}{\partial t} + \frac{\kappa_\infty(\mathbf{x})}{\Delta t}[e^{\mu_x(\mathbf{x})\Delta t/2} - e^{-\mu_x(\mathbf{x})\Delta t/2}] p_{x,i}(\mathbf{x},t) - \sum_\nu \eta_\nu^x(\mathbf{x}) S_\nu^{x,i}(\mathbf{x},t^+),$$

$$F_y(\mathbf{x},t^+) = [e^{\mu_y(\mathbf{x})\Delta t/2}\kappa_\infty(\mathbf{x}) - \kappa_0] \frac{\partial p_{y,i}(\mathbf{x},t^+)}{\partial t} + \frac{\kappa_\infty(\mathbf{x})}{\Delta t}[e^{\mu_y(\mathbf{x})\Delta t/2} - e^{-\mu_y(\mathbf{x})\Delta t/2}] p_{y,i}(\mathbf{x},t) - \sum_\nu \eta_\nu^y(\mathbf{x}) S_\nu^{y,i}(\mathbf{x},t^+),$$

$$F_z(\mathbf{x},t^+) = [e^{\mu_z(\mathbf{x})\Delta t/2}\kappa_\infty(\mathbf{x}) - \kappa_0] \frac{\partial p_{z,i}(\mathbf{x},t^+)}{\partial t} + \frac{\kappa_\infty(\mathbf{x})}{\Delta t}[e^{\mu_z(\mathbf{x})\Delta t/2} - e^{-\mu_z(\mathbf{x})\Delta t/2}] p_{z,i}(\mathbf{x},t) - \sum_\nu \eta_\nu^z(\mathbf{x}) S_\nu^{z,i}(\mathbf{x},t^+), \tag{40}$$

$$S_\nu^x(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^x(\mathbf{x},t^-) + \Delta t \frac{p_x(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right],$$

$$S_\nu^y(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^y(\mathbf{x},t^-) + \Delta t \frac{p_y(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right],$$

$$S_\nu^z(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^z(\mathbf{x},t^-) + \Delta t \frac{p_z(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right], \tag{41}$$

and

$$S_\nu^{x,i}(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^{x,i}(\mathbf{x},t^-) + \Delta t \frac{p_{x,i}(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right],$$

$$S_\nu^{y,i}(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^{y,i}(\mathbf{x},t^-) + \Delta t \frac{p_{y,i}(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right],$$

$$S_\nu^{z,i}(\mathbf{x},t^+) = e^{-\Delta t/[2\tau_\nu(\mathbf{x})]}\left[e^{-\Delta t/[2\tau_\nu(\mathbf{x})]} S_\nu^{z,i}(\mathbf{x},t^-) + \Delta t \frac{p_{z,i}(\mathbf{x},t)}{\tau_\nu(\mathbf{x})} \right]. \tag{42}$$

III. NUMERICAL METHODS

The methods in this paper not only have the goal of performing accurate calculations of propagation over a large scale but also have the goal of utilizing efficiently a computing resource that is often shared among many users and on which access time is typically limited. Therefore, methods that allow optimal selection of calculation parameters are important. Examination of accuracy used media with simple geometries in which exact solutions to the forward problem were available for comparison. The *k*-space method was shown in two dimensions to be more accurate than other finite-difference time-domain methods.[1] Therefore, comparisons here were only made with exact methods. The codes used in this study for numerical solution of the forward problem in three dimensions were developed concurrently by the collaborating institutions with some interaction.[b] The two codes, one that computed total pressure and one that computed scattered pressure, were shown to produce exactly the same scattering given an identical medium.

Because prospective applications of the *k*-space method are in the general area of biomedical ultrasound, accuracy and stability were evaluated with wideband plane-wave pulses characteristic of typical clinical and high-frequency biomedical ultrasound transducers. The clinical-frequency band is representative of bands used for diagnostic instrumentation and the high-frequency band is representative of bands used to study small animals. The clinical-frequency pulse was a Gaussian-windowed sinusoid with a 2.5-MHz center frequency and a −6-dB bandwidth of 1.7 MHz. The high-frequency pulse was a Gaussian-windowed sinusoid with a 40-MHz center frequency and a −6-dB bandwidth of 24 MHz.

A. Parameters of the PML

The PML allows the inhomogeneity to fill the entire computational domain by essentially eliminating reverberations but the matching region can extend computation time by increasing the size of the computational domain. Therefore, selection of PML parameters is important. In this study, the performance of the PML was evaluated in the high-frequency band as a function of the thickness of the PML and

the maximum absorption per grid point. A homogeneous medium was used with acoustic properties of water at body temperature ($c=1524$ m/s and $\rho=997$ kg/m^3). The size of the spatial step was four points per minimum wavelength (PPW). This wavelength was computed at the upper −40-dB frequency of the incident pulse and the time-step was chosen to result in a Courant–Friedrichs–Lewy (CFL) (Ref. 24) number of 0.5. Values of the CFL number reported here are based on the background sound speed while the time-step was determined by considering stability for a CFL number based on the maximum sound speed in the medium. The dispersionless PML attenuation parameters increased smoothly from the inner surface to the outer surface of the PML according to Eq. (27) in Ref. 1. The maximum absorption per grid point in the PML, $\alpha_{\max}$, was varied from 1 to 6 Np per grid point in increments of 0.5. The thickness of the PML was varied between 2 and 20 grid points in increments of 2. The pulse was propagated normally to the surface of the PML and the maximum reflection and transmission coefficients were determined as the PML thickness and absorption per PML point were varied.

B. Investigation of smoothing

Discontinuities in medium parameters, i.e., density, compressibility, and attenuation, produce, as previously noted, high spatial-frequency components in the medium spectrum and can cause aliasing when the first-order k-space operators given in Eq. (37) are evaluated. To reduce this effect, the medium parameters in some cases were lowpass filtered using a half-band filter.[16] The benefit from smoothing was evaluated by comparing scattering of a clinical-frequency pulse from a small high-contrast 1-mm diameter sphere that was half-band filtered with scattering from a sphere that was not smoothed. This size sphere was chosen because the minimum wavelength in the studied clinical temporal-frequency band associated with the highest temporal frequency is 0.3 mm; hence, the sphere is small in the sense that it has a size of the order of (or smaller than) the incident wavelength. This property along with the stated spatial sampling criteria means that about 26 spatial samples span the diameter of the sphere when the sampling increment $\Delta x=0.075$ mm. If smoothing causes the radius of the sphere to grow by just two spatial samples, a 25% increase in the sphere volume results. Since scattering from fluid spheres in the long-wavelength limit is proportional to the sphere volume,[25] smoothing small spheres will be more likely to degrade agreement with an exact calculation. Therefore, an unfiltered medium at this scale may be more desirable, making the comparison of scattering from the smoothed medium with that of the unsmoothed medium at this scale a worst-case comparison; smoothing can be expected to perform better for larger scales. The parameters of the small sphere (i.e., sound speed, density, and attenuation slope) were chosen to implement a high-contrast object because small-scale inhomogeneities in human tissues are likely to be associated with microcalcifications.[26] To arrive at the parameters, the sound speed of a lossless sphere was increased from a background $c=1524$ m/s (water at body temperature) by increments of 300 m/s until an L^2 error (i.e., the spectral norm of the scaled difference between measured and exact scattering[2]) measured relative to an exact calculation exceeded 0.03. The high-contrast sound speed was set to the value that resulted in the excess L^2 error minus 150 m/s. Next, absorption was added to the sphere and the density increased from the background 997 kg/m^3 in increments of 300 kg/m^3 until the L^2 error doubled. The high-contrast density was equal to the density closest to the point at which this occurred.

For a scattering medium such as a single sphere or collection of spheres having an analytic Fourier transform, the filtering is exact. For arbitrary media constructed by segmentation of digitized images acquired using another modality (e.g., magnetic resonance imaging), the Fourier transform must be evaluated numerically. Accuracy in each case is controlled by the sampling rate. The sampling rate appropriate for an arbitrarily constructed medium was determined in this work by increasing the amount of oversampling for the sphere with parameters described in Sec. III D until the scattering from the oversampled and half-band filtered medium agreed with scattering from the exactly half-band filtered medium. Also, scattering from a lossless high-contrast sphere ($c_1=3540$ m/s, $\rho_1=1990$ kg/m^3) in the same background with a volume equivalent to the volume of a single grid-point voxel, i.e., a so-called minimum-resolution sphere, was calculated for oversampling in multiples of 2, 4, 8, and 10.

C. Inclusion of absorption

The description of absorption caused by relaxation is for any finite number of relaxation processes. However, for each relaxation process, generalized compressibility-coefficient arrays for the medium require allocation of system memory. Thus, representation of attenuation with the minimum number of relaxation processes is desirable to model realistic absorption in tissue. Experience has shown that two processes satisfactorily model a linear dependence of absorption on frequency over the bands simulated. For clinical-frequency computations, relaxation times of 40 and 400 ns were used. For high-frequency computations, relaxation times of 2 and 20 ns were used. The long-duration relaxation time was chosen to be approximately the reciprocal of the lower −6-dB frequency of the pulse and the short-duration relaxation time was chosen to be approximately five times the maximum frequency contained in the pulse. Compressibility coefficients for a lossy background were found by a least-square-error fit of Eqs. (42) and (43) in Ref. 7 for an attenuation coefficient slope with a linear dependence on temporal frequency. Fitting was performed over a frequency range of $0<f<5$ MHz for clinical pulses and $9.1<f<70.9$ MHz for high-frequency pulses. Evaluation of relaxation absorption was accomplished by propagating high-frequency Gaussian weighted pulses in the lossy medium with $c_0=1524$ m/s, $\rho_0=993$ kg/m^3, and $\beta_0=0.5$ dB/(cm MHz) over a distance of the order of the maximum wavelength present in the pulse. Frequency-dependent attenuation derived from ratios of the pulse spectra and frequency-dependent phase velocities derived from

phase changes between the pulses at measurement points were compared with theoretical values from Ref. 7.

D. Accuracy and stability: L^2 vs CFL and PPW for a small sphere

A good compromise between CFL number and PPW was sought for large-scale calculations and comparisons with exact solutions. A low PPW is desirable to minimize spatial grid size because smaller grids required less storage and the three-dimensional FFTs can be evaluated more rapidly. Large values of CFL number are desirable because they require fewer temporal iterations to collect all non-negligible scattered waves. To determine the CFL and PPW values used for large-scale calculations, scattering from a test sphere was computed for selected values of PPW in the range 2–10 and CFL in the range 0.1–0.8. These ranges guarantee numerical stability and Nyquist temporal sampling. The sphere was 4 mm in diameter with sound speed $c_1=1460$ m/s, density $\rho_1=970$ kg/m^3, and attenuation slope $\beta_1=0.5$ dB/(cm MHz) in a lossless background with sound speed $c_0=1509$ m/s and density $\rho_0=997$ kg/m^3. The sphere was centered in the medium and excited by a clinical plane-wave pulse traveling in the $+\mathbf{e}_x$ direction. Rotational symmetry about the axis defined by the incident wave vector permits the L^2 error of the scattering calculation to be found from the time-domain waveforms observed at points on a ring situated in the $z=0$ plane 2.5 mm from the center of the sphere. These time-domain waveforms were examined to find an L^2 error that produced agreement with an exact solution. The L^2 errors from the set of calculations along with sampling constraints and medium representation considerations were used to guide selection of PPW (Δx) and CFL (Δt) for large-scale comparisons.

E. Impact of inhomogeneity scale

Using a fixed PPW and CFL number determined with methods just described, the impact of increasing the scale of the calculations was studied by scattering clinical pulses from each of two sets of centered spheres in a lossless background with sound speed $c_0=1509$ m/s and density $\rho_0=997$ kg/m^3. Both sets had radii ranging from 1 to 5 mm, which is the range of radii of spheres to be included in a large compound object discussed later. In one set, the spheres were lossless with sound speed $c_1=1570$ m/s and density $\rho_1=970$ kg/m^3. In the other set, the spheres were lossy with $\beta_1=0.5$ dB/(cm MHz). The L^2-error variation as a function of radius was tabulated.

Two more sets of spheres (0.75–4.55 mm) with "full contrast," meaning lossy, i.e., $\beta_1=0.5$ dB/(cm MHz), with higher sound-speed contrast $c_1=1650$ m/s and density $\rho_1=970$ kg/m^3 in a lossless background with sound speed $c_0=1509$ m/s and density $\rho_0=997$ kg/m^3 were excited by a clinical pulse using PPW=4 and CFL$\approx$0.4. The L^2 error was calculated. Next, the same numerical experiment was performed at half the CFL number to demonstrate how accuracy was recovered at larger scales. These full contrast spheres were used to construct a compound scattering object described next.

TABLE I. Properties of background and spheres for large-scale compound medium.

	Physical parameter		
Medium	Sound speed (mm/μs)	Density (g/cc)	Absorption [dB/(cm MHz)]
Background	1.509	997	0.0
Large sphere	1.570	970	0.3
Smaller spheres	1.650	970	0.5

F. Large-scale compound object

A large-scale inhomogeneity with tissue-like properties[27,28] was constructed by randomly positioning 12 non-overlapping spheres within a 24-mm diameter sphere. Calculation of scattering on such a scale required accurate wave propagation over nearly 300 wavelengths at the maximum temporal frequency contained in the pulse. The enclosed spheres all had the same acoustic parameters that differed from the acoustic parameters of the enclosing sphere. These acoustic parameters are reported in Table I. The radii and positions of the spheres are reported in Table II. Scattering was calculated from each centered unique sphere (eight different radii in all) as well as the compound sphere. The same data were collected from exact computations and the L^2 errors were tabulated. All scattering used four PPW and a CFL number approximately equal to 0.2 because all inhomogeneities were lossy and accurate large-scale calculation was the objective. For the compound medium, a comparison with an exact solution at a single temporal frequency ($f=2.5$ MHz) was made at 2943 points 14 mm from the center of the object by calculating the normalized root-mean-square error in the scattered-pressure magnitude.

IV. NUMERICAL RESULTS

A. PML parameters

The reflection and transmission coefficients for the PML parameter study are presented in Fig. 3 as functions of PML

TABLE II. Positions and radii of spheres forming the large-scale compound medium.

	Position		
Radius (mm)	x (mm)	y (mm)	z (mm)
12.00	0.00	0.00	0.00
4.55	−6.25	−2.50	−2.5768
3.90	−5.00	5.00	3.60
3.90	2.50	6.00	−3.60
2.50	7.00	1.00	2.00
2.50	2.50	−5.00	−2.00
2.50	7.00	6.00	1.50
2.50	−2.50	−9.00	−1.50
2.00	0.00	0.00	0.00
1.50	8.50	−3.50	0.00
1.25	3.75	−9.00	−1.00
1.25	−9.00	2.00	1.00
0.75	−2.00	10.00	0.00

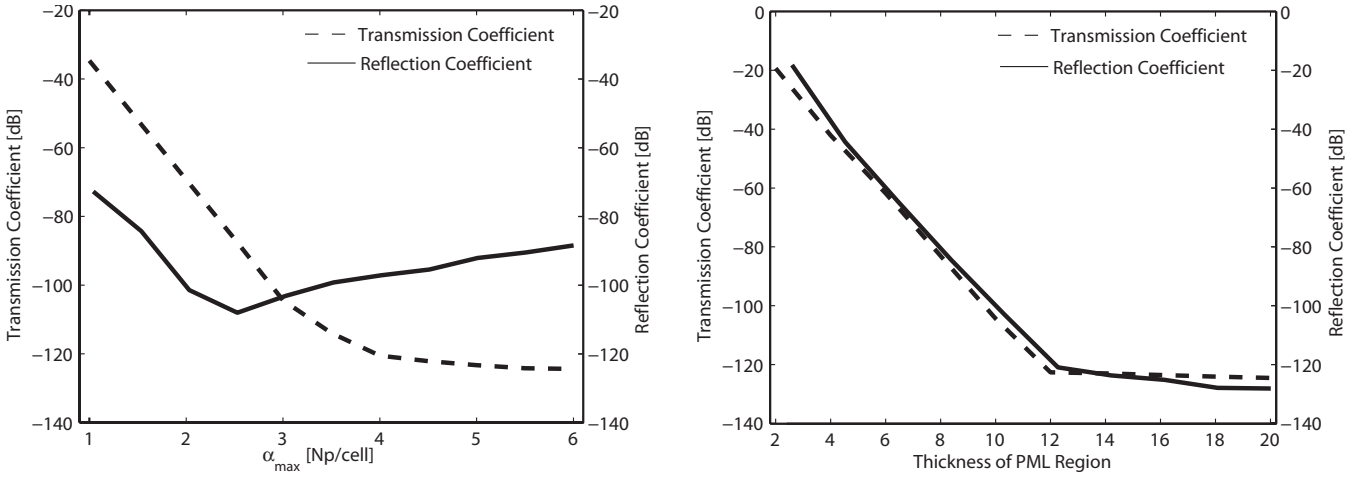


FIG. 3. Left panel: Behavior of a PML as the maximum attenuation per cell is varied with the number of PML layers fixed at 10. Right panel: Behavior of a PML as the number of PML layers is varied and the maximum attenuation per PML cell is fixed at 3 Np.

thickness and maximum absorption per PML point. The right panel of Fig. 3 shows reflection and transmission coefficients vs the number of PML points when the maximum absorption per PML cell is 3 Np. The reflection and transmission coefficients both decrease as the thickness of the PML region increases but the marginal benefit of including more than 12 layers is small. Reflection and transmission coefficients are plotted as functions of maximum absorption per PML point in the left panel of the figure with PML thickness fixed at ten points. The transmission coefficient decreases monotonically as maximum absorption increases up to 6 Np per PML point but the rate of decrease is small once the maximum absorption exceeds 3.5 Np/point. In contrast, the reflection coefficient reaches a minimum at 2.5 Np per PML point. Similar behavior of the PML reflection coefficient as a function of maximum absorption per point was demonstrated in the original electromagnetic computations by Berenger.[6] Figure 3 indicates that PML transmission and reflection coefficients less than −100-dB can be achieved simultaneously by setting the PML thickness to ten points and the maximum absorption to 3 Np per PML point. These PML parameters were employed in the simulations presented in this paper.

B. Medium smoothing

A worst-case comparison of scattering from a filtered and unfiltered high-contrast sphere with parameters $c_1 = 2874$ m/s, $\rho_1 = 2497$ kg/m^3, and $\beta_1 = 0.5$ dB/(cm MHz) resulted in L^2 errors of 0.054 and 0.060, respectively. The scattered waveforms are shown in Fig. 4. Little difference exists between the waveforms for the two cases. This is consistent with the approximately equal errors except that the scattered waves in the unfiltered medium have a low-level background artifact typical for scattering from unfiltered media.

For a weak-scattering 2-mm radius sphere, oversampling a space-domain constructed medium by a factor of 2 using a minimum integer multiple for a half-band filter practically eliminates low-level artifacts such as the kind seen in Fig. 4. The L^2 error was insensitive to the amount of oversampling. For the high-contrast so-called minimum-resolution sphere, oversampling improves agreement between scattering from the medium constructed exactly in the spatial-frequency domain with half-band filtering and the medium constructed by oversampling in the space domain followed by half-band filtering. The L^2 error is, however, lower for scattering computed from an unfiltered medium. This is attributed to increasing the effective volume of the scatterer.

In all cases, spatial filtering exactly or spatial filtering an oversampled domain by at least a factor of 2 practically eliminates artifact due to medium discontinuities. For small (grid size and smaller) inhomogeneities, spatial filtering broadens the spatial extent of the inhomogeneity and in-

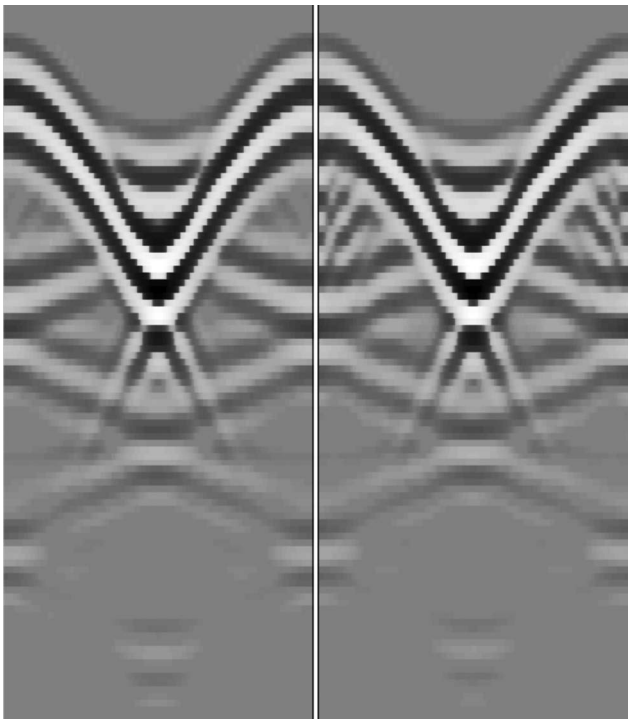


FIG. 4. The scattered waves from a filtered (left) and unfiltered (right) 1-mm radius sphere. The horizontal axes in each panel span 360° in the $z = 0$ plane with measurements taken at a distance of 3 mm from the sphere center and waveforms are shown on a bipolar logarithmic scale over a ±60 dB range. The vertical axes span approximately 4.5 μs.

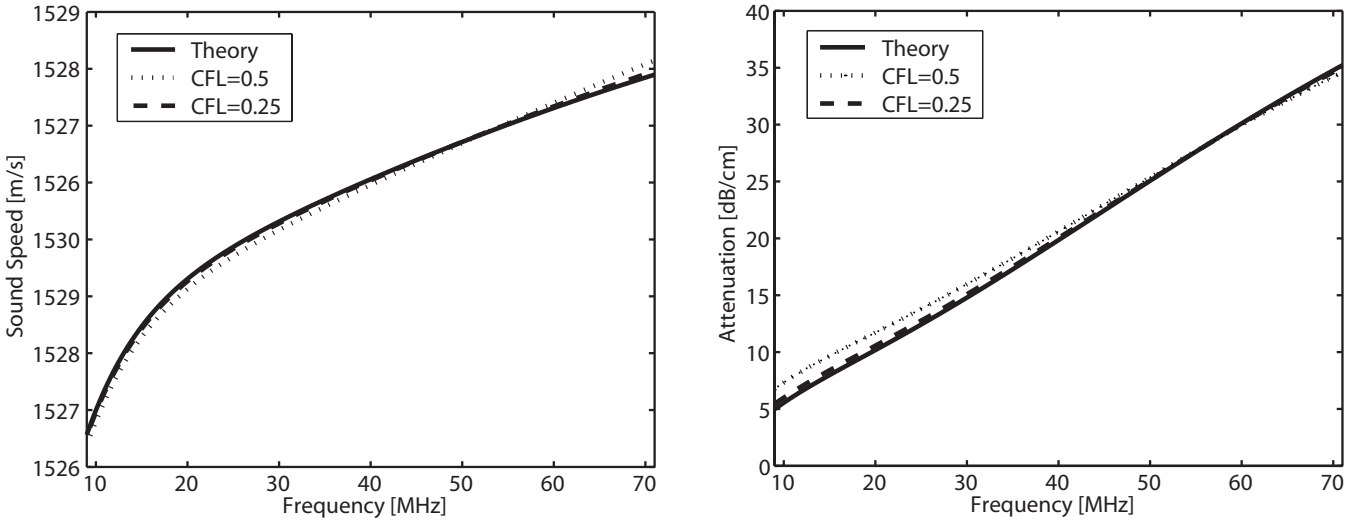


FIG. 5. Temporal-frequency dependence of sound speed (left) and attenuation (right) in the high-frequency band.

creases error. For large-scale high-contrast inhomogeneities, such as a 2-mm sphere with the properties of a calcification, half-band filtering has the effect of slightly decreasing the overall L^2 error because some loss of accuracy for scattering in the forward direction is compensated by a greater improvement in backscatter accuracy. Backscattering from the filtered medium agrees better with an exact solution than the unfiltered medium. Another possible approach to numerical evaluation of smoothing is to compare errors in regions where the exact solution is quiescent. Artifact from unfiltered media increases errors in these regions and shows the need for filtering. Additionally, the energy level of the artifacts can be compared to energy levels expected from scattering by minimum-resolution (single-pixel) objects to determine if artifacts are larger than the actual scattering in the medium.

C. Absorption

The right panel of Fig. 5 illustrates the ability of the first-order k-space method with second-order relaxation absorption to model an approximately linear frequency dependence of attenuation with $\beta=0.5$ dB/(cm MHz) in the high-frequency band. In this figure, the theoretical second-order relaxation absorption model is compared to numerical results using CFL numbers of 0.5 and 0.25. The two-process model produces small deviations from a linear frequency dependence. With a CFL of 0.5, the k-space method produces attenuation within a few dB/cm of the theoretical curve at all frequencies in the −6-dB bandwidth of the high-frequency pulse. Reducing the CFL number to 0.25 yields numerical attenuation very close to the theoretical curve. Fitting a linear dependence of attenuation on frequency produces a frequency dependence for sound speed shown in the left panel of Fig. 5. A comparison of numerical results to theoretical results shows good agreement for a CFL of 0.5 and excellent agreement for a CFL of 0.25. Analogous results are obtained at clinical frequencies.

D. Accuracy vs PPW and CFL for a small sphere

Scattering from a small sphere for a combination of grid sizes and time-steps yields the L^2 errors presented in Fig. 6. The results show that, as the grid spacing is decreased, smaller time-steps associated with lower CFL numbers are required to maintain accuracy. For these scattering experiments, the time-domain waveforms indicate that an L^2 error of approximately 0.02 or smaller is associated with excellent agreement between an exact solution and the k-space solution. Therefore, $L^2=0.02$ was used to guide selection of parameters for large-scale calculations.

A primary consideration for large-scale calculations is execution time. Execution-time results, presented in the Appendix, show that execution time is proportional to the number of time-steps and a power of the grid size. Therefore, reducing the grid size to the minimum acceptable value is desirable to reduce the computing system requirements. Nyquist sampling of a monochromatic wave at the highest temporal frequency of the clinical pulse would suggest Δx

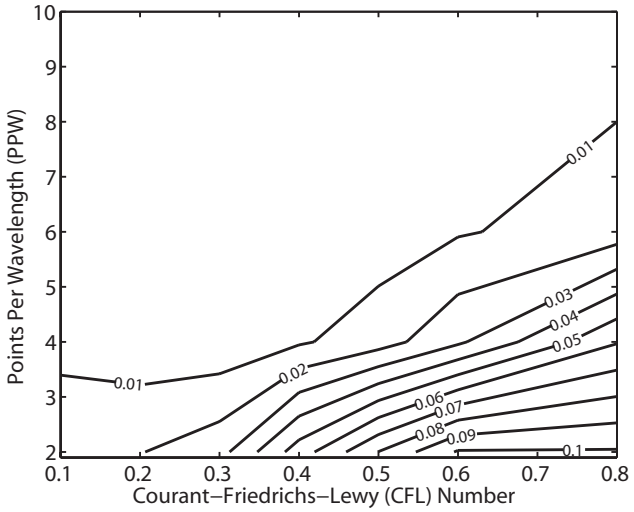


FIG. 6. L^2 error resulting from scattering of clinical pulse waveforms by a 4-mm diameter sphere.

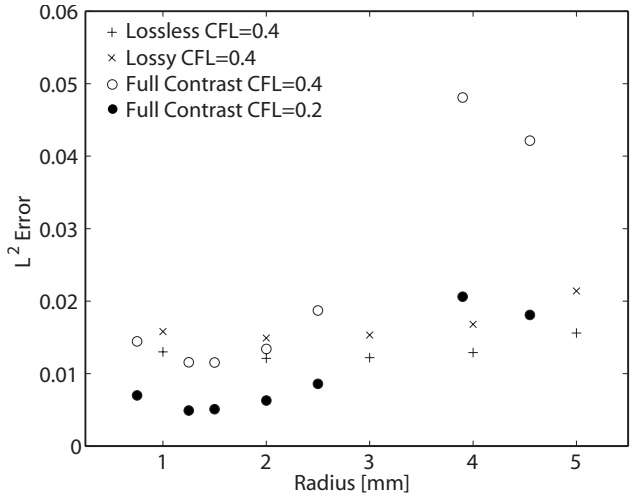


FIG. 7. Scattering from sets of spheres with progressively larger radii.

$=0.15$ mm (PPW$=2$) for large-scale calculations. However, because previously derived[2] stability criteria ignored density contrast that is included in the models used in the present study and no calculation for PPW$=2$ produced an L^2 error less than 0.02, the spatial sampling rate was doubled. An added benefit of doubling the spatial sampling rate is that discontinuities in the medium can be more accurately approximated and smaller inhomogeneities resolved. Given these considerations, a grid spacing of $\Delta x=0.075$ mm was selected for large-scale calculations of scattering of clinical pulses by spheres, and collections of spheres. With the selected grid spacing and given that $L^2<0.02$ produces accurate agreement with exact solutions, a time-step of $\Delta t=0.02$ μs (CFL≈ 0.55) is considered to be an upper bound for large-scale calculations.

E. Scaling results

Numerical scattering experiments with the two sets of spheres described in Sec. III E using prospective values for time-step and spatial grid size and using methods described in Sec. III D show that loss of accuracy can occur in both lossless and lossy media. The L^2 error as a function of scale size is plotted in Fig. 7 and shows that, as the sphere radius is varied from 1 to 5 mm, the accuracy of lossless scattering is weakly dependent on the scale size while lossy scattering shows a stronger dependence. As the size of the sphere is increased above 2 mm, accuracy decreases.

Also shown in Fig. 7 are L^2 errors in calculations of scattering from the inclusions in the large-scale compound object. Two different time-steps that result in CFL 0.4 and CFL 0.2 show that, at larger scales, discrepancies between the exact and k-space dispersion relationships enhance deviations from the exact solution. The L^2 errors associated with scattering at CFL 0.2 from the inclusions, the enclosing sphere, and the large-scale compound object are reported in Table III. Calculations for the large sphere and compound sphere were performed with a CFL number of 0.2 to maintain good agreement at large scales. Since error in the compound object is comparable to error observed in the inclusions, the error appears to be independent of the configuration of the scattering medium.

TABLE III. L^2 errors of propagation calculations for each radius sphere in the large-scale compound medium.

Radius (mm)	L^2 error
0.75	0.007
1.25	0.005
1.50	0.005
2.00	0.006
2.50	0.009
3.90	0.021
4.55	0.018
12.00	0.021
12.00[a]	0.025[b]

[a]Compound object.
[b]Single temporal frequency.

Snapshots of the scattered pressure in the $x,y,z=0$ planes at representative instants of time for the large-scale compound object are shown in Fig. 8. In this calculation, the incident wave (a clinical pulse) propagates in the $+\mathbf{e}_x$ direction and is offset at $x=-16$ mm at time $t=0$. Therefore, scattering is shown in the $y,z=0$ planes as the incident wave penetrates the sphere but no scattering in the $x=0$ plane occurs until the incident wave arrives at the center of the sphere. The entire video sequence (MPEG-2) is available at the webpage found in Ref. 29. The time histories allow appreciation of scattering into a plane from out-of-plane inhomogeneities.

V. DISCUSSION

A. Solutions for total pressure and scattered pressure

When the first-order wave equations are written in terms of the total pressure, the medium can be passively excited by specifying a propagating wave in the medium at time $t=0$. For this case, the medium must contain the incident wave as well as the inhomogeneities. The total-pressure forward-problem solution is, therefore, a natural choice for pulse-echo applications in which the initial position of the wave corresponds to receiver positions that must also be within the medium to sample spatially limited transmit and receive apertures. The total-pressure forward-problem solution can also be used to excite incident waves of infinite spatial extent. Such waves are created with wave vectors parallel to a Cartesian axis because reflections at the PML are produced for all incident directions that are not collinear with one of the Cartesian coordinate system axes. The angle of incidence can, however, be adjusted by rotating the inhomogeneity. Single-frequency and pulsatile incident plane waves can be implemented by specifying the values of pressure and velocity on the boundary of the grid. Such an excitation requires a time-domain calculation of the incident wave at every boundary point and values of the incident field are specified on the boundary at every time-step of the calculation. The incident pressure for the scattered-pressure form of the first-

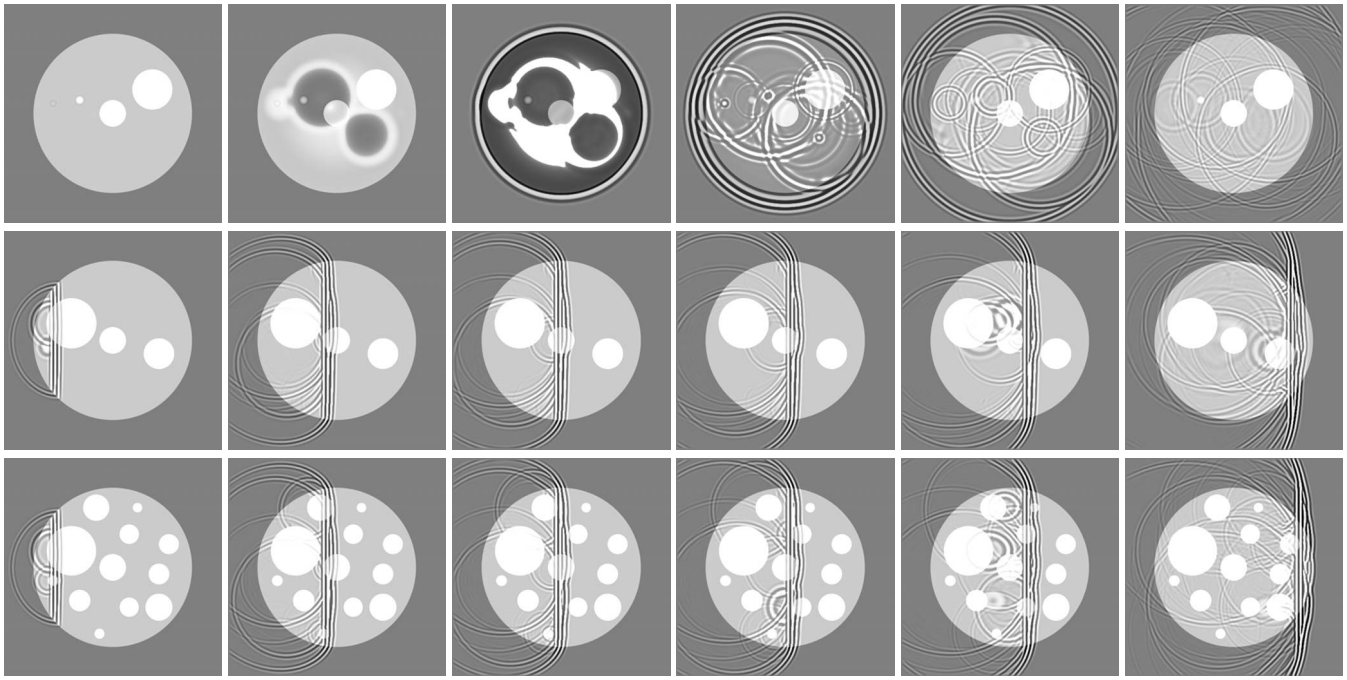


FIG. 8. Scattering from a 24-mm diameter compound sphere with properties similar to those of human breast in a background of water is shown in three orthogonal planes. At the top are time frames recorded at $x=0$. In the middle are time frames recorded at $y=0$. At the bottom are time frames recorded at $z=0$. The six frames are recorded at times 4.8, 9.6, 10.4, 11.2, 12.4, and 17.2 μs. The incident illumination is a plane-wave pulse offset at $x=-16$ mm at time $t=0$ with center frequency of 2.5 MHz and a Gaussian envelope with a -3 dB bandwidth of 1.7 MHz. The medium sound speed is overlayed on the pressure field with brighter areas corresponding to higher sound speed.

order equations is, in contrast, always active. This allows implementation of single-frequency and pulsatile plane waves from arbitrary directions. The incident wave is, however, needed not just on the boundary but also at each grid point. This added burden is not significant for simple incident waves but can add computational cost and storage for more complex incident waves such as focused beams.

B. Scaling and accuracy

Figure 7 suggests that large-scale media degrade accuracy. Why this occurs can be understood with the help of Fig. 5 that compares the *k*-space dispersion relationships with the exact theoretical values. For theoretical calculations, the dispersion relationships are exact. For *k*-space method calculations, the actual dispersion relationships present in the *k*-space results can, however, deviate from the exact theoretical values. The agreement between *k*-space and exact dispersion relationships is best for lower CFL numbers. As the scale of the inhomogeneity increases, the discrepancy between *k*-space and exact calculations increases because scattered waves accumulate. For example, absorption causes spectral components of a pulse to suffer amplitude reductions that accumulate as the propagation distance in the lossy medium increases. In general, calculation of scattering in real tissues[30] accurately using the *k*-space method requires close agreement between the dispersion relationships.

A more rigorous method for selection of relaxation times is to search for relaxation times of a specific order that minimize the mean square error between a linear frequency dependence of attenuation and the resulting theoretical curve. If this is done for the clinical-frequency range, relaxation times of 29 and 258 ns result. Although these are not the values chosen here, the comparisons with exact methods remain valid because exact codes used theoretical values based on the same relaxation times used in the *k*-space code. Two relaxation processes are justified (as opposed to three or more) because the rms error measured relative to the attenuation at 2.5 MHz (clinical center frequency) is only 1.3%. Inclusion of a third process only reduces the rms error measured relative to the attenuation at 2.5 MHz to 0.8%.

VI. CONCLUSIONS

The three-dimensional extension of the *k*-space code calculates scattering with an accuracy that is similar to that achieved with the two-dimensional code. Investigations with the two-dimensional code did not, however, include comparisons with exact scattering for lossy inhomogeneities. Therefore, because the *k*-space method is exact for lossless, homogeneous media, the departure of *k*-space calculated scattering from exact solutions as scale size increases was not previously observed. This behavior underscores the importance of using smaller CFL numbers to maintain accuracy for large-scale calculations that include loss when comparisons are made with results from exact methods. Similarly, if accurate comparisons are to be made with realistic models of tissue, then the distribution of *k*-space relaxation times must be chosen to model closely the dispersion relationships of the tissue to maximize agreement. The results of this paper show that calculation of propagation of ultrasound over hundreds of wavelengths is made feasible by the three-dimensional implementation of the *k*-space method on a cluster of computers such as those described. Increasing availability of suit-

able computing platforms will allow investigation, through simulation, of *b*-mode imaging and other applications such as ultrasound beam aberration caused by inhomogeneities in tissue. Furthermore, inverse problem solutions require a solution to the forward problem and the described method is a realistic approach to solving the forward problem for arbitrary configurations of tissue.

ACKNOWLEDGMENTS

David P. Duncan is thanked for discussions and code development. Jeffrey P. Astheimer and Jing Jin are thanked for participation in reviews throughout the duration of this work. Andrew J. Hesford is thanked for discussions during the final stages of the reported studies. Initial development of code used resources at Research Computing and the Department of Electrical Engineering, Multi-agent Bio-Robotics Lab, at the Rochester Institute of Technology. Gurcharan S. Khanna and Ferat Sahin are thanked for granting access to these resources. This research also used the SHARCNET facilities. Ge Baolai is thanked for technical assistance in the use of SHARCNET. Subsequent code development and computations used resources of the National Energy Research Scientific Computing Center (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Consulting staff at NERSC are thanked for help compiling benchmark test programs on the Cray XT4 (Franklin). M.I.D. is a Scholar of the Canadian Institutes of Health Research/University of Western Ontario Strategic Training Initiative in Cancer Research and Technology Transfer and holds a NSERC PGS-D scholarship. This research was funded in part by NIH Grant Nos. HL 50855, CA 74050, and EB 00280, Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Nos. 261323-03 and 261323-07, the Canadian Institutes of Health Research/University of Western Ontario Strategic Training Initiative in Cancer Research and Technology Transfer, and the University of Rochester Diagnostic Ultrasound Research Laboratory Industrial Associates.

APPENDIX: PARALLEL PERFORMANCE OF FFTW

Calculations were performed on two different clusters of computers. Performance results were compiled for both systems for comparison. The three-dimensional distributed FFT was the computational bottleneck so the execution time of a single three-dimensional FFT was analyzed on both systems as functions of the size of the computational grid and number of nodes using test programs that are distributed with the FFTW software. The grid size was varied from 32 to 2048 points in multiples of 2 simultaneously along each spatial dimension. The number of nodes was varied from 2 to 2048 in multiples of 2. The mean execution time for ten single-transform trials for both forward and backward transforms was noted for each combination of cluster, grid size, and number of nodes. The first cluster, Franklin (http://www.nersc.gov/nusers/systems/franklin), a Cray XT4 system at Lawrence Livermore National Laboratory, used nodes with 2.6 GHz dual-core AMD Opteron processors, 4 Gbytes of memory, and SeaStar2 (Refs. 31 and 32) interconnects

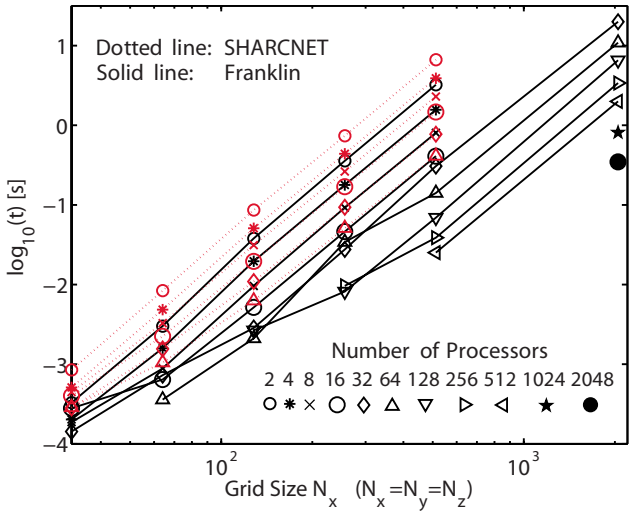


FIG. 9. (Color online) Execution times for the three-dimensional FFT on the Franklin and SHARCNET clusters.

running a Linux operating system. The second cluster, SHARCNET (http://www.sharcnet.ca), the Shared Hierarchical Academic Research Computing Network spread among several institutions in Ontario, used nodes with Opteron single-core 2.4-GHz processor, 32 Gbyte memory, and Quadrics Elan4 interconnects running an HP XC 3.1 Linux-based system. Since Franklin has many more compute nodes than SHARCNET, performance evaluations for some permutations in the set of possible grid sizes and nodes were not compiled on SHARCNET. The set of permutations of grid sizes and nodes selected ensured that adequate memory existed on each compute node to allocate arrays and minimize the probability of a node using a disk memory cache during the calculation.

A comparison of the execution time for a single distributed three-dimensional FFT on the two different computer clusters used to produce results in this paper is shown in Fig. 9. Runtimes on both clusters have scaling similarly with grid size for a fixed node count. The results show SHARCNET as slightly better scaling. However, for a fixed number of nodes, Franklin always performs the transform in less than half the time of SHARCNET. In all cases, for a fixed grid size, SHARCNET performs better by using more nodes. For Franklin, this is not the case. For some grid sizes, increasing the node count lengthens the execution time.

The difference in the slope of the execution-time lines between Franklin and SHARCNET for a fixed number of processors (two, for example) merits comment. The apparent lower slope of execution time on SHARCNET is attributed to node differences such as per-core memory. Faster execution on Franklin than on SHARCNET for any combination of nodes and grid size is attributed to Franklin's SeaStar2 high-performance, low-latency interconnects being available because the distributed three-dimensional FFT requires an all-to-all communication prior to performing the last set of two-dimensional transforms for the slab on each node. Execution time on Franklin can be different than on SHARCNET because performing a transform of fixed size with

fewer nodes is sometimes faster. The grid size at which this behavior occurs is proportional to the number of nodes performing the transform.

For a fixed number of processors, the time required to execute a single three-dimensional FFT scales as $O(N^3 \log N)$ for N^3 grid points. The benefit of using a cluster of computers results from increasing the number of nodes used to calculate the three-dimensional FFT as the grid size is increased. If the number of nodes is chosen to be equal to one dimension of a cubic medium, then the scaling exponent is reduced from 3 to 1.7. This benefit along with capabilities to calculate scattering in a large-scale medium that would otherwise be intractable because of the aggregate memory required for storage is an advantage of the parallel *k*-space method.

[a]In Eqs. (17) and (18), and in following analysis, the notation uses the quantum mechanics convention in which a symbol represents a conceptual object and the argument indicates the domain in which an object is evaluated. Thus, $p(\mathbf{x},t)$ is the scattered pressure in the space domain as a function of time t and $p(\mathbf{x},\omega)$ is the temporal Fourier transform as a function of temporal-frequency ω.

[b]Both institutions benefited from a previously developed two-dimensional algorithm from the University of Rochester and some initial exchanges concerning architecture but otherwise three-dimensional extensions of the code proceeded independently.

[1]M. Tabei, T. D. Mast, and R. C. Waag, "A *k*-space method for coupled first-order acoustic propagation equations," J. Acoust. Soc. Am. **111**, 53–63 (2002).

[2]T. D. Mast, L. P. Souriau, D.-L. D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, "A *k*-space method for large-scale models of wave propagation in tissues," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 341–354 (2001).

[3]B. Compani-Tabrizi, "*k*-space formulation of the absorptive full fluid elastic scalar wave equation in the time domain," J. Acoust. Soc. Am. **79**, 901–905 (1986).

[4]S. Finette, "Computational methods for simulating ultrasonic scattering in soft tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **34**, 283–292 (1987).

[5]S. Pourjavid and O. J. Tretiak, "Numerical solution of the direct scattering problem through the transformed acoustical wave equation," J. Acoust. Soc. Am. **91**, 639–645 (1992).

[6]J.-P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).

[7]A. I. Nachman, J. 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).

[8]N. N. Bojarski, "The *k*-space formulation of the scattering problem in the time domain," J. Acoust. Soc. Am. **72**, 570–584 (1982).

[9]N. N. Bojarski, "The *k*-space formulation of the scattering problem in the time domain: An improved single propagator formulation," J. Acoust. Soc. Am. **77**, 826–831 (1985).

[10]M. I. Daoud and J. C. Lacefield, "Distributed three-dimensional simulation of B-mode ultrasound imaging using a first-order k-space method," Phys. Med. Biol. (accepted for publication).

[11]T. L. Sterling, J. Salmon, D. J. Becker, and D. F. Savarese, *How to Build a Beowulf* (MIT, Cambridge, MA, 1999).

[12]M. Frigo and S. G. Johnson, "The design and implementation of FFTW3," Proc. IEEE **93**, 216–231 (2005).

[13]O. M. Al-Bataineh, T. D. Mast, E. Park, V. W. Sparrow, R. M. Keolian, and N. B. Smith, "Utilization of the *k*-space method in the design of a ferroelectric hyperthermia phased array," Ferroelectrics **331**, 103–120 (2006).

[14]F. Jensen, W. Kuperman, M. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).

[15]B. T. Cox and P. C. Beard, "Fast calculation of pulsed photoacoustic fields in fluids using *k*-space methods," J. Acoust. Soc. Am. **117**, 3616–3627 (2005).

[16]R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1983), pp. 155–157.

[17]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 6.

[18]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Woodbury, NY, 1989), Chap. 1.

[19]J. Meixner, "Absorption and dispersion of sound in gases with chemically reacting and excitable components," Ann. Phys. 5(**43**), 470–487 (1943).

[20]S. R. de Groot and P. Mazur, *Non-Equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962), Chap. 2.

[21]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Woodbury, NY, 1989), Chap. 10.

[22]X. Yuan, D. Borup, J. W. Wiskin, M. Berggren, and S. A. Johnson, "Simulation of acoustic wave propagation in dispersive media with relaxation losses by using FDTD method with PML absorbing boundary condition," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 14–23 (1999).

[23]X. Yuan, D. Borup, J. W. Wiskin, M. Berggren, R. Eidens, and S. A. Johnson, "Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation of acoustic scattering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 816–822 (1997).

[24]E. Turkel, "On the practical use of high-order methods for hyperbolic systems," J. Comput. Phys. **35**, 319–340 (1980).

[25]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968), Chap. 8.

[26]A. Thomas Stavros, *Breast Ultrasound* (Lippincott, Williams, and Wilkins, Philadelphia, PA, 2004), Chap. 12.

[27]F. T. D'Astous and F. S. Foster, "Frequency dependence of ultrasound attenuation and backscatter in breast tissue," Ultrasound Med. Biol. **12**, 795–808 (1986).

[28]Weiwad, W., Heinig, A., Goetz, L., Hartmann, H., Lampe, D., Buchmann, J., Millner, R., Spiekman, R. P., and Heywang-Köbrunner, S. H. "Direct measurement of sound velocity in various specimens of breast tissue," Invest. Radiol. **35**, 721–726 (2000).

[29]J. C. Tillett, "Scattering in three dimensions," http://www.ece.rochester.edu/projects/ultrasoundres/3dScattering.mpg, MPEG-2 Video (Last viewed 5/5/2009).

[30]A. Francis, *Duck, Physical Properties of Tissue* (Academic, San Diego, CA, 1990), Chap. 4.

[31]R. Alverson, "Red storm," an invited talk, Hot Chips, 15 August (2003); http://www.hotchips.org/archives/hc15/2_Mon/1.cray.pdf, Cray Red Storm system overview (Last viewed 12/31/2008).

[32]R. Brightwell, K. T. Pedretti, K. D. Underwood, and T. Hudson, "SeaStar interconnect: Balanced bandwidth for scalable performance," IEEE MICRO **26**, 41–57 (2006).

Nearfield acoustic holography using a laser vibrometer and a light membrane

Quentin Leclère[a] and Bernard Laulagnet
Laboratoire Vibrations Acoustique, INSA Lyon, F-69621 Villeurbanne Cedex, France

(Received 23 December 2008; revised 23 June 2009; accepted 23 June 2009)

This paper deals with a measurement technique for planar nearfield acoustic holography (NAH) applications. The idea is to use a light tensionless membrane as a normal acoustic velocity sensor, whose response is measured by using a laser vibrometer. The main technical difficulty is that the used membrane must be optically reflective but acoustically transparent. The latter condition cannot be fully satisfied because of the membrane mass, which has to be minimized to reduce acoustic reflections. A mass correction operator is proposed in this work, based on a two-dimensional discrete Fourier transform of the membrane velocity field. An academic planar NAH experiment is finally reported, illustrating qualitatively and quantitatively the feasibility of the method and the pertinence of the mass correction operator. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3180132]

PACS number(s): 43.60.Sx [OAS] Pages: 1245–1249

I. INTRODUCTION

Nearfield acoustic holography (NAH) has been developed in the early 1980s.[1] The principle of this technique is to measure the acoustic pressure on a surface in the nearfield of an acoustic source in order to get the sound pressure or acoustic velocity on the source surface using backpropagation formulations. The first surface geometry studied for acoustic holography is the plane,[1] but NAH has been extended to cylindrical and spherical coordinates,[2] then to arbitrarily shaped surfaces.[3] Techniques used to improve planar NAH are still currently under study, and concerns either the regularization approach,[4] necessary to get solutions with a physical sense, or methods used to lower errors induced by the use of discrete Fourier transform (DFT) decompositions.[5]

Initially based on acoustic pressure measurements, NAH has been recently formulated as based on particle velocity measurements,[6] using acoustic velocity probes.[7] The component of the acoustic velocity normal to the measurement plane is measured in place of the acoustic pressure, leading to normal acoustic velocity holograms. This approach is very interesting if the aim is to identify the velocity of the source, because velocity-to-velocity NAH is theoretically more robust than pressure to velocity NAH.

The aim of this paper is to study the possibility to use an alternative measurement device for velocity-to-velocity NAH. The basic idea is to place a membrane in the hologram plane. The membrane velocity is equal for continuity reasons to the normal acoustic velocity, and if the membrane response is measured using a laser vibrometer, it means that the acoustic velocity is directly estimated. This acoustic velocity measurement technique has been examined in a previous paper,[8] and has been investigated in the literature for underwater acoustics and underwater ultrasonic applications.[9–11] The major difficulty in air is that the approach is very intrusive: the acoustic fields are modified by the membrane. The solution to minimize these modifications is to use a tensionless membrane as light as possible, so as to obtain an acoustically transparent membrane. However, even with an ultra-light membrane, the mass effect cannot be completely avoided at high frequency. For example, the velocity of a 30 g/sq m membrane excited by a plane wave in normal incidence is attenuated by 10% (−0.5 dB) at 1500 Hz, and 50% (−3 dB) at 4500 Hz. A mass correction has been proposed for plane waves.[8] This correction is extended in this paper to any kind of wave using a DFT of the acoustic velocity field.

Section II concerns theoretical formulations of pressure-to-velocity and velocity-to-velocity planar NAH. Section III is devoted to the mass correction expression for the tensionless membrane in an anechoic environment. Then an experiment is reported, showing the feasibility of the approach and the pertinence of the mass correction operator.

II. THEORETICAL FORMULATIONS

A. Basics

Planar NAH is based on a two-dimensional DFT of the acoustic pressure measured on a planar surface. At a given frequency ω, this expansion can be written as follows:

$$P(x,y,z,\omega)=\sum_n\sum_m P_{nm}(z,\omega)e^{jk_{nx}x}e^{jk_{my}y}, \tag{1}$$

z being the direction normal to the measurement plane, with $k_{nx}=2\pi n/L_x$, $k_{my}=2\pi m/L_y$, and n, m positive or negative integers varying between limits defined by the spatial resolution.

Using expansion (1) into the Helmholtz equation,

$$\Delta p+(\omega/c)^2p=0, \tag{2}$$

and solving the differential equation leads to the following solution form:

$$P_{nm}(z,\omega)=P^+_{nm}(0,\omega)e^{jk_{nmz}z}+P^-_{nm}(0,\omega)e^{-jk_{nmz}z}, \tag{3}$$

with $k^2_{nmz}=(\omega/c)^2-k^2_{nx}-k^2_{my}$.

[a]Author to whom correspondence should be addressed. Electronic mail: quentin.leclere@insa-lyon.fr

If $k_{nx}^2+k_{my}^2>(\omega/c)^2$, then $k_{nmz}=j\sqrt{-k_{nmz}^2}$ is imaginary and waves are evanescent, either increasing or decreasing exponentially with z for respectively P^- and P^+. Assuming that the source is in the part of the space with z values lower than the measurement plane, acoustic waves cannot grow exponentially with z (Sommerfeld condition). Thus, $P_{nm}^-(0,\omega)=0$.

If $k_{nx}^2+k_{my}^2<(\omega/c)^2$, then k_{nmz} is real and waves are propagating. The propagation direction of the two components in Eq. (3) depends on the convention used for the harmonic time dependence. Assuming $p(\omega,t)=P(\omega)e^{-j\omega t}$, waves traveling along the positive z direction are represented by P^+. Thus, $P_{nm}^-(0,\omega)=0$.

Fourier components of the measured hologram can finally be expressed for either the propagating or evanescent part by the following equation:

$$P_{nm}(z,\omega)=P_{nm}(0,\omega)e^{jk_{nmz}z}. \tag{4}$$

This is the expression of pressure to pressure NAH; it allows the reconstruction of the acoustic field for any z value from the DFT components of the hologram $P_{nm}(0,\omega)$ measured at $z=0$. For instance, if the measurement plane is at a distance d from a planar acoustic source, the acoustic information in the source plane is assessed for $z=-d$.

It is also possible to obtain a pressure to velocity NAH expression. The (n,m) component of the normal acoustic velocity field in any plane parallel to the measurement surface can be deduced from measured pressures using the Euler equation:

$$V_{nm}^z(z,\omega)=\frac{-j}{\omega\rho}\frac{\partial}{\partial z}P_{nm}(z,\omega)=\frac{k_{nmz}}{\omega\rho}P_{nm}(0,\omega)e^{jk_{nmz}z}. \tag{5}$$

B. Extrapolation and regularization

The use of Fourier Series for NAH implies that the acoustic field is periodical in x and y dimensions, a period corresponding to the measured hologram. It is obviously false, but it can be admitted if the measurement surface is significantly larger than the source, and if the measurement surface is in the nearfield. The edge discontinuities created by the periodic assumption are very penalizing for NAH, because they affect the entire wavenumber domain. A windowing operation is thus necessary to put to zero the hologram at boundaries, ensuring a continuity between periodized fields. It is also possible to use zero-padding for the DFT: the measured hologram is extended by zero bands, lowering wraparound errors caused by the periodization of acoustic fields.[2] Another method consists in using iterative approaches to extend the measured hologram not by zeros, but by using a kind of extrapolation of the hologram.[5] This method is very interesting because the windowing operation inside the measured hologram is no more required, this continuity being satisfied by the extrapolated solution. A windowing operation remains necessary to put edges of the extrapolation to zero, in order to ensure the continuity of the periodized extrapolated field.

The back-propagation of evanescent waves [k_{nmz} imaginary and z negative in Eq. (4)] is a very sensitive operation: the resulting content of the exponential operator is real and positive. It means that evanescent components are exponentially amplified by the back-propagation, and the gain is more important for high spatial wavenumbers. The problem is that high spatial wavenumbers often contain more noise than real acoustic information. Thus, a low-pass filtering operation in the wavenumber domain is required to regularize the result.[4,12]

C. Velocity-based NAH

Until a few years ago, the only available sensor for NAH was the microphone. Recently, the apparition of a new acoustic velocity sensor[7] has made the acoustic velocity-based NAH possible.[6] The normal component of the acoustic velocity is measured in the hologram plane $z=0$ instead of sound pressures. The relation in the hologram plane between the acoustic pressure and normal velocity is, from Eq. (5),

$$P_{nm}(0,\omega)=\frac{\omega\rho}{k_{nmz}}V_{nm}^z(0,\omega). \tag{6}$$

Using this expression of $P_{nm}(0,\omega)$ in Eq. (5) gives the velocity-to-velocity propagation relation:

$$V_{nm}^z(z,\omega)=V_{nm}^z(0,\omega)e^{jk_{nmz}z}. \tag{7}$$

It can be noted, comparing Eqs. (5) and (7), that the expression of the normal velocity field in the source plane from the normal velocity in the measurement plane is simpler than from acoustic pressures, for which a multiplication by k_{nmz} is required in addition to the exponential function. It means that, for pressure-to-velocity NAH, a strong amplification of high wavenumbers is still realized for small (even null) back-propagation distances. Thus, velocity-to-velocity NAH is theoretically less sensitive to measurement noise affecting high wavenumbers than pressure-to-velocity NAH. This has been confirmed experimentally.[6,13]

III. A MASS CORRECTION OPERATOR FOR MEMBRANE-BASED NAH

This paper proposes a new sensing device for acoustic velocity-based NAH. The idea is to place a light membrane in the measurement plane and to measure its velocity using a scanning laser vibrometer. For continuity reasons, the measured velocity of the membrane is equal to the normal acoustic velocity on both sides of the membrane. The major difficulty is that this is a very intrusive approach: the acoustic velocity with the membrane is not equal to the acoustic velocity that would have existed without it. The aim of this section is to show how the membrane effect can be numerically removed. This correction has been presented in a previous work,[14] and is proposed here as a pre-processing for NAH.

Assuming a tensionless infinite membrane in an anechoic space with a mass per unit area μ, the relation between incident, transmitted, and reflected waves is given as follows:

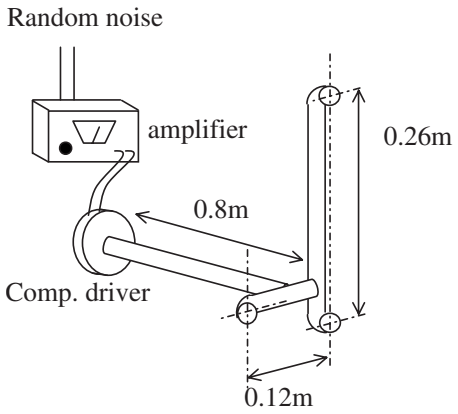


FIG. 1. Studied acoustic source.

$$P^i_{nm}(z,\omega) + P^r_{nm}(z,\omega) - P^t_{nm}(z,\omega) = -j\omega\mu V_{nm}(z,\omega) \quad (8)$$

On the other hand, normal acoustic velocities on both sides of the membrane must be equal, for continuity reasons, to the membrane velocity:

$$V^i_{nm}(z,\omega) + V^r_{nm}(z,\omega) = V^t_{nm}(z,\omega) = V_{nm}(z,\omega). \quad (9)$$

Or, using the Euler equation and taking care with wave directions:

$$P^i_{nm}(z,\omega) - P^r_{nm}(z,\omega) = P^t_{nm}(z,\omega), \quad (10)$$

$$P^t_{nm}(z,\omega) = \frac{\omega\rho}{k_{nmz}} V_{nm}(z,\omega). \quad (11)$$

The combination of Eqs. (8), (10), and (11) leads to the following relation:

$$P^i_{nm}(z,\omega) = V_{nm}(z,\omega)\frac{\omega}{2k_{nmz}}(2\rho - j\mu k_{nmz}). \quad (12)$$

Finally, the acoustic velocity that would have existed without the membrane is assessed by applying the Euler relation to the incident acoustic pressure:

$$V^c_{nm}(z,\omega) = \frac{k_{nmz}}{\omega\rho} P^i_{nm}(z,\omega) = V_{nm}(z,\omega)\left(1 - \frac{j\mu k_{nmz}}{2\rho}\right). \quad (13)$$

Equation (13) is valid if the incident wave is the same with and without the membrane. This is not true if the wave reflected by the membrane is once again reflected by any object (the studied source itself for instance), in direction of the membrane. Some standing waves can indeed appear between the membrane and the source surface, inducing erroneous velocity fields on the membrane.[8] This mass correction is thus theoretically not sufficient in all cases, and a special care remains necessary in minimizing the membrane mass to avoid reflections as much as possible.

IV. EXPERIMENTAL VALIDATION

An experiment is carried out to validate the membrane-based NAH approach. The studied source is a compression driver coupled to a copper tube (22 mm diameter) with three openings (see Fig. 1). The resulting acoustic source is equivalent to three correlated monopoles in the frequency range of interest. The volumetric velocity of each opening is obtained by using an acoustic velocity sensor, multiplying the measured normal velocity at the opening by the opening's section. A membrane is realized using a light textile material (35 g/sq m, thickness 70 μm) used in the conception of parachutes, provided by *Porcher Industries*. A rectangular piece of material (0.5×0.35 m^2) is fixed on a frame, with a minimum tension. It has been stated in a previous work[8] that the minimization of the tension in the membrane is important to avoid membrane modes. However, a residual tension, at least the tension generated by the gravity, cannot be totally suppressed. It has been shown experimentally[8] that this residual tension could be neglected. The frame is then placed in front of the acoustic source, 5 cm away from the openings plane (Fig. 2). The membrane velocity is measured with a scanning laser vibrometer. The measurement grid has $26\times19=494$ points, with a 14 mm step, resulting in a 35 $\times25$ cm^2 velocity hologram. The transfer function between the membrane velocity at each point and the excitation random signal is measured using the H1 estimator. The frequency resolution is 1 Hz, 15 time windows overlapping by 75% are used for the averaging, resulting in a total measurement time of about 35 mn.

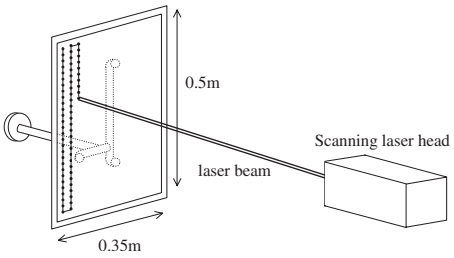


FIG. 2. Experimental setup.

The NAH is applied using regularization and extrapolation of the acoustic velocity field. The extrapolation is realized over 100% of the hologram[5] (26 points added on top and bottom sides and 19 points on right and left sides) using an iterative procedure. A k-space filter[12] is used to regularize the identification. The low-pass filter is arbitrarily centered on components of the DFT that are amplified by 25 dB.

The measured velocity fields and back-propagation results are given in Figs. 3–5 at 500, 1000, and 1700 Hz. All holograms, which are complex, are presented in real values for the phase angle maximizing the energy. It is clear that the velocity fields measured on the membrane (on the left of

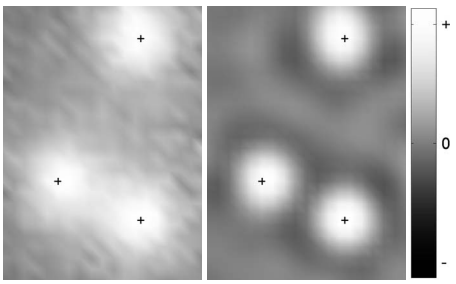


FIG. 3. 500 Hz. Left: measured velocity field Right: back-propagation result ($z=-5$ cm). Openings of the source are materialized using + markers.

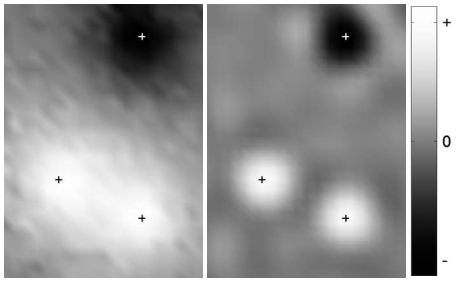


FIG. 4. 1000 Hz. Left: measured velocity field. Right: back-propagation result ($z=-5$ cm). Openings of the source are materialized using + markers.

figures) seem to have low signal-to-noise ratios: nonphysical high wavenumber are clearly disturbing the real hologram. However, the results of the back-propagation are satisfying. The three monopoles are clearly identified, with the phase relations corresponding to each frequency.

The possibility to use the membrane approach for NAH has been illustrated qualitatively. The quantitative aspect has also been studied, to test the pertinence of the mass correction operator proposed in Sec. III. The acoustic volumetric velocities of the three openings have been assessed using an acoustic velocity sensor. The acoustic velocity has been measured at the opening, and multiplied by the opening's section to get volumetric velocities. The volumetric velocities of NAH results have been assessed integrating the normal velocity on a 3 cm^2 around openings. Comparison between NAH results and direct measurements are given in Figs. 6–8 for openings situated respectively at the top, bottom and left of the source. The NAH results are presented with and without mass corrections in order to show the mass correction effects.

The comparison between microflown measurements and membrane-based NAH results with mass correction is very satisfying for the three openings: differences do not exceed 1–2 dB in the frequency range 500 Hz–2 kHz. These results validate the hypothesis of negligible tension used to formulate the correction. It is also clear that the mass correction is necessary even at low frequency: the errors without the correction are about 5 dB. It is interesting to note that, for this kind of membrane ($\mu=35$ g/sq m), the correction for a plane wave in a normal incidence would be of 2% (+0.1 dB) at 500 Hz. The correction obtained by using Eq. (13) for the measured velocity is about 5 dB at 500 Hz, which illustrates

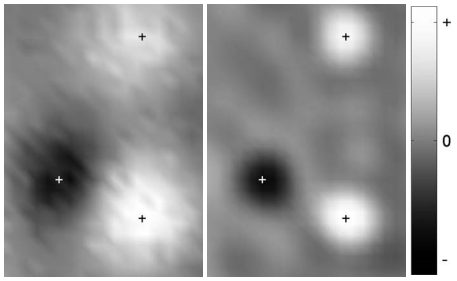


FIG. 5. 1700 Hz. Left: measured velocity field. Right: back-propagation result ($z=-5$ cm). Openings of the source are materialized using + markers.

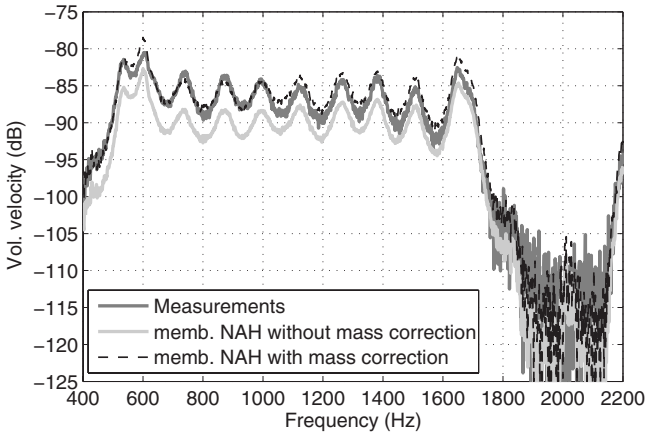


FIG. 6. Volumetric velocity of opening at the top.

the necessity of the DFT decomposition to take into account the mass effect correctly.

V. CONCLUSIONS

This paper examines the possibility to use a laser vibrometer scanning a light membrane for planar NAH. Planar NAH has been developed based on the acquisition of acoustic pressures, and more recently from acoustic normal velocities using acoustic velocity sensors. The principle of membrane-based NAH is that the velocity of the membrane, measured using a laser vibrometer, is equal to the normal acoustic velocity. Velocity-based NAH can thus be applied from laser measurements. The difficulty is that the membrane is not fully acoustically transparent, thus modifying the acoustic field in which it is inserted. A mass correction is proposed in this paper to numerically remove the membrane: the resulting velocity is the acoustic velocity that would have existed without the membrane. Meanwhile, this correction is limited to applications for which incident waves are not modified by the membrane. It is not the case if the waves reflected by the membrane are again reflected by the environment in direction of the membrane. An academic experiment is realized to illustrate the feasibility of the approach. The results show qualitatively and quantitatively the efficiency of the membrane-based NAH.

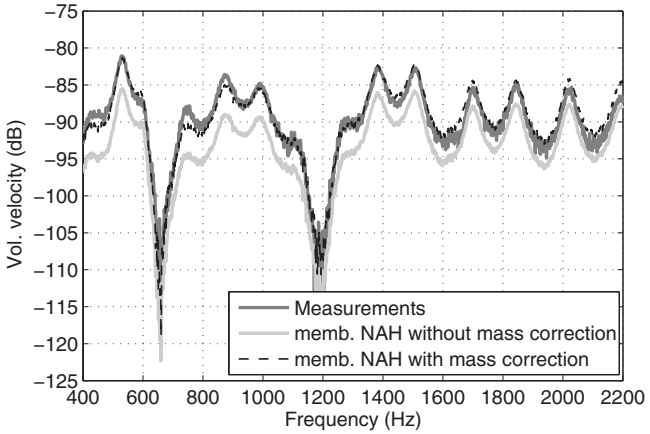


FIG. 7. Volumetric velocity of opening at the bottom.

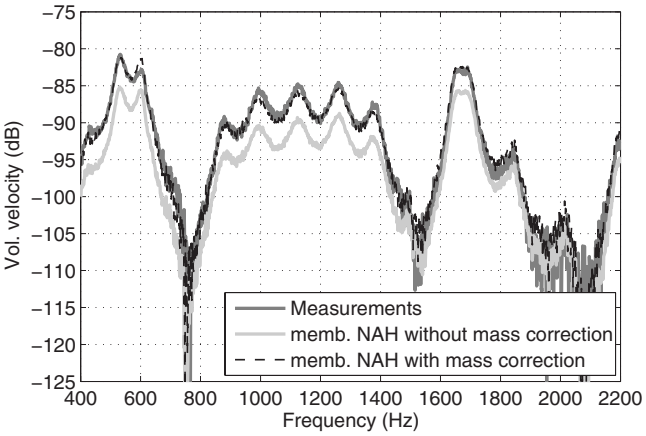


FIG. 8. Volumetric velocity of opening at the left.

ACKNOWLEDGMENT

The authors gratefully acknowledge the manufacturer *Porcher Industries* for providing them with various membrane materials.

[1]E. Williams, J. Maynard, and E. Skudrzyk, "Sound source reconstructions using a microphone array," J. Acoust. Soc. Am. **68**, 340–344 (1980).
[2]J. Maynard, E. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]W. Veronesi and J. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[4]E. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[5]E. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[6]F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," J. Acoust. Soc. Am. **118**, 3139–3144 (2005).
[7]H.-E. De Bree, P. Leussink, T. Korthorst, H. Jansen, T. Lammerink, and M. Elwenspoek, "The microflown; a novel device measuring acoustical flows," Sens. Actuators, A **SNA054/1–3**, 552–557 (1996).
[8]Q. Leclere and B. Laulagnet, "Particle velocity field measurement using an ultra-light membrane," Appl. Acoust. **69**, 302–310 (2008).
[9]D. G. Todoroff and D. Trivett, "Particle velocity detection using a thin membrane," J. Acoust. Soc. Am. **79**, S85 (1986).
[10]D. Bacon, "Primary calibration of ultrasonic hydrophone using optical interferometry," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 152–161 (1988).
[11]D. Royer, N. Dubois, and M. Fink, "Optical probing of pulsed, focused ultrasonic fields using a heterodyne interferometer," Appl. Phys. Lett. **61**, 153–155 (1992).
[12]E. Williams, *Fourier Acoustics, Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, 1999).
[13]C. Pezerat, Q. Leclere, N. Totaro, and M. Pachebat, "Identification of vibration excitations from acoustic measurements using near field acoustic holography (nah) and the force analysis technique (fat)," J. Sound Vib. **326**, 540–556 (2009).
[14]Q. Leclere and B. Laulagnet, "Imaging the acoustic field scanning an ultra light membrane using laser vibrometry," in Proceedings of Inter Noise 2007, Istambul, Turkey (2007).

Measurement of confined acoustic sources using near-field acoustic holography

Christophe Langrenne, Manuel Melon,[a] and Alexandre Garcia
Laboratoire d'Acoustique, Conservatoire National des Arts et Métiers, 292 rue Saint Martin, 75141 Paris Cedex 3, France

(Received 18 March 2009; revised 29 June 2009; accepted 30 June 2009)

Due to excessive reverberation or to the presence of secondary noise sources, characterization of sound sources in enclosed space is rather difficult to perform. In this paper a process layer is used to recover the pressure field that the studied source would have radiated in free space. This technique requires the knowledge of both acoustic pressure and velocity fields on a closed surface surrounding the source. The calculation makes use of boundary element method and is performed in two steps. First, the outgoing pressure field is extracted from the measured data using a separation technique. Second, the incoming field then scattered by the tested source body is subtracted from the outgoing field to recover free field conditions. The studied source is a rectangular parallelepiped with seven mid-range loudspeakers mounted on it. It stands at 40 cm from the rigid ground of a semi-anechoic chamber which strongly modifies the radiated pressure field, especially on the underside. After the measured data have been processed, the loudspeaker positions are recovered with a fairly good accuracy. The acoustic inverse problem is also solved to calculate the velocity field on the source surface. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3183594]

PACS number(s): 43.60.Sx, 43.20.Ye, 43.40.Sk [EGW] Pages: 1250–1256

I. INTRODUCTION

Because of their heavy weight or of any special required equipment, e.g., exhaust system, some industrial sources cannot be moved in an anechoic room and therefore must be measured *in situ*. Another possible configuration concerns the localization of the interior noise sources created by the air flow on the body of a vehicle. In that case, measured data have to be acquired in real driving conditions or in wind tunnels. Such *in situ* measurements can be very tricky to perform due to excessive reverberation and/or to the presence of other noise sources in the enclosure. Actually, the acoustic pressure field radiated by the studied source is mixed with other fields either generated by secondary sources or reflected by the walls and objects located in the room. Thus, measured data contain additional information which has to be removed. A possible solution consists in putting the probe in the very nearfield of the source where reverberated field is supposed to be negligible. Thus, sound intensity[1] and nearfield acoustical holography[2,3] (NAH) have proven to be very powerful tools for the localization of the most vibrating parts of sound sources.[4] In enclosed space, spherical arrays with large microphone number have many implementations. An interesting approach consists in measuring the pressure field with a spherical array placed near a vibrating surface.[5] Then, the acoustic vector intensity is reconstructed and mapped in a sphere which radius is larger than the one of the physical array. This method gave good results in measuring flow-noise-excited sources on the fuselage of a Boeing 757 aircraft in flight. More generally, an overview of methods for reconstructing acoustic quantities from pressure field measurements has been compiled by Wu.[6]

However, in some cases, the perturbing field has too much energy, and the classical methods fail to give reliable results. For measurement in enclosed spaces, many techniques have been developed over the past years. When dealing with enclosed spaces with simple shapes, the room influence can be taken into account in the measurement processing. For instance, Villot *et al.*[7] used a modification of NAH called phonoscopy in a rectangular parallelepiped room with one dead end to map velocity and intensity distribution on the planar surface facing the absorbent wall. However, this type of technique only deals with geometry expressed in separable coordinates and with simple boundary conditions. For small enclosures, a boundary element method (BEM) based acoustic holography[8] can be used. Bong-Ki and Jeong-Guon[9] successfully tested it on a half scale automotive cabin by using both singular value decomposition and regularization methods to inverse the so-called ill-posed problem.

Another solution consists in separating the field radiated by the tested source from the other pressure fields by using a double layer array.[10,11] This technique was used to compensate the defaults of an anechoic chamber at very low frequencies.[12] An extended version of statistically optimized nearfield acoustic holography[13] (SONAH) has also been tested showing better results than classical SONAH in noisy environment.[14]

Nevertheless, separation methods do not remove the scattered field from the outgoing field. A solution, based on doubled layer BEM (Ref. 15) for both separating the outgoing field from the incoming field and subtracting the scattered field, has been proposed and tested on simulated signals. This method will be afterward referred to as deconfined

[a]Author to whom correspondence should be addressed. Electronic mail: manuel.melon@cnam.fr

acoustic holography (DAH). DAH has been shown to give better simulation results than the ones given by a simple separation method at frequencies where scattering effects are no longer negligible. Please note that the proposed method can be seen as a first process which has to be computed before resolving the acoustical inverse problem. This approach will be tested here on a practical case with a single but perfectly reflective surface, i.e., the measurement room ground, and with no secondary sources.

In this article an experiment is performed on a rectangular parallelepiped with seven mid-range loudspeakers mounted on it. The reflecting ground strongly perturbs the acoustic fields radiated by the tested source, especially near the underside. DAH is applied to the measured data to remove the influence of the ground. The ability of the tested method is highlighted by showing results obtained with separation field process alone and with the complete process, i.e., with removal of the scattered field. Finally, a backward propagation of the processed field is performed to calculate the velocity field on the source structure. Results obtained without processing, with separation process only, with complete process, and with classical NAH are compared.

II. THEORY

A. Outline of the proposed method

The theory of DAH has already been described in Ref. 15; therefore only a brief summary will be given here. The problem consists in calculating the pressure field that would have been radiated by a primary sound source in free space from pressure and velocity measurements performed in a nonanechoic enclosure. Let S be a closed surface surrounding the source on which measurements are performed and let $\mathbf{s}$ be a point of S. Thus, measurements allow pressure $p^m(\mathbf{s})$ and pressure gradient $\partial_n p^m(\mathbf{s})$ fields to be known. The measured pressure $p^m(\mathbf{s})$ can be written as follows:

$$p^m(\mathbf{s}) = p^f(\mathbf{s}) + p^i(\mathbf{s}) + p^s(\mathbf{s}), \tag{1}$$

where $p^f(\mathbf{s})$ is the free field pressure radiated by the primary source, $p^i(\mathbf{s})$ is the incoming field radiated by all secondary sources and reflected by the enclosure, and $p^s(\mathbf{s})$ is the field scattered by the surface Γ of the tested source.

The first step of the method consists in separating the outgoing field from the incoming one. By using a Helmholtz standard integral formulation with a $e^{-i\omega t}$ dependence convention, the outgoing pressure field $p^o(\mathbf{s})=p^f(\mathbf{s})+p^s(\mathbf{s})$ is given by

$$p^o(\mathbf{s}) = \left(1 - \frac{\Omega_S}{4\pi}\right)p^m(\mathbf{s}) + \int_S [p^m(\mathbf{s}')\partial_n G(\mathbf{s},\mathbf{s}') - G(\mathbf{s},\mathbf{s}')\partial_n p^m(\mathbf{s}')]dS, \tag{2}$$

where $G(\mathbf{s},\mathbf{s}')$ is the free space Green's function for Helmholtz equation and $\Omega_S(\mathbf{s})$ is the solid angle coefficient given by[16]

$$\Omega_S(\mathbf{s}) = 4\pi + \int_S \partial_n\left(\frac{1}{|\mathbf{s}-\mathbf{s}'|}\right)dS. \tag{3}$$

One can see that $p^o(\mathbf{s})$ is obtained only from measured values of pressure and pressure gradient on S.

The second step of the method consists in removing the scattering of the incoming field. In general, the resolution of the scattering problem on complex elastic bodies is not easy to achieve. Fortunately, when dealing with rigid body machines, which is often the case for industrial sources, we can assume that $\partial_{\mathbf{n}_\Gamma} p(\mathbf{q})=0$, where $\mathbf{q}$ is a point of Γ. Then, the blocked pressure $p^b(\mathbf{q})$ on Γ can be computed using the following equation when the primary source is stopped:

$$\int_\Gamma p^b(\mathbf{q}')\partial_n G(\mathbf{q},\mathbf{q}')d\Gamma - \int_S [p^m(\mathbf{s}')\partial_n G(\mathbf{q},\mathbf{s}') - G(\mathbf{q},\mathbf{s}')\partial_n p^m(\mathbf{s}')]dS = \frac{\Omega_\Gamma}{4\pi}p^b(\mathbf{q}). \tag{4}$$

In the above equation, the second integral term represents the incoming field. Then, the scattered field $p^s(\mathbf{s})$ on S is determined by the following expression:

$$p^s(\mathbf{s}) = \int_\Gamma p^b(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}')dS. \tag{5}$$

Finally, the free field pressure with rigid body assumption $\tilde{p}^f(\mathbf{s})$ is given by

$$\tilde{p}^f(\mathbf{s}) = p^o(\mathbf{s}) - p^s(\mathbf{s}). \tag{6}$$

The resolution of the above equations is easily performed using BEM. Please note that the determination of the blocked pressure is a Fredholm integral equation of the second kind, thus giving no issues with the amplification of the noise components of the evanescent waves. Thus, no regularization methods[17] are needed, contrary to what usually encountered with ill-posed problems. Some difficulties may occur when the wave number coincides with a Dirichlet eigenvalue. In those circumstances, the usual Combined Helmholtz Integral Equation Formulation (CHIEF) is used to overcome this problem.[18,19]

B. Application to the tested case

In order to compare the results given by DAH to standard NAH, a simple experimental configuration has been chosen. The complex source is measured in a semi-anechoic room. Thus, the only acoustic perturbation is due to the rigid ground. The geometry of this configuration is shown in Fig. 1 where the rigid ground has been replaced by an image source. Under this condition, the measured pressure field is given by

$$p^m(\mathbf{s}) = p^f(\mathbf{s}) + p^s(\mathbf{s}) + p^{f\prime}(\mathbf{s}) + p^{s\prime}(\mathbf{s}), \tag{7}$$

where $p^{f\prime}(\mathbf{s})$ and $p^{s\prime}(\mathbf{s})$, respectively, are the free field pressure and the scattered field radiated by the image source. Note that $p^{f\prime}(\mathbf{s})+p^{s\prime}(\mathbf{s})=p^i(\mathbf{s})$.

When using a standard Helmholtz integral formulation of this problem, the measured pressure is then given by

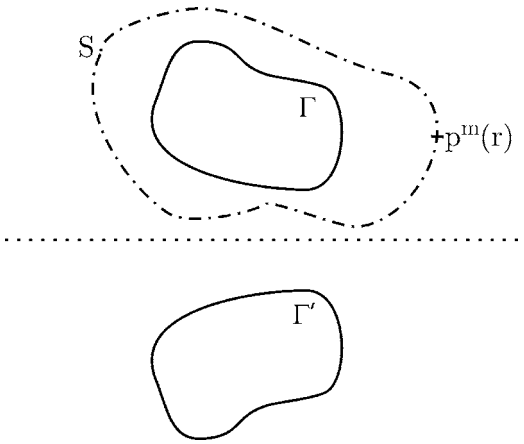


FIG. 1. Geometry of interest.

$$p^m(\mathbf{s}) = \int_\Gamma [p_{11}(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n p_{11}(\mathbf{q}')]d\Gamma + \int_{\Gamma'} [p_{22}(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n p_{22}(\mathbf{q}')]d\Gamma' + \int_\Gamma [p_{12}(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n p_{12}(\mathbf{q}')]d\Gamma + \int_{\Gamma'} [p_{21}(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n p_{21}(\mathbf{q}')]d\Gamma', \quad (8)$$

where Γ' is the surface of the image source, p_{11} and p_{22} are the pressure fields radiated by each source alone, while p_{12} and p_{21} are the coupled pressure fields, i.e., the fields radiated by a source on the other one.

With the rigid body assumption $[\partial_n p_{12}(\mathbf{q}')=0$ for $\mathbf{q}' \in \Gamma$ and $\partial_n p_{21}(\mathbf{q}')=0$ for $\mathbf{q}' \in \Gamma']$ and with the notations used in this paper, Eq. (8) can be rewritten as follows:

$$p^m(\mathbf{s}) = \int_\Gamma \{[\tilde{p}^f(\mathbf{q}') + p^b(\mathbf{q}')]\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n \tilde{p}^f(\mathbf{q}')\}dS + \int_{\Gamma'} \{[\tilde{p}^{f\prime}(\mathbf{q}') + p^{b\prime}(\mathbf{q}')]\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n \tilde{p}^{f\prime}(\mathbf{q}')\}dS. \quad (9)$$

When using symmetry properties of the problem, the measured pressure is also given by

$$p^m(\mathbf{s}) = \int_\Gamma \{[\tilde{p}^f(\mathbf{q}') + p^b(\mathbf{q}')]\partial_n G_r(\mathbf{s},\mathbf{q}') - G_r(\mathbf{s},\mathbf{q}')\partial_n \tilde{p}^f(\mathbf{q}')\}d\Gamma, \quad (10)$$

where $G_r(\mathbf{s},\mathbf{q}')$ is the rigid ground Green's function. Resolving the inverse acoustic problem given by Eq. (10) using BEM with rigid ground Green's function applied to the measured data yields the determination of $\partial_n \tilde{p}^f$ and $\tilde{p}^f + p^b$ on Γ.

In addition, the free field pressure calculated by DAH can be written as follows:

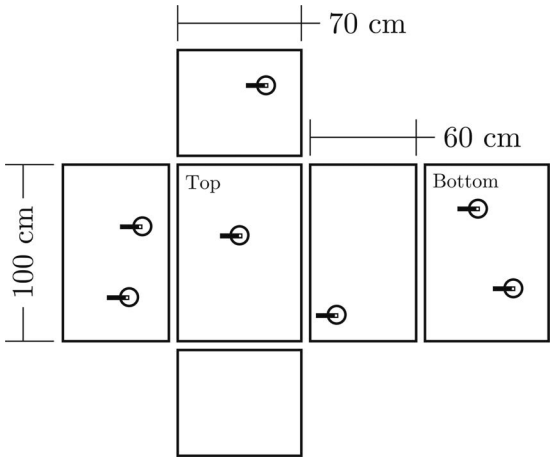


FIG. 2. Orthographic scaled projection of the tested source also showing reference microphone positions.

$$\tilde{p}^f(\mathbf{s}) = \int_\Gamma \{\tilde{p}^f(\mathbf{q}')\partial_n G(\mathbf{s},\mathbf{q}') - G(\mathbf{s},\mathbf{q}')\partial_n \tilde{p}^f(\mathbf{q}')\}d\Gamma. \quad (11)$$

Resolving the inverse acoustic problem given by Eq. (11) using BEM with a free space Green's function and with $\tilde{p}^f(\mathbf{s})$ calculated on S by DAH allows the determination of $\partial_n \tilde{p}^f$ and $\tilde{p}^f$ on Γ. Thus, velocity field on the tested source should be the same if calculated by any of these two last methods. This result will be discussed in Sec. III. Please note that DAH and NAH should lead to the same results for the determination of the normal velocity field in this particular configuration. However, for complex enclosed space with secondary sources, only DAH can remove the influence of the testing room.

III. MEASUREMENTS

A. Measurement set-up

The studied source is a rectangular parallelepiped with seven mid-range loudspeakers mounted on five of its six faces. Its orthographic scaled projection is shown in Fig. 2. The circles give loudspeakers' positions. The interior box has been designed to separate the wave fields radiated by the rear side of the speakers' membranes. The box is made of 19 mm thick pressed wood and stands on four legs at $d=40$ cm from the rigid ground of a semi-anechoic room. Seven bandwidth limited (200–2000 Hz) white noise signals are used to drive the loudspeakers. The seven electrical signals are uncorrelated. Acoustic pressure and pressure gradient are then measured on a closed surface S encompassing the source. These quantities are calculated from data measured by a p-p probe made of two KE4 Sennheiser microphones calibrated in amplitude and phase. The spacing between the two microphones is 3 cm yielding the following measurement bandwidth: 60–3450 Hz. Measurements on S are performed on a plane-parallel mesh (smoothed with round edges) made of 1050 points. Measurement surfaces lie at 11.5 cm from the source which leads to the following spacing between probe positions: $\Delta_{\text{length}}=7.69$ cm, $\Delta_{\text{width}}=7.75$ cm, and $\Delta_{\text{height}}=6.92$ cm. As the measurements are not performed at the same time, reference microphones are used. Positions of the

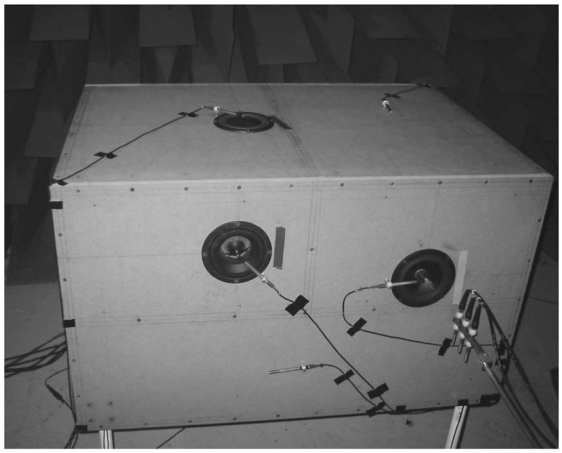


FIG. 3. Experimental set-up.

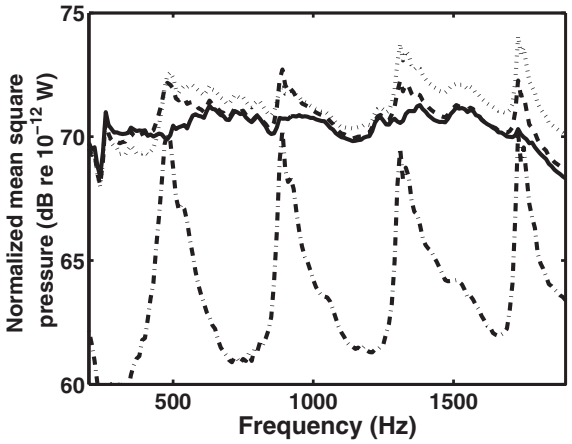


FIG. 4. Normalized mean square pressure estimator Π computed on S. Dotted line: Π^m, dashed line: Π^o, dashed-dotted line: and Π^s, solid line: Π^f.

seven reference microphones are indicated in Fig. 2. A picture of the wooden model is shown in Fig. 3. One can catch sight of three loudspeakers, five reference microphones (although only the ones in front of the loudspeakers were used), and three microphone probes (only the two nearest from the source were used here).

B. Results

Numerical resolution of the integral equations developed in the previous chapter is based on its discretization on surface elements. An isoparametric formulation based on Seybert *et al.*[16] using quadratic shape functions (six node curvilinear triangular elements) has been implemented. Although all electrical signals were uncorrelated, reference microphones at the underside were slightly correlated (the correlation coefficient for these two microphones fluctuates between 0.1 and 0.3 in the studied frequency band). To overcome this problem, a principal component analysis[20] has been performed. Thus, although DAH has been processed separately on every partial based on a reference microphone, all quantities plotted in this paper are recomposed from all partials.

A normalized mean square pressure estimator is computed for each pressure field type using the following equation:

$$\Pi^e = \int_S \frac{|p^e(s)|^2}{\rho_0 c} dS, \tag{12}$$

where p^e is the pressure field on which Π is estimated. The quantity p^e can either be the measured pressure field $p^m(\mathbf{s})$, the outgoing field $p^o(\mathbf{s})$, the scattered field $p^s(\mathbf{s})$, or the recovered free pressure field $\tilde{p}^f(\mathbf{s})$. Figure 4 shows the evolution of Π as function of frequency. Π^m shows oscillations which can be related to the modal behavior of the air volume between the base of the tested source and the rigid ground. Table I gives the deviation between the theoretical resonance frequencies: $f_n = nc/2d$, where n is a natural integer, and the observed frequencies. At low frequencies the deviation between the two values is significant (15.3% for the first mode). However, when the wavelength becomes small compared to the box dimensions, the discrepancy becomes negligible (1.8% for the fourth mode). After the separation process has been applied, the energy reflected by the rigid ground has been removed. However, the modal behavior is still visible on the Π^o curve. The plot of Π^f shows that this quantity is very large near resonance frequencies. When it is removed from Π^o, the Π^f curve is nearly flat.

To emphasize the outcomes obtained by the proposed process, pressure level maps plotted on the underside are quite useful. Figure 5(a) shows the measured pressure level at 11.5 cm from the underside at 890 Hz. One can see that the strongest levels do not coincide with the loudspeakers' positions which are symbolized with white circles. Then, the separation process alone is applied to the measured data which allows pressure level maps of $p^o(\mathbf{s})$ to be plotted [see Fig. 5(b)]. On this graph, loudspeakers' positions are more easily recovered although spots are very large compared to the white circles. Finally, removal of the scattered field [Fig. 5(c)] is performed and pressure level maps of $\tilde{p}^f(\mathbf{s})$ are shown in Fig. 5(d). The scattered field map exhibits large maximum pressure values at the center of the underside which prevent a correct identification of sound sources. When this last quantity is subtracted, the positions of the two loudspeakers appear with a greater accuracy [Fig. 5(d)].

The about 7 cm spacing between measurement points allows a maximum studying frequency of 814 Hz when using the well-known $\lambda/6$ criterion. Although the resonance frequency of the third mode is larger than the maximum authorized frequency value (1310 Hz versus 814 Hz), the complete process is still applied. Results are shown in Fig. 6 where pressure level map of $\tilde{p}^f(\mathbf{s})$ is plotted. Here, loudspeakers' positions are still clearly identifiable. To explain this result, remember that BEM processing is performed with quadratic shape functions which are supposed to give a better interpolation of acoustic quantities than linear functions do.

TABLE I. Deviation between theoretical resonance frequencies (longitudinal modes between two rigid planes) and observed frequencies.

Theoretical resonance frequency (Hz)	425	850	1275	1700
Observed resonance frequency (Hz)	490	891	1310	1730
Deviation (%)	15.3	4.8	2.8	1.8

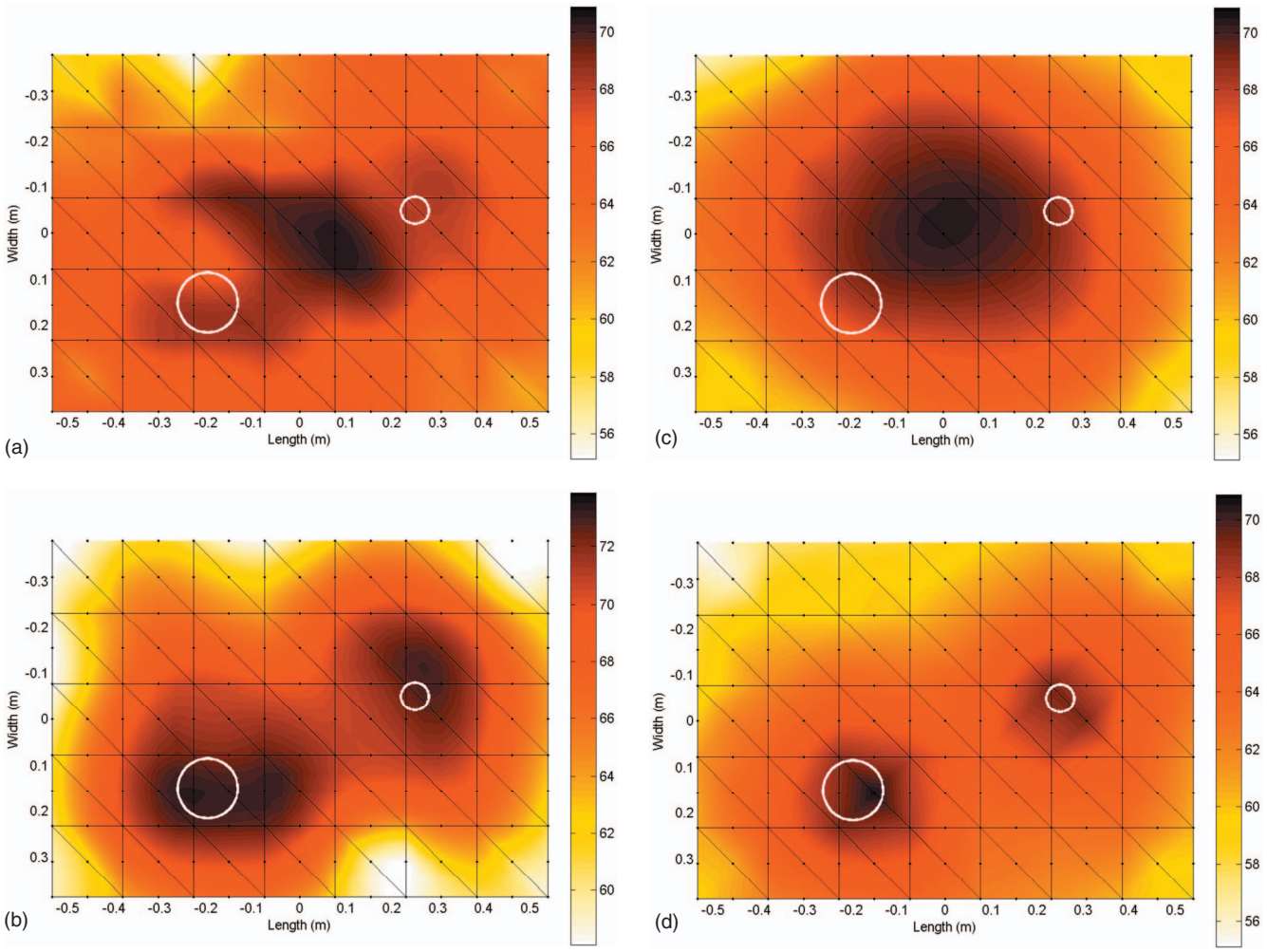


FIG. 5. Pressure field in dB for 2×10^{-5} at 890 Hz on the underside of the tested source: (a) measured pressure field, (b) outgoing pressure field, (c) scattered pressure field, and (d) recovered free pressure field.

C. Backward propagation

An illustration of the ability of DAH to characterize confined sources is shown on the reconstruction of the normal surface velocity v_n^Γ on Γ. The standard ill-posed inverse problem is solved using an inverse boundary element method (IBEM) with a Tikhonov[21] filter regularized with the L-curve[22] method. The target surface Γ is described by a 666 point mesh decomposed on four node quadrangle elements. The spacings between point positions are $\Delta_{\text{length}}=7.14$ cm, $\Delta_{\text{width}}=7$ cm, and $\Delta_{\text{height}}=7.5$ cm.

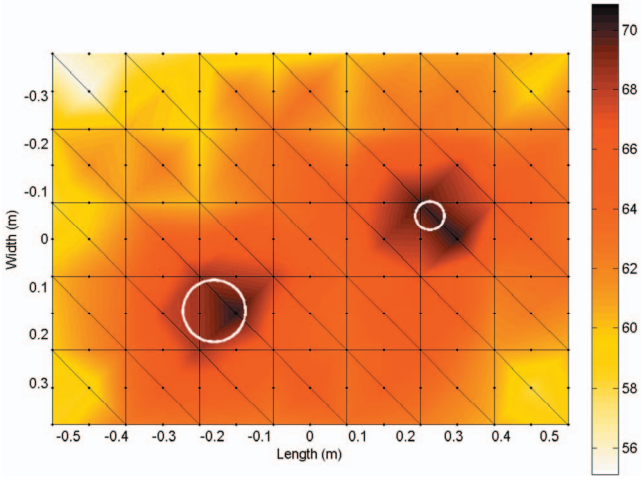


FIG. 6. Recovered free field pressure in dB for 2×10^{-5} at 1310 Hz on the underside of the tested source.

Figure 7(a) shows v_n^Γ maps at 890 Hz on the source underside computed from the measured pressure field on S with a free space Green's function. Even though the loudspeakers' positions are distinguishable, many other zones which show comparable velocity levels do not correspond to real sources. When v_n^Γ is computed from $p^o(\mathbf{s})$, the pressure field obtained by the separation method [see Fig. 7(b)], the "ghost" sources disappear but the dynamic range is lower than on the velocity map calculated with pressure field $\tilde{p}^f(\mathbf{s})$ obtained by DAH [Fig. 7(c)]. A map of v_n^Γ computed from measured acoustic data on S with a rigid ground Green's function is shown in [Fig. 7(d)]. As predicted by Sec. II B results, one can see that reconstructed velocities on the loudspeakers have nearly the same levels when calculated with DAH or standard NAH with double Green's function. Nevertheless, distribution of the less energetic zones is not the same depending on the chosen method. Note that low levels are strongly dependent on the regularization parameter. Considering that these zones are 30 dB below maximum levels,

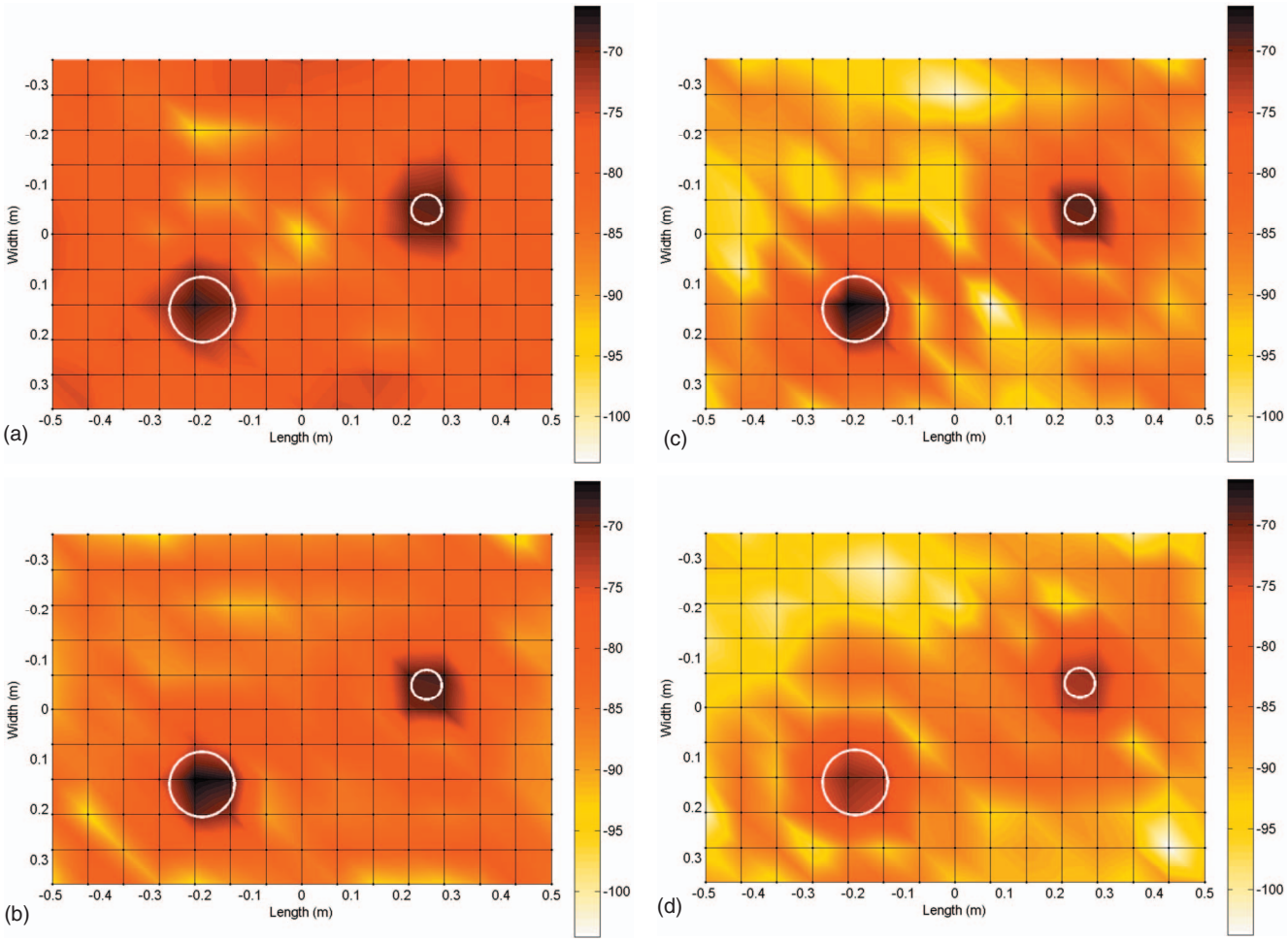


FIG. 7. Normal velocity field in dB (ref. 1 m s^{-1}) at 890 Hz on the underside of the tested source computed from the (a) measured pressure field with a free space Green's function, (b) outgoing pressure field with a free space Green's function, (c) recovered free pressure field with a free space Green's function, and (d) measured pressure field with a rigid ground Green's function.

this drawback is not very significant. Loudspeakers on other sides are less confined; thus their reconstruction is accurate within a half wavelength resolution without the use of DAH. However, DAH allows recovering maps with higher signal to noise ratio even for these loudspeakers.

Figure 8(a) shows the blocked pressure field in dB (re 2×10^{-5} Pa) at 890 Hz on the underside of the tested source computed from Eq. (4). An interesting outcome is that the locations of the speakers are not clearly visible on the maps which tends to confirm the efficiency of the DAH process and shows that the rigid source assumption is valid. Blocked pressure field has also been computed by solving the integral formulation on Γ using v_n^{Γ} obtained by IBEM with G_r. Resolution is performed twice, once using Eq. (10) and once using Eq. (11) on the structure Γ. Subtracting results from both calculations yields the determination of the blocked pressure [Fig. 8(b)]. As expected, the two maps shown in Fig. 8 look alike (with small differences due to the regularization process).

IV. SUMMARY

In this article, DAH has been applied to a complex three dimensional source. Measurements were performed in a semi-anechoic room. Results show that acoustic quantities at the underside of the source were strongly modified by the rigid ground. After DAH has been applied to the quantities collected on S, recovered free field pressure maps allowed a good localization of loudspeakers. When solving the acoustical inverse problem, results obtained with DAH were very similar to the ones obtained with classic NAH and rigid Green's function. Note that this outcome is specific to the rigid ground configuration and that DAH is supposed to also give good results in more general cases.

Concerning the practical implementation of DAH, one can expect measurements to be difficult to perform because of the large point number needed for industrial sources which typically have large dimensions. Fortunately, acquisition channels came down in price, and acoustic velocity can be estimated with a sufficient precision with well-calibrated cheap electret microphones. In fact, p-p or p-u arrays with typically more than 40 high quality probes are already commercialized. Thus, one can imagine that measurements can be performed all around the source using that kind of array with typically measurements performed at about 10 or 20 different times. In such a configuration, reference microphones will be needed, and the work carried out by Kim *et al.*[23] and Nam and Kim[24] on multireference processing and NAH should be applied to DAH. Multiple array acquisition

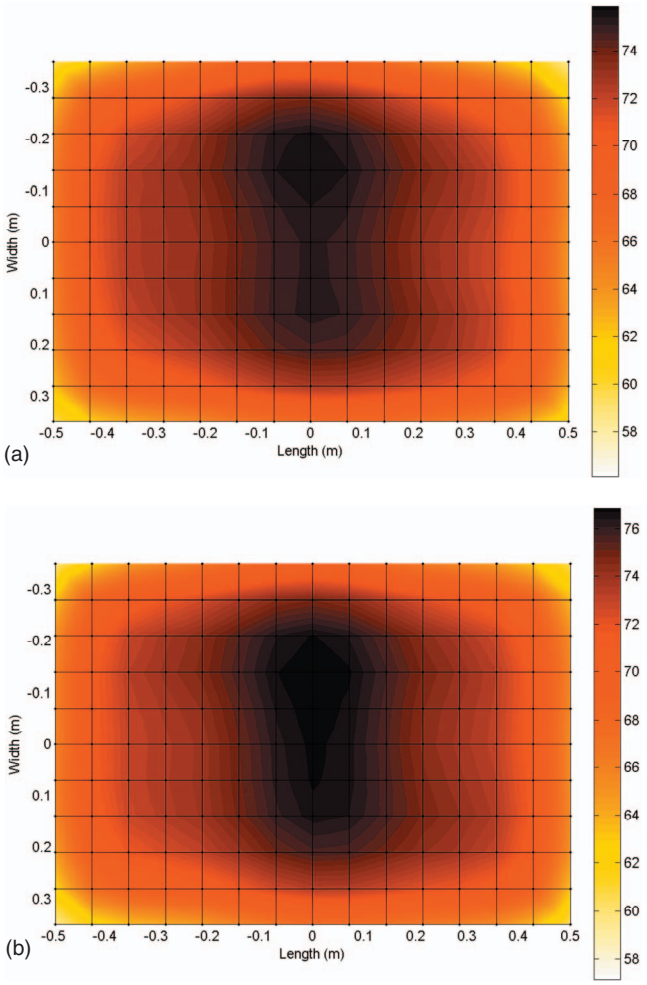


FIG. 8. Blocked pressure field in dB for 2×10^{-5} at 890 Hz on the underside of the tested source computed from (a) Eq. (4) and (b) normal velocity field on Γ.

can lead to uncorrelated sources with partly correlated reference signal issue. This problem is under study in our laboratory, and results will be reported in a forthcoming paper.

ACKNOWLEDGMENT

The authors wish to thank Jean Baptiste Legland for his meticulous efforts in building the tested source and for his enthusiastic efforts in performing part of the measurements.

[1]F. J. Fahy, *Sound Intensity* (Elsevier, London, 1989).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH) II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[4]T. E. Reinhart and M. J. Crocker, "Source identification of a diesel engine using acoustic intensity measurements," Noise Control Eng. **18**, 84–92 (1982).
[5]E. Williams, N. Valdivia, and P. Herdic, "Volumetric acoustic vector intensity imager," J. Acoust. Soc. Am. **120**, 1887–1897 (2006).
[6]S. F. Wu, "Methods for reconstructing acoustic quantities based on acoustic pressure measurements," J. Acoust. Soc. Am. **124**, 2680–2697 (2008).
[7]M. Villot, G. Chavériat, and J. Roland, "Phonoscopy: An acoustical holograhy technique for plane structures in enclosed spaces," J. Acoust. Soc. Am. **91**, 187–195 (1992).
[8]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am. **85**, 588–598 (1989).
[9]K. Bong-Ki and I. Jeong-Guon, "On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3014 (1996).
[10]G. Weinreich and E. B. Arnold, "Method for measuring acoustic radiation fields," J. Acoust. Soc. Am. **68**, 404–411 (1980).
[11]I. E. Tsukernikov, "Calculation of the field of a sound source in a bounded space," Sov. Phys. Acoust. **35**, 304–306 (1989).
[12]M. Melon, C. Langrenne, D. Rousseau, and P. Herzog, "Comparison of four subwoofer measurement techniques," J. Audio Eng. Soc. **55**, 1077–1091 (2007).
[13]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," in *Proceedings of the ISCV6* (1999), pp. 843–850.
[14]F. Jacobsen, X. Chen, and V. Jaud, "A comparison of statistically optimized near field acoustic holography using single layer pressure-velocity measurements and using double layer pressure measurements," J. Acoust. Soc. Am. **123**, 1842–1845 (2008).
[15]C. Langrenne, M. Melon, and A. Garcia, "Boundary element method for the acoustic characterization of a machine in bounded noisy environment," J. Acoust. Soc. Am. **121**, 2750–2757 (2007).
[16]A. Seybert, B. Soenarko, F. J. Rizzo, and D. J. Shippy, "An advance computational method for radiation and scattering of acoustic waves in three dimensions," J. Acoust. Soc. Am. **77**, 362–368 (1985).
[17]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[18]H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[19]W. Tobocman, "Calculation of acoustic wave scattering by means of the Helmholtz integral equation. I," J. Acoust. Soc. Am. **76**, 599–607 (1984).
[20]H.-S. Kwon, Y.-J. Kim, and J. Bolton, "Compensation for source nonstationarity in multireference, scan-based near-field acoustical holography," J. Acoust. Soc. Am. **113**, 360–368 (2003).
[21]A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola, *Numerical Methods for the Solution of Ill-Posed Problems* (Kluwer Academic, Dordrecht, 1995).
[22]P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," SIAM J. Sci. Comput. (USA) **14**, 1487–1503 (1993).
[23]Y.-J. Kim, J. S. Bolton, and H.-S. Kwon, "Partial sound field decomposition in multireference near-field acoustical holography by using optimally located virtual references," J. Acoust. Soc. Am. **115**, 1641–1652 (2004).
[24]K.-U. Nam and Y.-H. Kim, "A partial field decomposition algorithm and its examples for near-field acoustic holography," J. Acoust. Soc. Am. **116**, 172–185 (2004).

Near field acoustic holography based on the equivalent source method and pressure-velocity transducers

Yong-Bin Zhang[a)]
Institute of Sound and Vibration Research, Hefei University of Technology, Hefei 230009, China

Finn Jacobsen
Department of Electrical Engineering, Acoustic Technology, Technical University of Denmark, Building 352, Ørsteds Plads, DK-2800 Kongens, Lyngby, Denmark

Chuan-Xing Bi and Xin-Zhao Chen
Institute of Sound and Vibration Research, Hefei University of Technology, Hefei 230009, China

(Received 2 January 2009; revised 26 May 2009; accepted 18 June 2009)

The advantage of using the normal component of the particle velocity rather than the sound pressure in the hologram plane as the input of conventional spatial Fourier transform based near field acoustic holography (NAH) and also as the input of the statistically optimized variant of NAH has recently been demonstrated. This paper examines whether there might be a similar advantage in using the particle velocity as the input of NAH based on the equivalent source method (ESM). Error sensitivity considerations indicate that ESM-based NAH is less sensitive to measurement errors when it is based on particle velocity input data than when it is based on measurements of sound pressure data, and this is confirmed by a simulation study and by experimental results. A method that combines pressure- and particle velocity-based reconstructions in order to distinguish between contributions to the sound field generated by sources on the two sides of the hologram plane is also examined. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3179665]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Rz [EGW] Pages: 1257–1263

I. INTRODUCTION

Near field acoustic holography (NAH) is a powerful tool for visualizing sound fields radiated by complicated sound sources. In addition to the classical NAH technique based on spatial discrete Fourier transforms,[1–3] many alternative methods have been developed in the past years, e.g., the inverse boundary element method,[4–6] the statistically optimized method,[7,8] the Helmholtz equation least-squares method,[9,10] and the equivalent source method (ESM) (also known as the wave superposition method).[11–16] Usually the measured quantity is the sound pressure rather than the particle velocity, simply because pressure microphones are readily available and easy to calibrate whereas the particle velocity has been difficult to measure. However, in recent years, a particle velocity transducer called Microflown has appeared.[17] Conventional planar NAH based on measurement of particle velocity was first considered by Jacobsen and Liu.[18] As they demonstrated, NAH based on measurement of the particle velocity transducers performs very well compared with pressure-based NAH. Statistically optimized NAH based on measurement of particle velocity has also been investigated, and the results showed a similar advantage.[19] In addition, a so-called *p-u* method based on combined measurements of the pressure and the particle velocity was proposed. Measuring both quantities makes it possible to overcome the usual requirement of a source free region on the other side of the hologram plane.[19]

The purpose of this paper is to examine the performance of NAH based on the ESM and measurements with particle velocity transducers. The combination of the *p-u* method and NAH based on ESM is also examined.

II. OUTLINE OF THEORY

A. ESM based on measurement of sound pressure

ESM is based on the idea of modeling the sound field generated by a vibrating structure by a set of simple sources placed in the interior of the structure. Such a superposition has been proved to be mathematically equivalent to the Helmholtz integral formulation.[11] Given that M measurement points are selected on the hologram surface and the number of the equivalent sources is N, the pressure column vector $\mathbf{P}_h$ at the measurement positions can be represented in matrix form as

$$\mathbf{P}_h = i\rho ck\mathbf{G}_{hp}\mathbf{Q}, \tag{1}$$

where ρ is the density of the medium, c is the speed of sound, $k=\omega/c$ is the wave number, ω is the angular frequency, $\mathbf{Q}=[q(\mathbf{r}_{o1}),q(\mathbf{r}_{o2}),\ldots,q(\mathbf{r}_{oN})]^T$ is the column vector with the strengths of the equivalent sources $q(\mathbf{r}_{on})$, $\mathbf{r}_{on}$ is the location vector of the nth equivalent source, and $\mathbf{G}_{hp}$ is the complex transfer matrix obtained from Green's function,

$$\mathbf{G}_{hp}|_{m,n} = g(\mathbf{r}_{hm},\mathbf{r}_{on}) = -\frac{e^{ikr}}{4\pi r}, \quad r=|\mathbf{r}_{hm}-\mathbf{r}_{on}|, \tag{2}$$

in which $\mathbf{r}_{hm}$ is the location vectors of the mth measurement point and g is the free space Green's function with the $e^{-i\omega t}$

a)Author to whom correspondence should be addressed. Electronic mail: zybmy1997@163.com

sign convention. The unknown source strength vector $\mathbf{Q}$ can be obtained from the expression

$$\mathbf{Q} = \frac{1}{i\rho ck}\mathbf{G}_{hp}^{+}\mathbf{P}_h, \tag{3}$$

where the generalized inverse matrix $\mathbf{G}_{hp}^{+}$ is obtained from $\mathbf{G}_{hp}$ by singular value decomposition. Once the source strength vector has been determined, the pressure and the normal velocity on the surface of the acoustic source can be reconstructed as

$$\mathbf{P}_s = i\rho ck\mathbf{G}_{sp}\mathbf{Q}, \tag{4}$$

$$\mathbf{U}_{ns} = \mathbf{G}_{sv}\mathbf{Q}, \tag{5}$$

where $\mathbf{P}_s$ and $\mathbf{U}_{ns}$ are the reconstructed pressure and normal velocity vectors on the surface of the source, and $\mathbf{G}_{sp}$ and $\mathbf{G}_{sv}$ are complex transfer matrices,

$$\mathbf{G}_{sp}|_{m,n} = g(\mathbf{r}_{sm}, \mathbf{r}_{on}), \tag{6}$$

$$\mathbf{G}_{sv}|_{m,n} = \frac{\partial g(\mathbf{r}_{sm}, \mathbf{r}_{on})}{\partial \mathbf{n}_s}. \tag{7}$$

In these expressions $\mathbf{r}_{sm}$ is the location vector of the mth point on the surface of the source, and $\mathbf{n}_s$ is the outward normal of the source.

B. ESM based on measurement of particle velocity

It is a simple matter to modify the foregoing to ESM based on measurement of the particle velocity. If the particle velocity normal to the measurement surface is measured at M points, then Eq. (1) is replaced with

$$\mathbf{V}_{nh} = \mathbf{G}_{hv}\mathbf{Q}, \tag{8}$$

where $\mathbf{V}_{nh}$ is a column vector of normal components of the particle velocity at the measurement positions, and $\mathbf{G}_{hv}$ is a complex transfer matrix,

$$\mathbf{G}_{hv}|_{m,n} = \frac{\partial g(\mathbf{r}_{hm}, \mathbf{r}_{on})}{\partial \mathbf{n}_h}, \tag{9}$$

in which $\mathbf{n}_h$ is the outward normal of the hologram surface. Once $\mathbf{Q}$ has been determined the reconstruction can be realized by Eqs. (4) and (5).

C. ESM based on measurement of pressure and particle velocity

It is impossible to distinguish between sounds generated by sources on the two sides of the hologram plane only from measurement of pressure or particle velocity. A separation technique is needed as, e.g., the double layer method proposed by Bi *et al.*,[20] or the slightly different method based on measurement of pressure and a finite difference estimate of the particle velocity proposed by Bi and Chen.[21] Here the somewhat simpler idea proposed by Jacobsen and Jaud is introduced,[19] and a *p-u* method based on measurement of both the pressure and the normal component of the particle velocity is developed. From Eqs. (3)–(5) and (8), the reconstructed pressure and particle velocity can be expressed as follows:

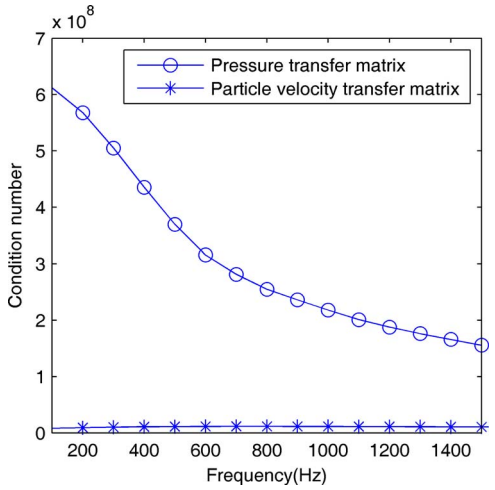


FIG. 1. (Color online) Condition number of transfer matrices.

$$\mathbf{P}_s = \frac{i\rho ck}{2}\mathbf{G}_{sp}\left(\frac{1}{i\rho ck}\mathbf{G}_{hp}^{+}\mathbf{P}_h + \mathbf{G}_{hv}^{+}\mathbf{V}_{nh}\right), \tag{10}$$

$$\mathbf{U}_{ns} = \frac{1}{2}\mathbf{G}_{sv}\left(\frac{1}{i\rho ck}\mathbf{G}_{hp}^{+}\mathbf{P}_h + \mathbf{G}_{hv}^{+}\mathbf{V}_{nh}\right). \tag{11}$$

The separation is possible because of the fact that the particle velocity is a vector component that changes sign, unlike the pressure, if the source is moved to a symmetrical position on the opposite side of the hologram plane. As in Ref. 19, one should perhaps not expect the same accuracy in the general case where the disturbing noise is not coming from a source placed symmetrically with respect to the hologram plane.

D. Error analysis and comparison of condition numbers

In practice, measured data are always contaminated by errors. Suppose that the real measured pressure and the normal component of the particle velocity can be written as follows:

$$\mathbf{P}_h = (\mathbf{P}_h)_r + (\mathbf{P}_h)_e, \tag{12}$$

$$\mathbf{V}_{nh} = (\mathbf{V}_{nh})_r + (\mathbf{V}_{nh})_e, \tag{13}$$

where the subscripts r and e denote the exact value and the error. Substituting Eqs. (12) and (13) into Eq. (5), the reconstructed surface normal velocity becomes

$$\mathbf{U}_{ns} = (\mathbf{U}_{ns})_r + (\mathbf{U}_{ns})_e = \frac{1}{i\rho ck}\mathbf{G}_{sv}\mathbf{G}_{hp}^{+}(\mathbf{P}_h)_r + \frac{1}{i\rho ck}\mathbf{G}_{sv}\mathbf{G}_{hp}^{+}(\mathbf{P}_h)_e \tag{14}$$

if it is based on pressure measurements, and

$$\mathbf{U}_{ns} = (\mathbf{U}_{ns})_r + (\mathbf{U}_{ns})_e = \mathbf{G}_{sv}\mathbf{G}_{hv}^{+}(\mathbf{V}_{nh})_r + \mathbf{G}_{sv}\mathbf{G}_{hv}^{+}(\mathbf{V}_{nh})_e \tag{15}$$

if it is based on velocity measurements. According to Ref. 16, an upper bound of the relative error of the reconstructed normal velocity can be expressed as

$$\frac{\|(\mathbf{U}_{ns})_e\|}{\|(\mathbf{U}_{ns})_r\|} \leq \begin{cases} \text{cond}(\mathbf{G}_{sv})\text{cond}(\mathbf{G}_{hp})\dfrac{\|(\mathbf{P}_h)_e\|}{\|(\mathbf{P}_h)_r\|} & \text{(pressure measurement)} \\ \text{cond}(\mathbf{G}_{sv})\text{cond}(\mathbf{G}_{hv})\dfrac{\|(\mathbf{V}_{nh})_e\|}{\|(\mathbf{V}_{nh})_r\|} & \text{(particle velocity measurement)}, \end{cases} \tag{16}$$

where $\|\cdot\|$ denotes the norm of a matrix, and cond $(\cdot)$ denotes the condition number of a matrix. Equation (16) demonstrates that the errors will be magnified by the condition number of the transfer matrix. It follows that if the two measurements have the same error level, the condition numbers of the transfer matrices $\mathbf{G}_{hp}$ and $\mathbf{G}_{hv}$ determine the influence of measurement error on the reconstructed velocity. A similar error expression can be obtained for the reconstructed pressure by replacing $\mathbf{G}_{sv}$ with $\mathbf{G}_{sp}$.

The condition number of $\mathbf{G}_{hv}$ is much smaller than the condition number of $\mathbf{G}_{hp}$ in near fields where NAH is used, which leads to the conclusion that NAH based on the particle velocity is less sensitive to measurement errors than pressure-based NAH. This is demonstrated by an example. Suppose that a planar acoustic source is located at $z=0$ with dimensions 0.6×0.6 m^2 and modeled with a grid of 21×21. The hologram plane and the equivalent source plane are located at $z=0.05$ m and $z=-0.03$ m, respectively, with the same dimension as the source and a grid of 31×31. The condition numbers of $\mathbf{G}_{hp}$ and $\mathbf{G}_{hv}$ are shown in Fig. 1. It is obvious that the condition number of $\mathbf{G}_{hv}$ is smaller than the condition number of $\mathbf{G}_{hp}$ in the entire frequency range from 100 to 1500 Hz. At low frequencies the condition number of $\mathbf{G}_{hp}$ is more than 70 times larger than that of $\mathbf{G}_{hv}$. The ratio decreases with the frequency but is still significant at the upper limiting frequency. Figure 2 shows all the singular values of $\mathbf{G}_{hp}$ and $\mathbf{G}_{hv}$ at 200 Hz. It can be seen that the singular values of $\mathbf{G}_{hp}$ decay faster than the singular values of $\mathbf{G}_{hv}$ (from similar high values), which explains why the pressure-based approach is more ill-posed than the particle velocity-based approach. Similar results have been obtained at other frequencies.

The condition number (the ratio of the largest to the smallest singular value) is a measure of the sensitivity of the solution of a system of linear equations to errors in the input data. If two columns of the corresponding matrix are nearly identical, the matrix is ill-conditioned, the condition number is large, and errors will be amplified. Thus the explanation for the observation that $\mathbf{G}_{hp}$ tends to be more ill-conditioned than $\mathbf{G}_{hv}$ is that the elements of the former are Green's function, whereas the elements of the latter are the *derivative* of Green's function. The derivative is much more sensitive to small changes in the argument of the function, $|\mathbf{r}_{hm}-\mathbf{r}_{on}|$, in agreement with the fact that whereas the pressure is inversely proportional to the distance to a monopole, the particle velocity is inversely proportional to the square of the distance very near the source.

In order to reduce the influence of the measurement errors, Tikhonov regularization is used to stabilize the computational process; and the regularization parameter is chosen by the L-curve method.[22]

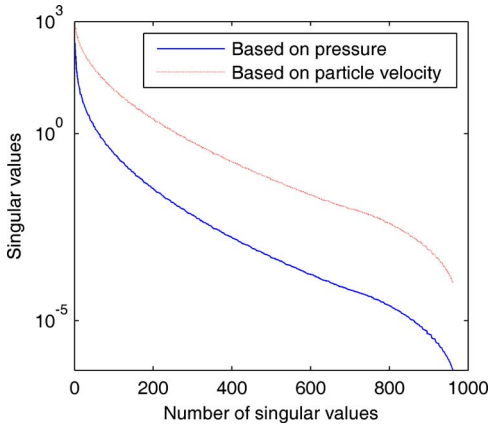


FIG. 2. (Color online) Singular values of the transfer matrices at 200 Hz.

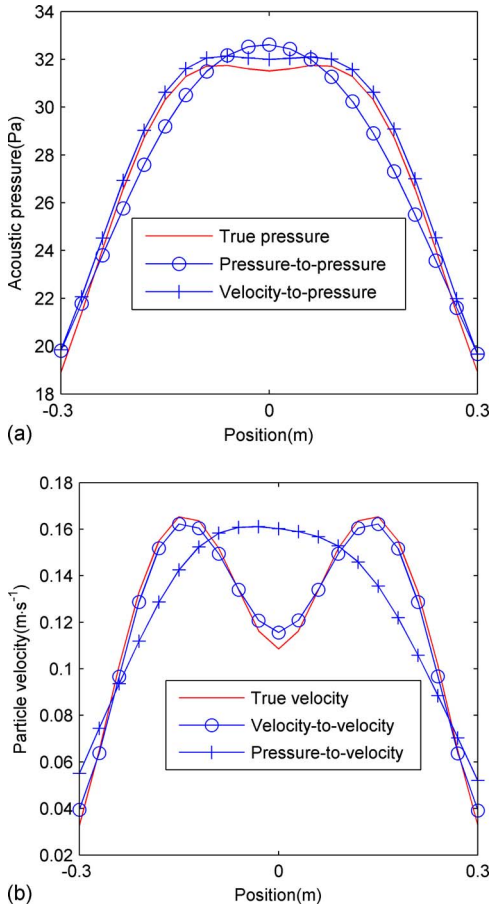


FIG. 3. (Color online) True and predicted pressure (a) and particle velocity (b) in the reconstruction plane, generated by a baffled vibrating panel at 100 Hz.

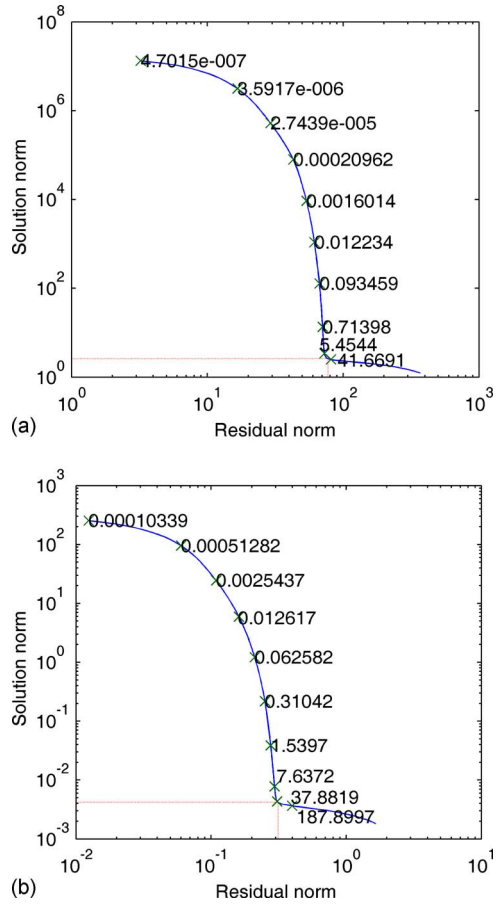


FIG. 4. (Color online) L-curves for determination of regularization parameter based on pressure (a) and on particle velocity (b). The position of the selected parameter is marked by dotted lines.

III. A SIMULATION STUDY

The condition number ratio found in the foregoing seems to indicate that it should be advantageous to measure the normal component of the particle velocity rather than the pressure in the hologram plane. To examine the matter a simulation study has been carried out. The test case was a point driven simply supported 3-mm-thick aluminum plate mounted in an infinite baffle. The dimensions, grid, and positions of the vibrating plate, the hologram plane, and the equivalent source plane were all the same as in the example presented above. The excitation of the plate was a harmonic force with an amplitude of 100 N acting at the center of the plate. The displacement and the normal velocity on the surface of the plate were calculated by modal superposition, and the radiated sound field was calculated from a numerical approximation to Rayleigh's first integral.[23] The reconstructed plane was located at $z=0.03$ m. In what follows, the "true" data have been calculated from the numerical approximation to Rayleigh's integral, and the reconstructed values have been calculated from the "measured" pressure or particle velocity.

A. ESM based on measurement of pressure or particle velocity

Figure 3(a) compares the true pressure along the x-axis at 100 Hz with reconstructions based on the pressure and

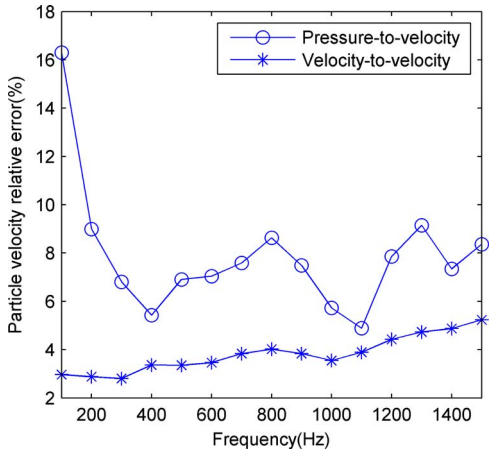


FIG. 5. (Color online) Relative error of reconstructed particle velocity.

based on the normal component of the particle velocity. To make the simulation study more realistic noise has been added to the measured data corresponding to a signal-to-noise ratio of 20 dB. It can be seen that the pressure reconstructed from velocity data is in better agreement with the true pressure than the pressure reconstructed from pressure data, which demonstrates that the method is less sensitive to

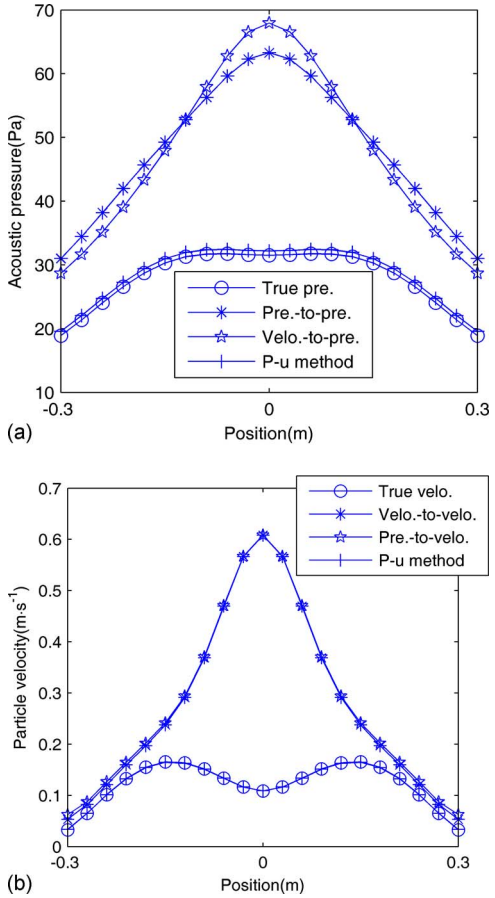


FIG. 6. (Color online) True and predicted undisturbed pressure (a) and particle velocity (b) in the reconstruction plane at 100 Hz. The primary source is a vibrating panel in a baffle, and the disturbing source is a monopole.

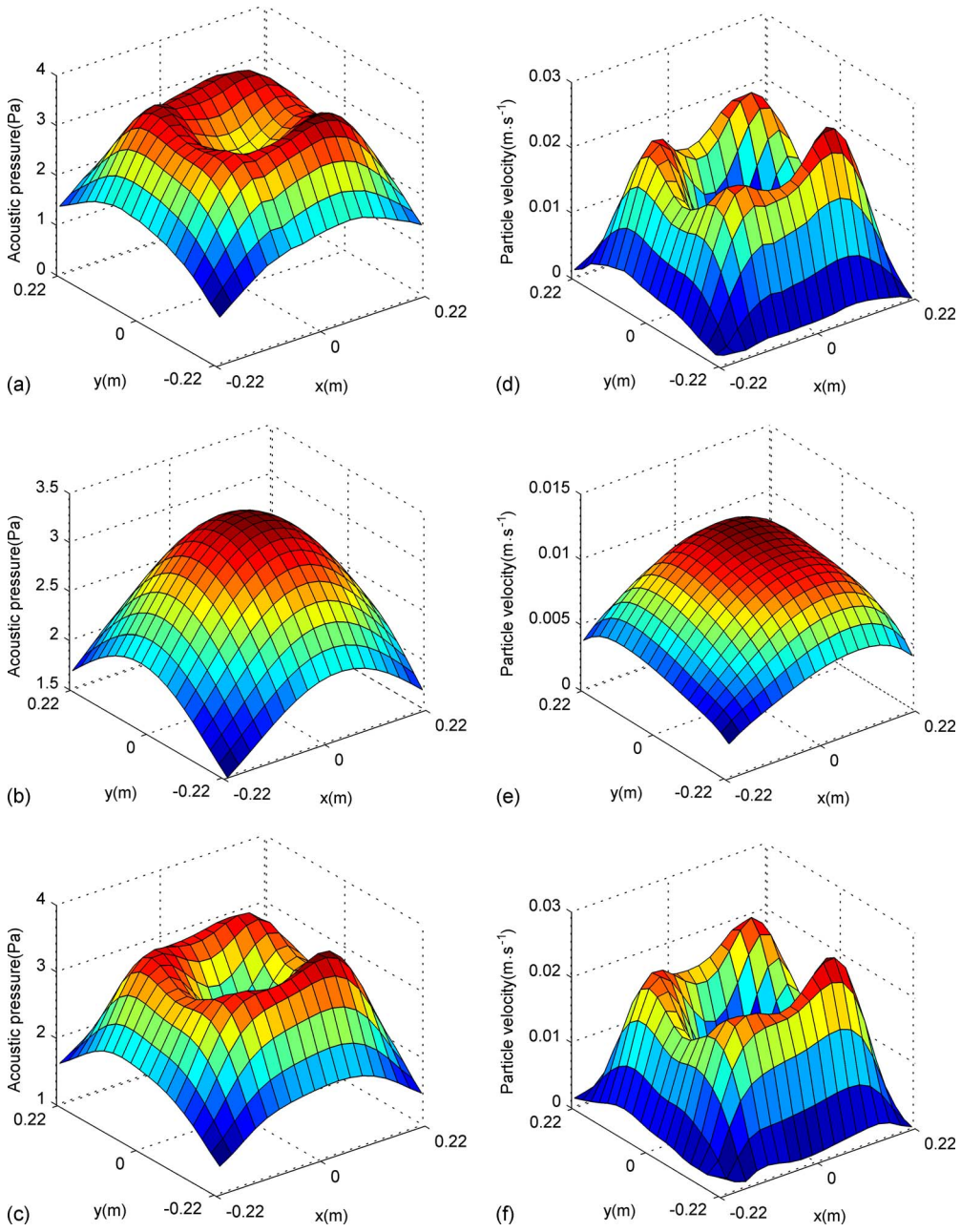


FIG. 7. (Color online) Sound field generated by a vibrating panel at 160 Hz. True pressure (a) and pressure predicted from pressure (b) and from particle velocity (c), and true particle velocity (d) and particle velocity predicted from pressure (e) and from particle velocity (f).

measurement errors if it is based on velocity data. This tendency is demonstrated even more clearly in Fig. 3(b), in which the true velocity and reconstructions based on the pressure and on the normal component of the particle velocity are compared; the reconstruction based on the particle velocity is by far the best. The L-curves for this case are shown in Fig. 4. The regularization parameters based on the pressure and on the velocity are both chosen reasonably near the corner of the curves. Similar results (not shown) have been found at other frequencies. Figure 5 shows the relative error as a function of the frequency. This quantity is defined as

$$\eta = \left(\sqrt{\sum_{i=1}^{N} |v_i - \hat{v}_i|^2} \Big/ \sqrt{\sum_{i=1}^{N} |v_i|^2} \right) \times 100\%, \qquad (17)$$

where N is the number of points on the reconstructed plane, and v_i and $\hat{v}_i$ are true and reconstructed normal velocities of the ith point. It is apparent that velocity-to-velocity results are better than pressure-to-velocity results in the entire frequency range, although the difference decreases with the frequency in agreement with the error analysis presented above.

B. ESM based on the *p-u* method

The *p-u* method is examined with a sound field generated by the same plate as described above and a monopole located at (0, 0, 0.2) (coordinates in meter), that is, on the other side of the hologram plane. (Reflections in the baffled panel have been ignored.) The disturbing source generates a higher pressure and a larger particle velocity than the primary source, the difference being up to 40%. The corresponding reconstructions in the prediction plane are shown in Figs. 6(a) and 6(b), which demonstrate that reconstructions based only on pressure or particle velocity are completely wrong, whereas the *p-u* method successfully separates the sound fields and gives results in good agreement with the true undisturbed values.

IV. EXPERIMENTAL RESULTS

Two experiments have been carried out in a large anechoic room at the Technical University of Denmark. In the first experiment the source was a 3-mm aluminum plate with dimensions of 44×44 cm^2 mounted as one of the surfaces of a box of heavy fiberboard and excited by a loudspeaker inside the box. The sound pressure and the particle velocity were measured at 18×19 points in two planes of dimensions 42.5×45 cm^2 using a $\frac{1}{2}$-in. *p-u* intensity probe produced by Microflown. The transducer was calibrated as described in Ref. 24. The two measurement planes were 8 and 4.5 cm from the plate, and the measured data in the plane nearest the source were regarded as the true reference data. The equivalent sources were distributed in a plane 3 cm behind the plate. A Brüel & Kjær (B&K) "PULSE" analyzer (type 3560) was used for measuring the frequency responses between the pressure and particle velocity signals from the transducer and the signal generated by the PULSE analyzer (pseudorandom noise) for driving the source.

Figure 7 shows the results at 160 Hz. Parts (a)–(c) in the left column show the true pressure and the pressure predicted from measurements of pressure and predicted from measurements of particle velocity. It can be seen that the reconstruction based on the particle velocity is much better than the reconstruction based on the pressure. Parts (d)–(f) in the right column show the true particle velocity and the particle velocity predicted from measurements of pressure and predicted from measurements of particle velocity. It is apparent that the particle velocity reconstructed from the pressure is not very accurate, whereas the reconstruction based on measurement of particle velocity is far better. Similar results (not shown) have been obtained at other frequencies. In some cases (not shown) the pressure-to-pressure reconstruction was found to be slightly better than the velocity-to-pressure reconstruction, though, but velocity-to-velocity reconstruction was invariably found to be considerably better than the pressure-to-velocity reconstruction.

In the second experiment there were two sources. The primary source was a "coincident-source" loudspeaker unit produced by KEF mounted in a rigid plastic sphere with a diameter of 27 cm, and the disturbing source placed on the opposite side of the hologram plane was a B&K 4299 "volume velocity adaptor" (a 10-cm-long tube with an internal

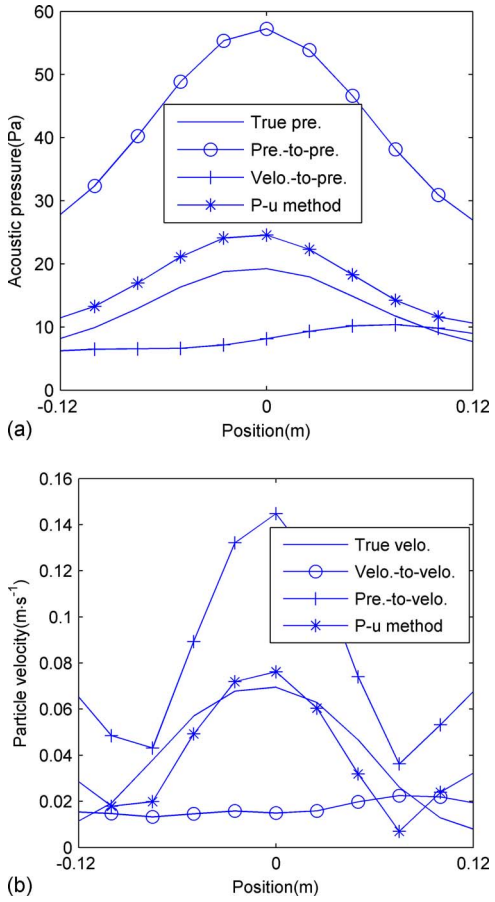


FIG. 8. (Color online) Sound field generated by a loudspeaker mounted in a sphere disturbed by sound from an experimental monopole at 528 Hz. True undisturbed pressure (a) and particle velocity (b) compared with predictions based on measurements of the pressure, the particle velocity, and both quantities using the *p-u* method.

diameter of 4 cm), mounted at the end of a long tube driven by loudspeaker (a B&K 4275 "OmniSource"). Because of its small opening this source is a good approximation to a monopole. The two sources were placed symmetrically with respect to the center of the measurement plane (of 25×25 cm^2) with a distance of 10 cm between them. The sound pressure and the particle velocity were measured at 11×11 points using the Microflown $\frac{1}{2}$-in. *p-u* intensity probe in the case when the two sources were operating together and the case when only the primary source was operating. The latter case provided the true reference data. In this case the hologram plane also served as the prediction plane, and the equivalent sources were placed 3 cm behind the front surface of the loudspeaker.

Figure 8 compares the true undisturbed pressure and particle velocity with predictions based on the pressure and based on the particle velocity, and predictions based on the *p-u* method at 528 Hz. The reconstructions based only on pressure or particle velocity give wrong results, but the reconstructions using the *p-u* method agree well with the true values. Similar results (not shown) have been found at other frequencies. However, because of resonances in the tube driven by the OmniSource the level of the disturbing sound varied strongly with the frequency, and at frequencies where this source was relatively weak compared with the primary

source reconstructions based on measurement of the particle velocity (not shown) were better than reconstructions based on the *p-u* method. The observation that the advantage of the *p-u* method vanishes when the disturbing sound is relatively weak has also been made with the statistically optimized version of NAH.[25] The explanation is that the *p-u* method relies critically on pressure- and particle velocity-based estimates being identical.

V. CONCLUSIONS

NAH based on the ESM and measurements with pressure-velocity transducers has been examined. Error sensitivity considerations demonstrate an advantage in using the normal component of the particle velocity rather than the pressure in the hologram plane as the input, and this advantage has been confirmed both by a simulation study and by experimental results. A variant of the method that combines pressure- and particle velocity-based estimates has also been examined and shown to perform well. This method makes it possible to distinguish between sounds coming from the two sides of the hologram plane.

ACKNOWLEDGMENTS

The authors would like to thank Microflown for lending a *p-u* sound intensity probe. This work was supported by the National Natural Science Foundation of China (Grant Nos. 50675056 and 10874037), the Fok Ying Tung Education Foundation (Grant No. 111058), and the Program for New Century Excellent Talents in University (Grant No. NCET-08-0767). Additionally, the China Scholarship Council is acknowledged for financial support.

[1]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[2]W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH): II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[3]E. G. Williams, H. D. Dardy, and K. B. Washburn, "Generalized nearfield acoustical holography for cylindrical geometry: Theory and experiment," J. Acoust. Soc. Am. **81**, 389–407 (1987).
[4]M. R. Bai, "Application of BEM (boundary element method)-based acoustic holography to radiation analysis of sound sources with arbitrarily shaped geometries," J. Acoust. Soc. Am. **92**, 533–549 (1992).
[5]B.-K. Kim and J.-G. Ih, "On the reconstruction of the vibro-acoustic field over the surface enclosing an interior space using the boundary element method," J. Acoust. Soc. Am. **100**, 3003–3016 (1996).
[6]S.-C. Kang and J.-G. Ih, "Use of nonsingular boundary integral formulation for reducing errors due to near-field measurements in the boundary element method based near-field acoustic holography," J. Acoust. Soc. Am. **109**, 1320–1328 (2001).
[7]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Int. J. Acoust. Vib. **6**, 83–89 (2001).
[8]Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized nearfield acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).
[9]Z. Wang and S. F. Wu, "Helmholtz equation-least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[10]S. F. Wu and J. Yu, "Reconstructing interior acoustic pressure fields via Helmholtz equation least-squares method," J. Acoust. Soc. Am. **104**, 2054–2060 (1998).
[11]G. H. Koopman, L. Song, and J. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[12]L. Song, G. H. Koopman, and J. Fahnline, "Numerical errors associated with the method of superposition for computing acoustic fields," J. Acoust. Soc. Am. **89**, 2625–2633 (1991).
[13]A. Sarkissian, "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**, 1593–1596 (2004).
[14]A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).
[15]C.-X. Bi, X.-Z. Chen, and J. Chen, "Nearfield acoustic holography based on the equivalent source method," Sci. China, Ser. E: Technol. Sci. **48**, 338–353 (2005).
[16]R. Jeans and I. C. Mathews, "The wave superposition method as a robust technique for computing acoustic fields," J. Acoust. Soc. Am. **92**, 1156–1166 (1992).
[17]F. Jacobsen and H.-E. de Bree, "A comparison of two different sound intensity measurement principles," J. Acoust. Soc. Am. **118**, 1510–1517 (2005).
[18]F. Jacobsen and Y. Liu, "Near field acoustic holography with particle velocity transducers," J. Acoust. Soc. Am. **118**, 3139–3144 (2005).
[19]F. Jacobsen and V. Jaud, "Statistically optimized near field acoustic holography using an array of pressure-velocity probe," J. Acoust. Soc. Am. **121**, 1550–1558 (2007).
[20]C.-X. Bi, X.-Z. Chen, and J. Chen, "Sound field separation technique based on equivalent source method and its application in nearfield acoustic holography," J. Acoust. Soc. Am. **123**, 1472–1478 (2008).
[21]C.-X. Bi and X.-Z. Chen, "Sound field separation technique based on equivalent source method using pressure-velocity measurements and its application in nearfield acoustic holography," Proceedings of Inter-Noise 2008, Shanghai, China.
[22]P. C. Hansen and D. P. O' Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," SIAM J. Sci. Comput. (USA) **14**, 1487–1503 (1993).
[23]E. G. Williams, *Fourier Acoustics—Sound Radiation and Nearfield Acoustical Holography* (Academic Press, San Diego, 1999).
[24]F. Jacobsen and V. Jaud, "A note on the calibration of pressure-velocity sound intensity probes," J. Acoust. Soc. Am. **120**, 830–837 (2006).
[25]F. Jacobsen, X. Chen, and V. Jaud, "A comparison of statistically optimized near field acoustic holography using single layer pressure-velocity measurements and using double layer pressure measurements," J. Acoust. Soc. Am. **123**, 1842–1845 (2008).

Patch near-field acoustic holography: Regularized extension and statistically optimized methods

Jean-Claude Pascal,[a] Sébastien Paillasseur, and Jean-Hugh Thomas
Laboratoire d'Acoustique de l'Université du Maine (CNRS UMR 6613) and Ecole Nationale Supérieure d'Ingénieurs du Mans (ENSIM), Université du Maine, rue Aristote, 72000 Le Mans, France

Jing-Fang Li
Visual VibroAcoustics, 51 rue d'Alger, 72000 Le Mans, France

(Received 29 September 2008; revised 10 June 2009; accepted 7 July 2009)

The patch holography method allows one to make measurements on an extended structure using a small microphone array. Increased attention has been paid to the two techniques, which are quite different at first glance. One is to extrapolate the pressure field measured on the hologram plane while the other is to use statistically optimized processing. A singular value decomposition formulation of the latter is proposed in this paper. The similarity of the two techniques is shown here. Both use a convolution of the measured pressure patch to obtain a better estimate of the wavenumber spectrum backward propagated on the structure. By using the Morozov discrepancy principle to compute the regularization parameter, the two methods lead to very close results.
 [DOI: 10.1121/1.3192349]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.20.Ye, 43.40.Sk [EGW] Pages: 1264–1268

I. INTRODUCTION

A fundamental assumption of planar near-field acoustic holography (NAH) is that the measurement plane in the near-field of sources is infinite. In practice it is enough that the measurement plane is significantly larger than the radiating surface so that the field reconstruction can be done under good conditions. As soon as the dimensions of the hologram decrease, the truncation effect of the measured field creates distortions mainly due to the use of the discrete spatial Fourier transform (DSFT). These distortions can be reduced by using a Tukey window[1] but with the impossibility to use the whole reconstructed field, by selectively filtering the edge effects using wavelets,[2] or by regularizing an inversion technique of the transfer matrix representing the propagation,[3,4] eventually associated with a condensation method.[5,6] The analysis of an extended emitting area with a small microphone array is often found in applications of acoustic holography, and the fundamental assumption is rarely satisfied. To overcome this important limitation, methods known as "patch holography" have been proposed.[7–11] When the array (patch) is smaller than the radiating area, for example, in case of a large vibrating structure, these techniques allow one to reconstruct the field on the projected area from the hologram with a minimum of distortion caused by the edge effects and the sound field emitted by the surfaces not covered by the array.

Essentially two methods are used to solve this problem: (i) a method by which the field on the hologram is extrapolated over a larger area by using an iterative process,[7–9,12–14] and (ii) a method that uses statistical optimization for estimating the wavenumber spectrum from an acquisition on a small aperture hologram.[10,11,15–17]

The first method was introduced by Saijyou and Yoshikawa.[7] They showed that a sound field measured over the patch can be extended into the exterior region by using an iterative data restoration algorithm that increases the aperture size by limiting the bandwidth of the wavenumber spectrum. This method has been further optimized by Williams *et al.*[8,9] by using a modified Tikhonov filter with a regularization technique, and the discrete Fourier transform formulation has also been extended in terms of singular value decomposition (SVD). Since this technique was implemented for cylindrical geometries[12] and applied to recovery a source distribution from hologram pressures was measured over multiple unconnected patches.[14] Variants for performing patch NAH were also proposed: a one-step procedure using Tikhonov regularization with generalized cross validation[13] and methods using a sound field model in terms of spherical harmonics[18] or equivalent sources.[5,6]

The second technique introduced by Steiner and Hald[10] optimizes the NAH process by realizing a spatial convolution to have a wavenumber spectrum produced by only the source region covered by the patch. The spatial convolution is obtained by imposing constraints on the wavenumber spectrum. This method has been adapted to different experimental configurations[11,16,17] and cylindrical geometries.[15]

The scope of this paper is to compare the two methods. So the processing algorithms are formulated with the same notations, then the performances of the two methods are illustrated by a simple example.

II. SOME NAH PROBLEMS

Consider a planar surface Γ_s corresponding to the source area on which the pressure and normal velocity fields are,

[a]Author to whom correspondence should be addressed. Electronic mail: jean-claude.pascal@univ-lemans.fr

respectively, $p(\mathbf{y})$ and $u_z(\mathbf{y})$. Another surface Ω is located at a small distance d in the near-field, on which the acoustic pressure $p(\mathbf{x})$ is measured. The discrete space Fourier transform of the Helmholtz equation allows one to write as

$$\mathbf{p}(\mathbf{x}) = \mathbf{W}\mathbf{P}(\mathbf{k}) = \begin{cases} \mathbf{W}\mathbf{G}_N\mathbf{W}^+\mathbf{u}(\mathbf{y}) \\ \mathbf{W}\mathbf{G}_D\mathbf{W}^+\mathbf{p}(\mathbf{y}), \end{cases} \quad (1)$$

where $\mathbf{p}(\mathbf{x})$ is the column vector of M elements of the measured pressure on the mesh of a finite surface Ω, and $\mathbf{u}(\mathbf{y})$ and $\mathbf{p}(\mathbf{y})$ are, respectively, the vectors of velocity and pressure on Γ_s. $\mathbf{W}$ is the matrix of the backward Fourier transform that allows one to obtain the vector of pressure $\mathbf{p}(\mathbf{x})$ from its discrete wavenumber spectrum denoted by the vector $\mathbf{P}(\mathbf{k})$. The elements of the matrix $\mathbf{W}$ are written by $W_{mn} = \exp(-j\mathbf{k}_n \cdot \mathbf{x}_m)/\sqrt{M}$. The forward Fourier transform is defined here as the generalized inverse $\mathbf{W}^+$ of the backward Fourier transform, such that $\mathbf{W}\mathbf{W}^+ = \mathbf{I}$. This definition allows one to obtain an estimate of wavenumber spectrum from an irregular sampling of the hologram,[19] as it is done for the statistically optimal near-field acoustic holography (SONAH) method.[11] In this case, the matrix of the backward transform used for the inverse of matrix corresponds to an overdetermined system, which leads to the estimate of the wavenumber spectrum in the least square sense. In the particular case of a regular grid and a square matrix (as used in the examples of Sec. IV), the authors have the equivalent equation $\mathbf{W}^+ \equiv \mathbf{W}^H$. In practice, the matrix $\mathbf{W}$ used to recover the field from the wavenumber spectrum is often different from that which is used to calculate the forward Fourier transform. In fact, independently of the measurement mesh, the field is presented on a regular grid of which the number of points is strongly increased to visually get a better resolution. $\mathbf{W}$ is a rectangular matrix with the number of rows (points of the field) being much more significant than the number of columns (points of the wavenumber spectrum). This operation corresponds to a Shannon interpolation and no more information is added. This method gets the same results as the iterative procedure proposed in Ref. 14, while being much faster.

The diagonal matrices $\mathbf{G}_N$ and $\mathbf{G}_D$ are the propagators, the elements of which are expressed as $G_{N,n} = -\rho c k e^{-jk_z d}/k_z$ and $G_{D,n} = e^{-jk_z d}$, respectively, with $k_z = \sqrt{k^2 - \mathbf{k}_n^2}$ ($k = \omega/c$ and $\sqrt{-1} = -j$). The inversion of Eq. (1) allows one to reconstruct with a good accuracy, the pressure $\mathbf{p}(\mathbf{y})$ and the velocity $\mathbf{u}(\mathbf{y})$ from the measured pressure $\mathbf{p}(\mathbf{x})$ provided that the hologram is larger than the source region and that the evanescent waves amplified by the inverse propagator are filtered by a low-pass filter represented by the diagonal matrix $\mathbf{F}_\alpha = \mathrm{diag}(F_\alpha)$ as follows:[20,21]

$$\tilde{\mathbf{u}}(\mathbf{y}) = \mathbf{W}\mathbf{G}_N^{-1}\mathbf{F}_\alpha\mathbf{W}^+\mathbf{p}(\mathbf{x}), \quad (2)$$

$$\tilde{\mathbf{p}}(\mathbf{y}) = \mathbf{W}\mathbf{G}_D^{-1}\mathbf{F}_\alpha\mathbf{W}^+\mathbf{p}(\mathbf{x}). \quad (3)$$

The reconstruction of $\tilde{\mathbf{u}}(\mathbf{y})$ and $\tilde{\mathbf{p}}(\mathbf{y})$ is thus smoothed by the removal of the components whose spatial oscillations have short wavelengths. Williams *et al.*[1,9] described in detail the consequences of the discretization in the spatial and wavenumber domains and the role of the spatial periodicity of the pressure $\mathbf{p}(\mathbf{x})$ on adjacent regions to Ω (aperture replication problem). The operation of reconstruction of the velocity on Γ_s is a convolution of this periodic pressure field, which causes an expansion and a recovering of the border areas on adjacent regions. The authors now consider that the measurement surface Ω is reduced to a patch Ω_p. This patch is smaller than the source area Γ_s (see Fig. 1). Γ_p is the normal projection of the patch on the source plane where the pressure and velocity fields are to be reconstructed. The reduction in the hologram size makes the problem of spatial periodicity more critical. The technique of zero-padding has long been used as a solution[1] for the effect of convolution mentioned above. It does not, however, allow one to reduce the influence of the part of Γ_s outside Γ_p or the edge discontinuity effects.

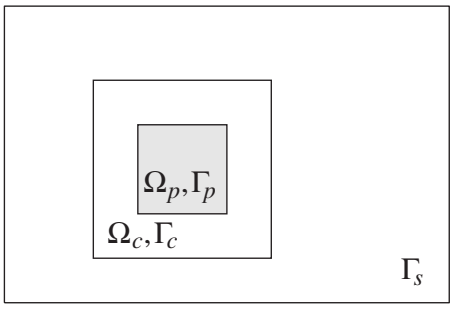


FIG. 1. Ω_p, pressure measurement surface (patch); Ω_c, band around the patch; Γ_p and Γ_c, corresponding to normal projection area on source surface; and Γ_s, whole radiating surface.

III. PATCH NAH METHODS

The objective is, in the case of an extended source region Γ_s, to reconstruct as accurately as possible the field on Γ_p from the pressure measurements on Ω_p (see Fig. 1).

A. Iterative method for extension of the hologram

The iteration method for extension of the hologram was proposed by Saijyou and Yoshikawa[7] and then improved by Williams *et al.*[8,9] who also showed that DSFT and SVD approaches provide comparable results. It is the DSFT approach that is used here. The method is to extrapolate the pressure $\mathbf{p}(\mathbf{x} \in \Omega_p)$ over the domain Ω_c. Initially the band Ω_c is filled with zeros. The zero-padding technique is implemented using a $M_1 \times M$ rectangular matrix $\mathbf{R}$ with elements equal to 1 or 0 (where $M_1 > M$), such as

$$\mathbf{p}_0(\mathbf{x}) = \mathbf{R}\mathbf{p}(\mathbf{x}) = \begin{cases} \mathbf{p}(\mathbf{x} \in \Omega_p) & \text{(measurements)} \\ 0(\mathbf{x} \in \Omega_c) & \text{(zero-padding)}. \end{cases} \quad (4)$$

The transpose of the matrix $\mathbf{R}$ allows one to extract the pressure on the surface of the patch Ω_p: $\mathbf{p}(\mathbf{x}) = \mathbf{R}^T\mathbf{p}_0(\mathbf{x})$. After zero-padding and filtering, the smoothed pressure

$$\tilde{\mathbf{p}}_0(\mathbf{x}) = \mathbf{W}\mathbf{F}_{N,\alpha}\mathbf{W}^+\mathbf{p}_0(\mathbf{x}) = \mathbf{W}\mathbf{F}_{N,\alpha}\mathbf{W}^+\mathbf{R}\mathbf{p}(\mathbf{x}) \quad (5)$$

is extended on the whole field $\Omega_p \cup \Omega_c$ by the effect of a low-pass filter in the wavenumber domain represented by the diagonal matrix $\mathbf{F}_{N,\alpha}$. To refine the process of extrapolation, several iterations are used as follows:

$$\tilde{\mathbf{p}}_{i+1}(\mathbf{x}) = \mathbf{W}\mathbf{F}_{N,\alpha_i}\mathbf{W}^{+}[(\mathbf{I}-\mathbf{R}\mathbf{R}^{T})\tilde{\mathbf{p}}_i(\mathbf{x}) + \mathbf{R}\mathbf{p}(\mathbf{x})], \tag{6}$$

where $\mathbf{I}-\mathbf{R}\mathbf{R}^T$ is a diagonal matrix that selects the points on the band Ω_c. The iteration procedure starts with $i=0$ and $\tilde{\mathbf{p}}_0(x)$ given by Eq. (5). The process is stopped at $i=I$ when the desired convergence $\|\tilde{\mathbf{p}}_{i+1}-\tilde{\mathbf{p}}_i\| < \varepsilon$ is reached. For each new iteration, the measured pressure field on the patch $\mathbf{p}(\mathbf{x} \in \Omega_p)$ replaces the corresponding part of the smoothed field $\tilde{\mathbf{p}}_i(x)$, without changing the estimate in Ω_c. A fundamental point of the method[8,9] is the use of the modified Tikhonov regularization filter[21]

$$F_{N,\alpha_i}(\mathbf{k}_n) = \lambda_n^2[\lambda_n^2 + \alpha_i(\alpha_i/[\alpha_i + \lambda_n^2])^2]^{-1}, \tag{7}$$

where $\lambda_n^2=|G_N(\mathbf{k}_n)|^2$ for $\mathbf{F}_{N,\alpha}$. This filter contains the information about the propagation process and depends on the distance d separating the Ω_p and Γ_p planes. The parameter α_i is updated at each iteration according to the standard deviation σ_i of the noise, which is estimated in the wavenumber domain with the assumption that beyond a cut-off value $\mathbf{k}^2 > k_c^2$, the spectrum only holds noise

$$\sigma_i \approx \|\mathbf{D}\mathbf{P}_i(\mathbf{k})\|_F/\|\mathbf{D}\|_F, \tag{8}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\mathbf{D}=\mathrm{diag}(D_n)$, with $D_n(\mathbf{k})=1$ when $\mathbf{k}^2 \geq k_c^2$ and $D_n(\mathbf{k})=0$ otherwise. k_c corresponds to the maximum wavenumber in agreement with the sampling criterion.[21] According to Eq. (6), the non-smoothed pressure is $\mathbf{p}_i(\mathbf{x})=(\mathbf{I}-\mathbf{R}\mathbf{R}^T)\tilde{\mathbf{p}}_{i-1}(\mathbf{x})+\mathbf{R}\mathbf{p}(\mathbf{x})$. Once the standard deviation is computed, an automatic selection of the regularization parameter α_i is obtained by verifying the following relation called the Morozov discrepancy principle:

$$\|\tilde{\mathbf{p}}_i(\mathbf{x}) - \mathbf{p}_i(\mathbf{x})\|_F/\sqrt{M_1} = \|(\mathbf{F}_{N,\alpha_i}-\mathbf{I})\mathbf{P}_i(\mathbf{k})\|_F/\sqrt{M_1} \equiv \sigma_i. \tag{9}$$

The regularization parameter depending on the bounds of the exact solution is not known in advance. A classical strategy due to Morozov determines this regularization parameter by solving a non-linear scalar equation,[22] in this case by an iterative computation. Morozov's principle is established when the discrepancy of the corresponding regularized solution is just equal to the measurement error. Williams[21] described the use of this principle to determine the regularization parameter from the estimated noise by considering the wavenumber spectrum outside the circle of radius k_c. In this procedure noise is considered in a general sense; it also includes the distortions due to the truncation of the field and the errors of spatial undersampling (Shannon's criterion is never ensured in NAH). Finally the reconstructed velocity on Γ_p can be written by

$$\tilde{\mathbf{u}}(\mathbf{y}) = \mathbf{R}^T\mathbf{W}\mathbf{G}_N^{-1}\mathbf{W}^{+}\tilde{\mathbf{p}}_I(\mathbf{x}). \tag{10}$$

The filter of Eq. (7) uses the propagator G_N that supposes that the extrapolated field is optimized for reconstructing the velocity by using Eq. (2). To reconstruct the pressure using Eq. (3), it will be necessary to substitute G_D for G_N and $\mathbf{F}_{D,\alpha}$ for $\mathbf{F}_{N,\alpha}$ in the iteration procedure of Eq. (6).

B. Method of the statistically optimized backpropagation

The method of the statistically optimized backpropagation was proposed by Hald and co-worker[10,11] for the plane geometry under the name of SONAH then adapted to cylindrical geometries[15] and to the reconstruction of other quantities.[16,17] In this method, the transfer matrix $\mathbf{H}_D$ used to compute the pressure on the source plane from a small hologram aperture

$$\mathbf{p}(\mathbf{y}) = \mathbf{H}_D\mathbf{p}(\mathbf{x}) \tag{11}$$

is obtained using the approach completely different from that given by Eq. (3). An optimized solution is searched to correct, as well as possible, all the distortions of the reconstruction process such as small size of the patch, discontinuity at the edges, and influence of noise in the restoration of evanescent waves. For this purpose, a set of elementary solutions for which one knows the correspondence on the Ω_p and Γ_p planes is employed. In fact, these elementary solutions are projections of plane waves whose angle of incidence is associated with each point $\mathbf{k}_n$ of the wavenumber spectrum. For example, $\mathbf{H}_D$ should satisfy the following expression:

$$\mathbf{p}_n(\mathbf{y}) = \mathbf{H}_D\mathbf{p}_n(\mathbf{x}), \tag{12}$$

where $\mathbf{p}_n(\mathbf{y})=[e^{-j\mathbf{k}_n\cdot\mathbf{y}}]=\mathbf{R}^T\mathbf{W}\sqrt{M}\ \mathrm{diag}(\delta_{\ell n})$ and $\mathbf{p}_n(\mathbf{x})=[e^{-jk_z d}e^{-j\mathbf{k}_n\cdot\mathbf{x}}]=\mathbf{R}^T\mathbf{W}\mathbf{G}_D\sqrt{M}\ \mathrm{diag}(\delta_{\ell n})$, with $k_z=\sqrt{k^2-\mathbf{k}_n^2}$. $\mathrm{diag}(\delta_{\ell n})$ is a diagonal matrix having only one element equal to 1 (when $\ell=n$, $\delta_{\ell n}=1$) corresponding to the plane wave with wavenumber vector $\mathbf{k}_n$. All other points of the wavenumber spectrum must also satisfy Eq. (12). All these constraints are synthesized in a matrix form by the following equation:

$$[\mathbf{p}_1(\mathbf{y})\cdots\mathbf{p}_n(\mathbf{y})\cdots\mathbf{p}_N(\mathbf{y})] = \mathbf{H}_D[\mathbf{p}_1(\mathbf{x})\cdots\mathbf{p}_n(\mathbf{x})\cdots\mathbf{p}_N(\mathbf{x})],$$

$$\mathbf{R}^T\mathbf{W} = \mathbf{H}_D\mathbf{R}^T\mathbf{W}\mathbf{G}_D. \tag{13}$$

It is the transposed form of Eq. (13) that was used in Ref. 11. By introducing a $M\times N$ rectangular matrix $\mathbf{W}_R=\mathbf{R}^T\mathbf{W}$ (where $N>M$), Eq. (13) can be written as $\mathbf{W}_R=\mathbf{H}_D\mathbf{A}$, where $\mathbf{A}=\mathbf{W}_R\mathbf{G}_D$. The authors can see that the number of points of the wavenumber spectrum is larger than that on the patch. This important issue discussed below leads to a system to which the solution $\mathbf{H}_{D,\alpha}$ is obtained by computing the (regularized) Moore–Penrose generalized right inverse[23] of matrix $\mathbf{A}$,

$$\mathbf{H}_{D,\alpha} = \mathbf{W}_R\mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \alpha\mathbf{I})^{-1}. \tag{14}$$

Substituting Eq. (14) and the expression for $\mathbf{A}$ into Eq. (11) yields

$$\tilde{\mathbf{p}}(\mathbf{y}) = \mathbf{W}_R\mathbf{G}_D^H\mathbf{W}_R^H(\mathbf{W}_R\mathbf{G}_D\mathbf{G}_D^H\mathbf{W}_R^H + \alpha\mathbf{I})^{-1}\mathbf{p}(\mathbf{x}). \tag{15}$$

Each element of the matrix $\mathbf{A}\mathbf{A}^H=\mathbf{W}_R\mathbf{G}_D\mathbf{G}_D^H\mathbf{W}_R^H$ represents the sum of the whole wavenumber spectrum. If $\mathbf{W}_R$ is a square matrix, the operation is an ordinary DSFT on the patch, without optimization. It is the number of points of the K-spectrum higher than the number of points of the hologram ($N>M$), which allows an optimization for the small size of the patch by increasing the number of constraints

imposed on the wavenumber spectrum. By considering that $N \to \infty$, Hald[11] gave analytical expressions for the elements of the matrix $\mathbf{AA}^H$. The regularization factor can be obtained in the same way as for the iterative method, by estimating the standard deviation of the noise using Eq. (8) and by evaluating α by means of the Morozov discrepancy principle.

An alternative formulation of the SONAH method can be obtained by carrying out the SVD of matrix $\mathbf{A}$, such as $\mathbf{A}=\mathbf{U}\ \mathrm{diag}(\lambda_n)\mathbf{V}^H$. Substituting this expression into Eq. (14) results in an expression equivalent to Eq. (15) as follows:

$$\tilde{\mathbf{p}}(\mathbf{y}) = \mathbf{W}_R \mathbf{V} \mathbf{F}_\alpha\ \mathrm{diag}(\lambda_n^{-1}) \mathbf{U}^H \mathbf{p}(\mathbf{x}), \tag{16}$$

where $\mathbf{F}_\alpha$ is the diagonal matrix of a filter whose elements are given by $F_\alpha = \lambda_n^2/(\alpha+\lambda_n^2)$, where λ_n^2 are the eigenvalues of the decomposition of the matrix $\mathbf{A}$. It should be noted that unlike the use of the SVD to compute the inverse of the projection matrix between the hologram and the source plane,[3,13,21] here the inverse is applied only to the model of the transfer matrix between the points of the patch and those of the wavenumber spectrum backpropagated to the Γ_s plane. A statistically optimized solution for the reconstruction of the velocity can be obtained in the same way by substituting G_N for G_D, but because of the bad conditioning introduced by the use of propagator G_N, the choice of the direct calculation of the derivative of Eq. (15) was often made, as was done in Ref. 16.

IV. RESULTS AND DISCUSSION

A. Common expressions for the two techniques

It is noteworthy that from the previous results [Eqs. (10), (15), and (16)], the two methods can be presented in one formulation as follows:

$$\tilde{\mathbf{s}}(\mathbf{y}) = \mathbf{R}^T \mathbf{W} \tilde{\mathbf{S}}(\mathbf{k}) \quad \text{with} \quad \tilde{\mathbf{S}}(\mathbf{k}) = \mathbf{C}_{\eta,\alpha} \mathbf{p}(\mathbf{x}), \tag{17}$$

where $\tilde{\mathbf{s}}(\mathbf{y})$ can be the pressure $\tilde{\mathbf{p}}(\mathbf{y})$ or the velocity $\tilde{\mathbf{u}}(\mathbf{y})$ on the area Γ_p. In all cases, the matrix $\mathbf{C}_{\eta,\alpha}$ represents a convolution of the pressure $\mathbf{p}(\mathbf{x})$ on the patch Ω_p followed by a transformation that allows one to estimate the regularized wavenumber spectrum $\tilde{\mathbf{S}}(\mathbf{k})$ of the velocity or the pressure on Γ_p (η could be D or N according to the propagator). For the method of extension of the patch, the convolution matrix is implicitly defined in recursive form by Eq. (6),

$$\mathbf{C}_{\eta,\alpha} \mathbf{p}(\mathbf{x}) = \mathbf{G}_\eta^{-1} \mathbf{W}^+ \tilde{\mathbf{p}}_I(\mathbf{x}). \tag{18}$$

In the case of the SONAH method, the convolution matrix is defined from Eqs. (15) and (16) as

$$\begin{aligned} \mathbf{C}_{\eta,\alpha} &= \mathbf{G}_\eta^H \mathbf{W}^H \mathbf{R} (\mathbf{R}^T \mathbf{W} \mathbf{G}_\eta \mathbf{G}_\eta^H \mathbf{W}^H \mathbf{R} + \alpha \mathbf{I})^{-1} \\ &= \mathbf{V} \mathbf{F}_\alpha\ \mathrm{diag}(\lambda_n^{-1}) \mathbf{U}^H, \end{aligned} \tag{19}$$

where η is the propagator of the pressure or the velocity adapting to the reconstruction process.

Both methods are clearly the techniques of optimization of the wavenumber spectrum reconstructed on the plane Γ_p, and the following example gives an illustration.

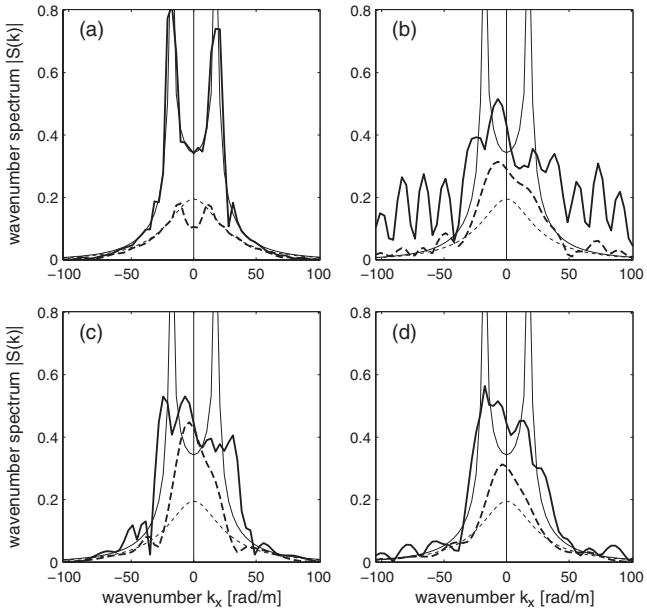


FIG. 2. Computed values (thick lines) and theoretic values obtained by Eq. (20) (thin lines) of cross-section slices of wavenumber spectra backpropagated from $d=3$ cm to the plane Γ_s for a point source located 5 cm from the hologram (1000 Hz, SNR: 40 dB) for $k_y=0$ (solid lines) and $k_y=1.5\ k$ (dashed lines). (a) Hologram of 60×60 grid points, Tukey window at 50%, modified Tikhonov filter and DSFT; (b) patch of 12×12 and zero-padding of 60×60 grid points, modified Tikhonov filter and DSFT; (c) patch of 12×12 and expansion on a 60×60 mesh with modified Tikhonov filter and 800 iterations; and (d) patch of 12×12, SONAH with regularization.

B. Illustration by a simple example

Consider a point source ($f=1000$ Hz) at $(x_0,y_0)=(5\ \mathrm{cm},\ 9\ \mathrm{cm})$ with respect to the origin at the center of the hologram. The hologram plane is at 5 cm from the source. The field is backpropagated 3 cm toward the source (the plane Γ_s is therefore at $z_0=2$ cm from the source center). A white noise with a signal-to-noise ratio (SNR) of 40 dB is added. Two cross-section slices of the wavenumber spectra on the plane Γ_s, versus k_x for $k_y=0$ (solid lines) and $k_y=1.5\ k$ (dashed lines) within the region of evanescent waves, are shown in Figs. 2(a)–2(d). The computed values are shown by thick lines. The theoretical wavenumber spectrum for the unit point source[24] obtained by

$$S(\mathbf{k}) = \frac{-\rho c k e^{-j(k_x x_0 + k_y y_0 + k_z z_0)}}{k_z} \tag{20}$$

is shown in Fig. 2 by thin lines. The reference K-spectrum shown in Fig. 2(a) is computed from 60×60 grid points with a spatial step size of 3 cm. A Tukey window at 50% is applied before Fourier transformation (DSFT), filtering, and backpropagation, according to Eq. (3). Figures 2(b)–2(d) show the results reconstructed from measurements on a patch of 12×12 grid points by the use of different processing that allows a K-spectrum on a 60×60 mesh to be obtained. The results shown in Fig. 2(b) is obtained by simply applying DSFT after zero-padding and filtering: $\mathbf{G}_D^{-1}\mathbf{F}_\alpha\mathbf{W}^+\mathbf{R}\mathbf{p}(\mathbf{x})$. For results shown in Fig. 2(c), the method of extension of the patch is employed [Eq. (18)] with 800 iterations of Eq. (6). To obtain the results shown in Figs. 2(a)–2(c), a modified Tikhonov filter [Eq. (7)] is used with a coefficient of regu-

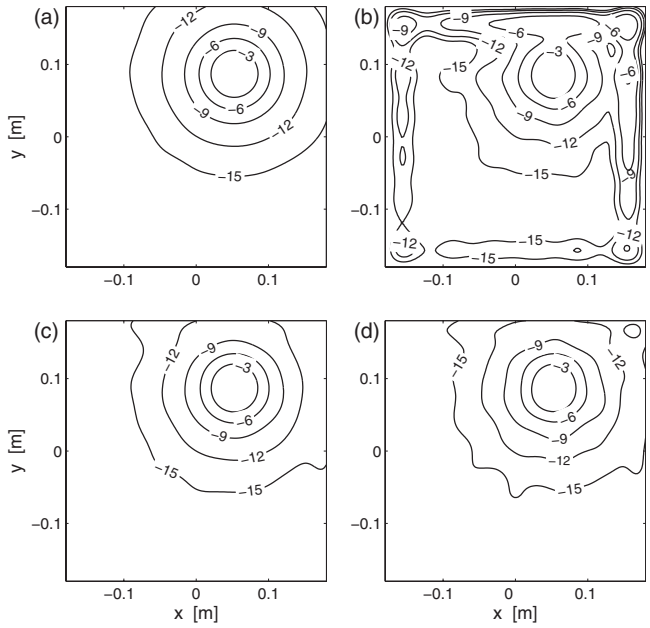


FIG. 3. Reconstruction on Γ_p area (normal projection of the patch) for the four processing methods shown in Fig. 2. Contours of pressure amplitude in decibels.

larization α determined by using Eqs. (8) and (9). Figure 2(d) corresponds to the SONAH method using the SVD formulation, but the two formulas of this method strictly check the equality of Eq. (19). The same number of points (60×60) is used for the wavenumber spectrum in the four configurations. Thus, in all the cases, the matrix $\mathbf{R}$ in Eqs. (6), (17), and (19) is of size $60^2 \times 12^2$. Finally, Fig. 3 shows the pressure on the normal projection Γ_p of the patch, by performing the backward Fourier transform of the K-spectra shown in Fig. 2. The important distortions of K-spectrum in Fig. 2(b) are caused by severe aliasing effects on the projected field [see Fig. 3(b)]. Although the cross-section slice of wavenumber spectra shown in Figs. 2(c) and 2(d) still deviates from the theoretical values or from the curves in Fig. 2(a), the projected fields shown in Figs. 3(c) and 3(d) allow one to obtain the pressure fields without too many distortions.

V. CONCLUSIONS

The same notation is used to express two methods of patch holography: the regularized extension method by iteration process and the statistically optimized method. A fundamental aspect is associated with the rectangular matrix $\mathbf{R}$, which describes the extension of the initial field by zero-padding for the iterative method and of the additional constraints imposed on the wavenumber spectrum by increasing the density of points. For the latter technique, an alternative formulation using the SVD is established. By determining the regularization parameter using the Morozov discrepancy principle, the methods of iterative expansion and SONAH provide similar results with pressure to pressure propagator. The iterative method is more expensive in computing times but seems to be more robust when the SNR decreases. Both make significant improvements to the standard NAH.

[1] E. G. Williams, *Fourier Acoustics* (Academic, London, 1999).
[2] J.-H. Thomas and J.-C. Pascal, "Wavelet preprocessing for lessening truncation effects in nearfield acoustical holography," J. Acoust. Soc. Am. **118**, 851–860 (2005).
[3] P. A. Nelson and S. H. Yoon, "Estimation of acoustic sources strength by inverse methods: Part I, Conditioning of the inverse problem," J. Sound Vib. **233**, 634–668 (2000).
[4] E. G. Williams, "Comparison of SVD and DFT approaches for NAH," Proceedings of the Inter-Noise 2002, Dearborn, MI, 19–21 August (2002).
[5] A. Sarkissian, "Method of superposition applied to patch near-field acoustic holography," J. Acoust. Soc. Am. **118**, 671–678 (2005).
[6] C. Bi, X. Chen, L. Xu, and J. Chen, "Patch nearfield acoustic holography based on the equivalent source method," Sci. China, Ser. E: Technol. Sci. **51**, 100–110 (2008).
[7] K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[8] E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[9] E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[10] R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Sixth International Congress on Sound and Vibration, Copenhagen, Denmark, 5–8 July (1999), pp. 843–850.
[11] J. Hald, "Planar near-field acoustical holography with arrays smaller than the sound source," 17th International Congress on Acoustics, Rome, Italy, 2–7 September (2001), Vol. **I**, Pt. A.
[12] M. Lee and J. S. Bolton, "Patch near-field acoustical holography in cylindrical geometry," J. Acoust. Soc. Am. **118**, 3721–3732 (2005).
[13] M. Lee and J. S. Bolton, "A one-step patch near-field acoustical holography procedure," J. Acoust. Soc. Am. **122**, 1662–1670 (2007).
[14] M. Lee and J. S. Bolton, "Reconstruction of source distributions from sound pressures measured over discontinuous regions: Multipatch holography and interpolation," J. Acoust. Soc. Am. **121**, 2086–2096 (2007).
[15] Y. T. Cho, J. S. Bolton, and J. Hald, "Source visualization by using statistically optimized near-field acoustical holography in cylindrical coordinates," J. Acoust. Soc. Am. **118**, 2355–2364 (2005).
[16] F. Jacobsen and V. Jaud, "Statistically optimized near field acoustic holography using an array of pressure-velocity probes," J. Acoust. Soc. Am. **121**, 1550–1558 (2007).
[17] J. Gomes, F. Jacobsen, and M. Bach-Anderson, "Statistically optimised near field acoustic holography and the Helmholtz equation least squares method: A comparison," Eighth International Conference on Theoretical and Computational Acoustics, Heraklion, Greece, 2–6 July (2007).
[18] Z. Wang and S. F. Wu, "Helmholtz equation–least-squares method for reconstructing the acoustic pressure field," J. Acoust. Soc. Am. **102**, 2020–2032 (1997).
[19] J. R. F. Arruda, "Analysis of non-equally spaced data using a regressive discrete Fourier series," J. Sound Vib. **156**, 571–574 (1992).
[20] B.-K. Kim and J.-G. Ih, "Design of an optimal wave-vector filter for enhancing the resolution of reconstructed source field by near-field acoustical holography (NAH)," J. Acoust. Soc. Am. **107**, 3289–3297 (2000).
[21] E. G. Williams, "Regularization method for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[22] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer-Verlag, New York, 1996), Chap. 2.
[23] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. (Johns Hopkins University Press, Baltimore, MD, 1996).
[24] C. H. Harrison and P. L. Nielsen, "Plane wave reflection coefficient from near field measurements," J. Acoust. Soc. Am. **116**, 1355–1361 (2004).

Efficient estimation of decay parameters in acoustically coupled-spaces using slice sampling[a)]

Tomislav Jasa
Thalgorithm Research, Toronto, Ontario L4X 1B1, Canada

Ning Xiang[b)]
Graduate Program in Architectural Acoustics, School of Architecture, Rensselaer Polytechnic Institute, Troy, New York 12180

(Received 1 September 2008; revised 27 May 2009; accepted 30 May 2009)

Room-acoustic energy decay analysis of acoustically coupled-spaces within the Bayesian framework has proven valuable for architectural acoustics applications. This paper describes an efficient algorithm termed slice sampling Monte Carlo (SSMC) for room-acoustic decay parameter estimation within the Bayesian framework. This work combines the SSMC algorithm and a fast search algorithm in order to efficiently determine decay parameters, their uncertainties, and inter-relationships with a minimum amount of required user tuning and interaction. The large variations in the posterior probability density functions over multidimensional parameter spaces imply that an adaptive exploration algorithm such as SSMC can have advantages over the exiting importance sampling Monte Carlo and Metropolis–Hastings Markov Chain Monte Carlo algorithms. This paper discusses implementation of the SSMC algorithm, its initialization, and convergence using experimental data measured from acoustically coupled-spaces.
 [DOI: 10.1121/1.3158934]

PACS number(s): 43.60.Uv, 43.60.Jn, 43.55.Mc [EJS] Pages: 1269–1279

I. INTRODUCTION

Bayesian probabilistic inference has been increasingly applied in acoustics applications ranging from architectural acoustics,[1,2] geo-acoustic inversion, source tracking,[3–6] and underwater acoustics applications.[7–9] The Bayesian formalism specifically applied to decay time evaluation in acoustically coupled-spaces has proven to be a useful framework for analyzing Schroeder decay functions[10] from room impulse response measurements. This framework allows one to estimate not only the decay parameters from the Schroeder decay model,[1] but also to determine the decay order,[2] quantify uncertainties of decay estimates, and determine the inter-relationship between multiple decay parameters.[11] Due to computational demands, it is common to use Markov chain Monte Carlo (MCMC) and Monte Carlo (MC) algorithms such as importance sampling Monte Carlo (ISMC) integration[3,6,11] for numerical calculation within the Bayesian framework. ISMC integration and MCMC algorithms represent effective approaches to estimate the decay parameters, quantify the estimate uncertainties, and determine decay inter-relationships in cases where it is possible to properly initialize these algorithms. Using an analytic example and sample posterior probability density functions (PPDFs) of acoustically coupled rooms, this paper discusses the difficulty in choosing a good ISMC sampling or MCMC proposal distributions, which often require significant user effort. A deterministic fast search (FS) algorithm,[12] which is less dependent on user initialization, is only able to estimate decay parameters; however, it cannot quantify uncertainties in the estimates nor can it determine inter-relationships between decay parameters. For data analysis, the uncertainties and inter-relationships of relevant parameters are of as the same importance as the parameters themselves.

This paper describes an efficient algorithm termed slice sampling Monte Carlo (SSMC), recently introduced by Neal,[13] as a generic sampling method. The paper shows how the SSMC algorithm combined with the FS algorithm[12] can be applied to Bayesian analysis of acoustically coupled-spaces. The SSMC algorithm has not yet been documented (at least to the best knowledge of the authors) in acoustic applications. As Bayesian inferential methods have increasingly found applications in acoustics research, the introduction of the SSMC algorithm in the context of architectural acoustics may also benefit acousticians who are working on Bayesian methods in other acoustics applications. Specifically, the significance of this work for architectural acousticians is that an increased accuracy, higher efficiency, and critically less user-interaction within the Bayesian framework can be achieved for sound energy decay analysis, particularly for multiple-slope decays, often encountered in acoustically coupled-spaces.[14–17] High efficiency is required in practice, since architectural acousticians often need to analyze multiple decay times and related parameters, along with their uncertainties over 6–8 octave bands or over 10–22 third-octave bands. Reducing the required user tuning and interaction is beneficial to acousticians who are unaccustomed with ISMC and MCMC algorithms but who still

[a)] Aspects of this work have been presented at the 152nd ASA Meeting in Honolulu 2006 and at the 155th ASA Meeting in Paris.
[b)] Author to whom correspondence should be addressed. Electronic mail: xiangn@rpi.edu

require the benefits of Bayesian analysis.

This paper is organized as follows, Sec. II briefly describes Bayesian formulation of a PPDF over the decay parameter space. Section III discusses the difficulties in choosing an appropriate ISMC sampling distribution and the related problem in choosing an appropriate proposal distribution in MCMC algorithms (focusing on the commonly used Metropolis–Hastings algorithm). Section IV discusses implementation aspects of the SSMC algorithm used in this work. Section V shows the results of the SSMC/FS algorithm applied to experimentally obtained Schroeder decay functions. Section VI concludes the paper.

II. BAYESIAN FORMULATION

A detailed explanation in the Bayesian framework is given in the papers by Xiang and Jasa;[12] this paper begins with a brief review on the Schroeder decay function model for determining the decay parameters in acoustically coupled-spaces. A linear parametric model **GA** approximates the experimental data **D** as follows:

$$\mathbf{D} = \mathbf{GA} + \mathbf{e}, \tag{1}$$

with an error **e** where **A** is a vector of m weighting coefficients and **G** is a $K \times m$ discretized model matrix, with the jth column of **G** given by

$$G_{kj}(T_j, t_k) = \begin{cases} t_K - t_k & \text{for } j = 0 \\ \exp(-13.8 t_k / T_j) & \text{for } j = 1,2,\ldots,m-1. \end{cases} \tag{2}$$

T_j in Eq. (2) is the jth decay time parameter to be determined for $0 \le j \le m-1$, $T_0 = \infty$, $0 \le k \le K-1$, and t_K represents the upper limit of Schroeder's integration and K is the number of data points of the Schroeder decay function. The validity of this model for determining the decay times in acoustically coupled-spaces has been experimentally verified (especially when t_K is large) in Ref. 11. This work applies a Bayesian analysis to the decay model in Eq. (1) as briefly summarized below. Prior to analysis, the error components e_i are only known to be of a finite amount of energy. With this being the only information I available, the application of the principle of maximum entropy[18] leads to assignment of a likelihood function

$$l(\mathbf{T}, \mathbf{A}, \sigma | \mathbf{D}, I) = (\sqrt{2\pi}\sigma)^{-K} \exp\left(\frac{-\mathbf{e}^{\mathrm{Tr}}\mathbf{e}}{2\sigma^2} \right), \tag{3}$$

where **T** is the vector of the decay times and **A** is a vector of linear coefficients and Tr denotes a matrix transpose. Both **A** and **T** are decay parameters that the authors wish to find. The parameter σ^2 in Eq. (3) represents a finite but unspecified error variance. In room-acoustics practice, acousticians are challenged to estimate decay parameters for multiple-sloped sound energy decays. For a double-sloped decay [$m=3$ in Eq. (2)] this results in a likelihood function over a six-dimensional parameter space, with σ being one additional parameter along with three linear (A_j) and two (nonlinear) decay time (T_j) parameters. At this point it is possible to marginalize over the error variance leaving the likelihood in terms of the decay times and the linear coefficients as shown in Appendix A. This results in a likelihood function in the form of a student-T distribution

$$l(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) = (2\pi)^{-K/2} \Gamma\left(\frac{K}{2} \right) \frac{Q^{-K/2}}{2} \tag{4}$$

with

$$Q = \frac{\mathbf{e}^{\mathrm{Tr}}\mathbf{e}}{2} \tag{5}$$

and gamma function $\Gamma(x)$.

The PPDF of $\mathbf{T}, \mathbf{A}$ given data **D** and the available background information I as noted by $p(\mathbf{T}, \mathbf{A} | \mathbf{D}, I)$ are defined by the likelihood and prior probability $\pi(\mathbf{T}, \mathbf{A} | I)$ of the decay parameters

$$p(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) = \frac{1}{Z} l(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) \pi(\mathbf{T}, \mathbf{A} | I), \tag{6}$$

where

$$Z = \int l(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) \pi(\mathbf{T}, \mathbf{A} | I) d(\mathbf{T}, \mathbf{A}). \tag{7}$$

As $p(\mathbf{T}, \mathbf{A} | \mathbf{D}, I)$ of Eq. (6) cannot be represented in closed form due to the nonlinear nature of the model given in Eq. (2), it is convenient to form a compact representation based on $\mathbf{T}, \mathbf{A}$ [one example being the mean and/or covariance of $\mathbf{T}, \mathbf{A}$ as was done in Ref. 11]. In order to simplify the notation in the remainder of the paper, a compact representation will be denoted by

$$\begin{aligned} L &= \int f(\mathbf{T}, \mathbf{A}) p(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) d(\mathbf{T}, \mathbf{A}) \\ &= \frac{1}{Z} \int f(\mathbf{T}, \mathbf{A}) l(\mathbf{T}, \mathbf{A} | \mathbf{D}, I) \pi(\mathbf{T}, \mathbf{A} | I) d(\mathbf{T}, \mathbf{A}), \end{aligned} \tag{8}$$

where $f(\mathbf{T}, \mathbf{A})$ of Eq. (8) is used to define a particular compact representation. Details of these compact representations (specifically the mean and covariance of $\mathbf{T}, \mathbf{A}$) using experimentally obtained data will be shown in Sec. V. In order to simplify the notation the authors combine the decay time and linear coefficients $\mathbf{T}, \mathbf{A}$ into a single parameter vector **X** when there is no need to distinguish between the two and the background information I is also dropped for the remainder of the paper for simplicity.

III. DIFFICULTIES WITH TWO COMMON MONTE CARLO ALGORITHMS

ISMC and MCMC algorithms both rely on choosing initial probability distributions representing the prior knowledge of the PPDF to be estimated. This section discusses potential difficulties in choosing these initial distributions. The initial distributions are denoted by either a *sampling distribution* for the ISMC algorithm or *proposal distribution* for the MCMC algorithm.

A. ISMC integration algorithms

The work by Xiang *et al.*[11] used ISMC integration in which the ISMC sampling distribution $g(\mathbf{X})$, with a support greater than $p(\mathbf{X}|\mathbf{D})$, is applied to the integral of Eq. (8) as follows:

$$L = \int f(\mathbf{X}) \frac{p(\mathbf{X}|\mathbf{D})}{g(\mathbf{X})} g(\mathbf{X}) d\mathbf{X} = \int f(\mathbf{X}) w(\mathbf{X}) g(\mathbf{X}) d\mathbf{X}, \quad (9)$$

with $w(\mathbf{X}) = p(\mathbf{X}|\mathbf{D})/g(\mathbf{X})$. The ISMC sampling distribution $g(\mathbf{X})$ can be effectively replaced with the approximation (see Appendix B)

$$\hat{g}(\mathbf{X}) = \frac{1}{M} \sum_{r=0}^{M-1} \delta(\mathbf{X} - \mathbf{X}_r), \quad (10)$$

where $\delta(\mathbf{X}-\mathbf{X}_r)$ is a Dirac delta function centered at the sample $\mathbf{X}_r$ drawn from the sampling distribution $g(\mathbf{X})$, and M is the number of such samples used. Using the approximation $\hat{g}(\mathbf{X})$ the representation of Eq. (9) is given by

$$L \approx \frac{1}{M} \sum_{r=0}^{M-1} \int f(\mathbf{X}) w(\mathbf{X}) \delta(\mathbf{X} - \mathbf{X}_r) d\mathbf{X}$$

$$= \frac{1}{M} \sum_{r=0}^{M-1} f(\mathbf{X}_r) w(\mathbf{X}_r). \quad (11)$$

The formalism of Eq. (11) using Eq. (10), termed ISMC integration, has been presented as opposed to defining "estimators" of L as is commonly found in statistical literature,[19] for the benefit of readers who are not well versed with statistical terminology.

The difficulty of choosing an appropriate ISMC sampling distribution $g(\mathbf{X})$ in high dimensions has been shown in Ref. 19. The difficulty can be demonstrated with an explicit numerical example (see Appendix C) to show that the difficulty still exists even in low dimensions. The following illustrative example shows how ISMC estimates can be very sensitive to poor choices of sampling distributions in terms of placement as well as variance. To highlight the difficulties, Fig. 1 illustrates a marginal PPDF $p(\mathbf{T}|\mathbf{D})$ evaluated for two different room impulse responses experimentally measured in real halls. Figure 1(a) shows a very sharply peaked PPDF while Fig. 1(b) shows a very narrow, elongated PPDF along one dimension. The ellipsoids marked in the figures conceptually indicate sampling distributions for ISMC. The ISMC sampling distribution marked by A, with a support greater than but similar to the actual PPDF $p(\mathbf{T}|\mathbf{D})$, is ideal for precise, unbiased estimations using ISMC integration. The ISMC sampling distribution B, with a support much greater than the actual PPDF $p(\mathbf{T}|\mathbf{D})$ still results in a reasonable ISMC integration estimate, but is less efficient as more samples will be required to obtain a good result. The ISMC sampling distribution C, with a support less than the actual PPDF $p(\mathbf{T}|\mathbf{D})$, will lead to failure of ISMC integration estimates as the variance of the estimates as given by Eq. (C2) will likely be unbounded. Figure 1 uses two actual PPDFs evaluated from experimentally measured results to demonstrate that a sampling distribution marked by A is in practice hardly possible without any prior knowledge on the sharpness (spreading), orientation, location, and size of actual PPDFs. In light of these difficulties, creating an efficient automated procedure for determining decay parameters and their reliability estimates from Schroeder decay curves using ISMC integration will be difficult, since the location of the PPDF mode, its shape, its orientation, and its size may not be known when the ISMC sampling distribution has to be selected.

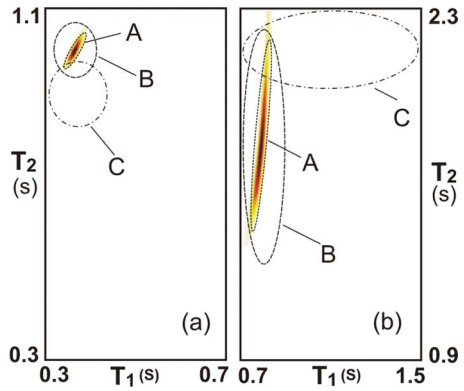


FIG. 1. (Color online) Marginalized PPDF $p(\mathbf{T}|\mathbf{D})$ evaluated for two different room impulse responses experimentally measured in real halls with ellipsoids indicating proposal distributions for ISMC integration. Proposal distributions marked by A, with support greater than the PPDF $p(\mathbf{T}|\mathbf{D})$, is ideal for precise, unbiased estimations using ISMC integration. Proposal distribution B, with a support greater than the PPDF $p(\mathbf{T}|\mathbf{D})$ still results in reasonable ISMC estimates, but is less efficient. Proposal distributions C, with a support less than the PPDF $p(\mathbf{T}|\mathbf{D})$, will lead to failure of ISMC estimates.

B. MCMC algorithms

An alternative to an ISMC integration approach is a MCMC algorithm such as the popular Metropolis–Hastings algorithm.[19] The Metropolis–Hastings algorithm generates *dependent* samples $\mathbf{X}_r$ from the PPDF $p(\mathbf{X}|\mathbf{D})$ using only knowledge of the likelihood and the prior $l(\mathbf{X}|\mathbf{D})\pi(\mathbf{X})$. As was done with the ISMC algorithm in Eq. (10), these samples can then be used to form an approximation of the PPDF $p(\mathbf{X}|\mathbf{D})$ by

$$p(\mathbf{X}|\mathbf{D}) \approx \frac{1}{M} \sum_{r=0}^{M-1} \delta(\mathbf{X} - \mathbf{X}_r), \quad (12)$$

which can estimate the representation of Eq. (8) by a Monte Carlo approximation

$$L \approx \frac{1}{M} \sum_{r=0}^{M-1} \int f(\mathbf{X}) \delta(\mathbf{X} - \mathbf{X}_r) d\mathbf{X} = \frac{1}{M} \sum_{r=0}^{M-1} f(\mathbf{X}_r). \quad (13)$$

The Metropolis–Hastings algorithm generates a sequence of samples $\mathbf{X}_r$ through the parameter space by a random walk. At each step of the algorithm a sample $\mathbf{S}_r$ in the parameter space is chosen with probability distribution given by $h(\mathbf{S}_r - \mathbf{X}_r)$, where $h(\mathbf{X})$ is a user defined proposal distribution. The sample $\mathbf{S}_r$ is accepted (i.e., $\mathbf{X}_{r+1} = \mathbf{S}_r$) with a probability

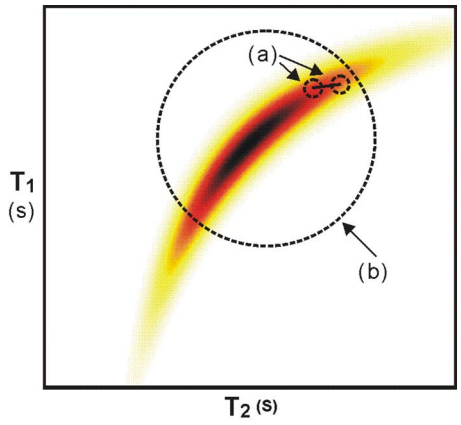


FIG. 2. (Color online) Example of potential problems in choosing a proposal distribution for the Metropolis–Hastings MCMC algorithm with symmetric normal proposal distribution. Circles show one standard deviation distance from the mean. (a) Small tailed proposal distribution $h(\mathbf{T})$ with too small support in comparison to the PPDF $p(\mathbf{T}|D)$ results in most of the proposed MCMC samples $\mathbf{S}_r$ being accepted but with a slow exploration of the PPDF $p(\mathbf{T}|D)$. (b) Heavy tailed proposal distribution with excessive support in comparison to the PPDF $p(\mathbf{T}|D)$ results in majority of the proposed MCMC samples $\mathbf{S}_r$ being rejected, which also results in a slow exploration of the PPDF $p(\mathbf{T}|D)$.

$$\min\left[1, \frac{p(\mathbf{S}_r|\mathbf{D})h(\mathbf{S}_r - \mathbf{X}_r)}{p(\mathbf{X}_r|\mathbf{D})h(\mathbf{X}_r - \mathbf{S}_r)}\right], \tag{14}$$

and rejected (i.e., $\mathbf{X}_{r+1}=\mathbf{X}_r$) otherwise. Both $p(\mathbf{S}_r|\mathbf{D})$ and $p(\mathbf{X}_r|\mathbf{D})$ of Eq. (14) are evaluated from the posterior distribution. The proposal distribution $h(\mathbf{X})$ is commonly chosen to be a symmetric function in which case $h(\mathbf{S}_r-\mathbf{X}_r)=h(\mathbf{X}_r-\mathbf{S}_r)$ and so Eq. (14) reduces to

$$\min\left[1, \frac{p(\mathbf{S}_r|\mathbf{D})}{p(\mathbf{X}_r|\mathbf{D})}\right] \tag{15}$$

[see Appendix D for details on the validity of the acceptance probability of Eq. (14)]. While the proposal distribution $h(\mathbf{X})$ is not explicitly present in Eq. (15), the choice of $h(\mathbf{X})$ will have a large impact on the efficiency of the Metropolis–Hastings algorithm. If $h(\mathbf{X})$ has much thicker tails than the PPDF $p(\mathbf{X}|\mathbf{D})$ then the acceptance probability of Eq. (15) will be very low for most of the proposed samples $\mathbf{S}_r$, which results in $\mathbf{X}_{r+1}=\mathbf{X}_r$ for most r, and so the algorithm will not explore the parameter space efficiently. Similarly if $h(\mathbf{X})$ has much thinner tails than the PPDF $p(\mathbf{X}|\mathbf{D})$ then $p(\mathbf{S}_r|\mathbf{D})/p(\mathbf{X}_r|\mathbf{D})\approx 1$ and so almost all of the proposed samples $\mathbf{S}_r$ will be accepted; however, the algorithm will still explore the parameter space very slowly as $p(\mathbf{X}_{r+1}|\mathbf{D})=p(\mathbf{S}_r|\mathbf{D})\approx p(\mathbf{X}_r|\mathbf{D})$. Figure 2 conceptually shows an example of two symmetric normal proposal distributions superimposed on two different PPDFs with dashed-line circles indicating a distance of one standard deviation from the mean. Figure 2(a) illustrates a thin tailed proposal distribution in comparison with the PPDF shown, which results in a slow exploration of the PPDF even though most of the proposed samples are accepted. Figure 2(b) shows a case where the proposal distribution has a thicker tail in one dimension than the PPDF, which will cause most of the proposed samples to be rejected, again resulting in a slow exploration of the PPDF.

IV. THE SLICE SAMPLING MONTE CARLO ALGORITHM

As shown in Sec. III, both the Metropolis–Hastings MCMC algorithm and the ISMC algorithm suffer from the requirement of a good initialization of proposal/sampling distributions. For efficiently determining the representation of Eq. (8) it is important to use an algorithm that is less dependent on good initialization than either the Metropolis–Hastings MCMC or ISMC integration algorithms. The SSMC algorithm as presented by Neal[13] was developed, in part, to minimize the effect of the proposal distributions on efficiency of the algorithm.

A. The SSMC algorithm

The fundamental principle of the SSMC algorithm is to introduce an auxiliary probability distribution, which will aid generating samples from the desired distribution. As an example consider a one-dimensional PPDF $p(X|\mathbf{D})$. One can define an auxiliary distribution[13] by

$$p(X,y|\mathbf{D}) = \begin{cases} 1 & \text{if } 0 < y < p(X|\mathbf{D}) \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Marginalization over the variable y results in

$$\int p(X,y|\mathbf{D})dy = \int_0^{p(X|\mathbf{D})} 1dy = p(X|\mathbf{D}), \tag{17}$$

which is the desired posterior distribution. As in any Monte Carlo approach the marginalization can be implemented by sampling from the joint distribution $p(X,y|\mathbf{D})$ and ignoring the parameter y. The multidimensional PPDF $p(\mathbf{X},y|\mathbf{D})$ of Eq. (6) can be handled in the same manner by simply applying the auxiliary distribution of Eq. (16) to each component X_j individually

$$p(X_j,y|\mathbf{D}) = \begin{cases} 1 & \text{if } 0 < y < p(X_j|\mathbf{D}) \\ 0 & \text{otherwise,} \end{cases} \tag{18}$$

as is done in Gibbs sampling approach.[3,7,19] A key element of the SSMC algorithm is that it effectively replaces the sampling distribution (ISMC) or proposal distribution (MCMC) algorithms with a uniform proposal distribution; its spreading is adaptively constrained by the PPDF to be sampled. The principle benefit of this approach is that it allows for an adaptive tuning within the SSMC algorithm (which is difficult to achieve with other MCMC algorithms). This paper presents a simplified SSMC algorithm as discussed in Ref. 13, in which the adaptive tuning of the auxiliary variable is achieved using an interval doubling technique. Other more elaborate versions of the SSMC algorithm, as well as proofs of validity of the algorithm, are described in the original Ref. 13. As with the Metropolis–Hastings algorithm, the SSMC algorithm has a update rule, which defines a new sample $\mathbf{X}_{r+1}$ in the parameter space given the current sample $\mathbf{X}_r$. The authors present the simplified SSMC algorithm and update rule below.

Algorithm 1. Simplified slice sampling: Return sample X_{r+1} given sample X_r drawn from the distribution $p(X|\mathbf{D})$.

1: $y=a$ random value from uniform distribution $[0,p(X_r|\mathbf{D})]$
2: $u=a$ random value from uniform distribution [0,1]
3: $x_l=X_r-(1-u)w$
4: $x_r=X_r+uw$
5: **while** $p(x_l|\mathbf{D})\geq y$ **do**
6: $x_l=x_l-w$
7: **end while**
8: **while** $p(x_r|\mathbf{D})\leq y$ **do**
9: $x_r=x_r+w$
10: **end while**
11: **while** 1 **do**
12: $x'=a$ random value from uniform distribution $[x_l,x_r]$
13: **if** $p(x'|\mathbf{D})\geq y$ **then**
14: **return** $X_{r+1}=x'$
15: **else**
16: **if** $p(x'|\mathbf{D})\leq p(x_l|\mathbf{D})$ **then**
17: $x_l=x'$
18: **end if**
19: **if** $p(x'|\mathbf{D})\geq p(x_r|\mathbf{D})$ **then**
20: $x_r=x'$
21: **end if**
22: **end if**
23: **end while**

B. A graphical illustration of the SSMC algorithm

Figure 3(a) shows a unimodal marginal PPDF $p(X|\mathbf{D})$ with an initial starting sample given by X_0 and the value of y randomly chosen from the uniform distribution defined over $[0,p(X_0|\mathbf{D})]$, which corresponds to step 1 of Algorithm 1. Figure 3(b) shows the "slice" S of the parameter space defined as the region where $p(X|\mathbf{D})>y$. Figure 3(c) shows the random stepping out procedure of steps 3–10, which have been done through an interval doubling approach.[13] The result is that the bounding interval B contains the slice S. Figure 3(d) shows a randomly selected point X' chosen within the bounding interval B. As the point X' is not inside the slice S the bounding interval is shrunk to where the point X' defines a new boundary point [in this case the new value for x_l for the interval B' as is shown in Fig. 3(e) corresponds to steps 11–23 in Algorithm 1]. Finally, Fig. 3(f) shows a randomly selected point X'' drawn from B' also contained in the slice S. This point X'' is accepted as a new sample X_1. The process is then iterated with X_1 as the initial sample of the algorithm.

An important feature of the SSMC algorithm is that it generates samples relying only on a uniform proposal distribution whose variance is determined adaptively. In other words, SSMC can explore the PPDF efficiently by updating knowledge from the PPDF to be sampled while the sampling is in progress.

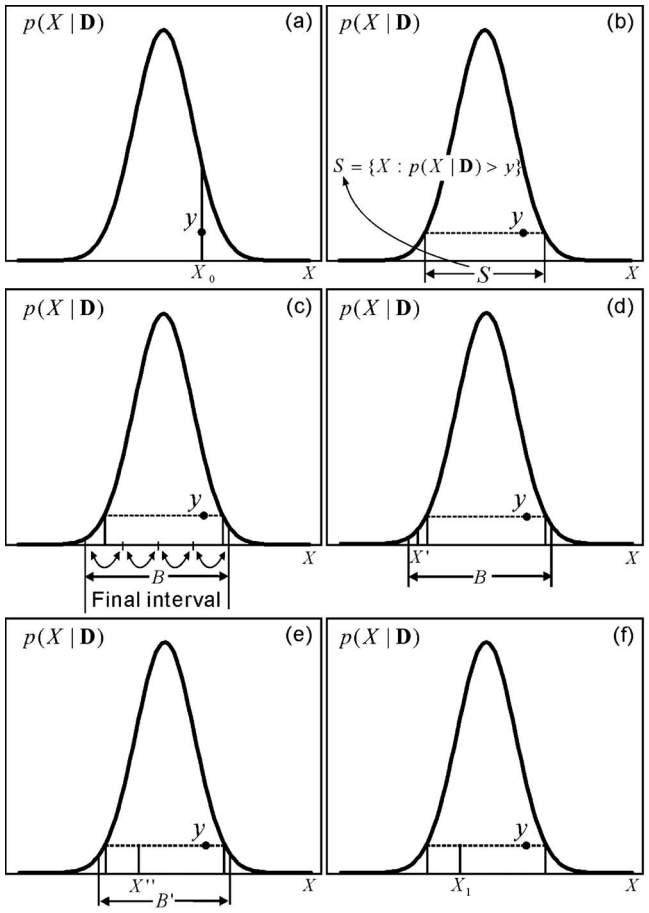


FIG. 3. Iterative steps of the slice sampling illustrated using an exemplary unimodal PPDF. One-dimensional parameter X is used for an illustrative discussion. For the experimental data discussed in Sec. V $X=A_j$ (with $j=0,1,2$) or $X=T_k$ (with $k=1,2$), respectively. (a) Unimodal PPDF $p(X|\mathbf{D})$ with an initial starting sample given by X_0 and the randomly chosen value of y, which corresponds to step 1 of Algorithm 1. (b) Slice S of the parameter space defined as the region where $p(X|\mathbf{D})>y$, namely, $S=\{X:p(X|\mathbf{D})>y\}$. (c) Random stepping out procedure of steps 3–10, which have been done through an interval doubling approach. (d) Randomly selected point X' chosen within the bounding interval B. As the point X' is not inside the slice S the bounding interval is shrunk to where the point X' defines a new boundary point [in this case the new value for x_l for the interval B' as is shown in (d) and corresponds to steps 11–23]. (e) Randomly selected point X'' drawn from B' also contained in the slice S. This point X'' is accepted as a new sample X_1. The process is then iterated with X_1 as the initial sample of the algorithm.

C. Initialization and convergence of the SSMC algorithm

The SSMC algorithm still requires the user to specify the interval doubling parameters w_{T_j} and w_{A_j}, where w_{T_j} and w_{A_j} correspond to the estimate of the spread of the PPDF $p(\mathbf{X}|\mathbf{D})$ in each of the T_j and A_j parameters, respectively. The choice of w_{T_j} can, in principle, be chosen fairly well based on the expected precision in architectural acoustics practice. In the practical implementation of SSMC, a rough estimation of the reverberation time, in case of a single-slope energy decay, can be easily deduced using a small early-decay portion of the decay function, while its standard deviation τ is expected about 1% of the reverberation time T to be determined, which leads to a proper choice of w_{T_j}. In case of a double-slope energy decay, the primary decay time T_1

can be easily estimated in the same way as for the reverberation time in the single-slope case, and the secondary decay time T_2 is expected to be larger than T_1 (see Ref. 2); however, the standard deviation τ_2 of T_2 is expected to be in the same order or even larger than τ_1 of T_1, which means a similar order of w_{T_1} and w_{T_2} can be straightforwardly chosen in the practical implementation of the SSMC algorithm in order to reach the expected precision in architectural acoustics practice. An important advantage of the SSMC algorithm over ISMC integration or conventional MCMC algorithms (such as Metropolis–Hastings) is that the w_{T_j} and w_{A_j} are adjusted dynamically by the algorithm, and so the SSMC algorithm is less sensitive to a poor choice of these values. In fact, w_{T_j} and w_{A_j} can be updated dynamically from sample to sample based on information gained during shrinking steps of the previous samples. This fact is especially beneficial as choosing values for the w_{A_j} parameters is more difficult than for w_{T_j} parameters and good rules of thumb are as of yet unknown. Section V will elaborate on the initialization of the interval doubling parameter $\mathbf{w}$ using experimentally measured data. As with all other MCMC algorithms the efficiency of the SSMC algorithm will depend on the initial starting sample $\mathbf{X}_0$. If the sample $\mathbf{X}_0$ is in a region of low PPDF values, the algorithm will take longer to converge to the PPDF and so many of the initial samples do not represent the PPDF well. This problem is often alleviated using a "burn-in" phase, in which sample $\mathbf{X}_0$ is chosen after a certain amount of initial samples $\mathbf{X}_{-m},\ldots,\mathbf{X}_{-1}$ are discarded. The burn-in phase of the algorithm can be avoided using the FS algorithm[12] to choose the initial sample $\mathbf{X}_0$ for both the linear $\mathbf{A}$ and decay time $\mathbf{T}$ parameters. The ability to choose the initial parameters $\mathbf{T}$ and $\mathbf{A}$ using the FS algorithm and the ability of the SSMC algorithm to overcome poor choices for w_{T_j} and w_{A_j} allows for a combined algorithm, which is especially useful in architectural acoustics practice. As with other MCMC algorithms, the dependent samples $\mathbf{X}_r$ generated by the SSMC algorithm can be used to approximate the representation of Eq. (8); however, proving when the calculated representation of Eq. (8) has converged is an open research problem (a problem shared by all MCMC and MC algorithms). Section V B discusses one particular heuristic method to detect convergence.

V. EXPERIMENTS

An intimate performance hall (Susan Howorth Theater) in Powerhouse Arts Center, Oxford, MS is coupled to a reverberant gallery. The gallery and the theater measures are $21.3\times12.2\times7.4$ m^3 and $21.3\times16.2\times7.63$ m^3, respectively (see a sketch in Fig. 4). With doors closed, the natural reverberation times averagely amount to 1.5 s for the primary space (theater) and 3.9 s for the secondary space (gallery), respectively. In the acoustically coupled-spaces, when the primary space possesses shorter nature reverberation times than the secondary space, energy decays often exhibit double-sloped decay behaviors.[2] To investigate the energy decay characteristics, an omni-directional sound source is placed at the middle of the stage, whereas an omni-directional microphone as a sound receiver is located at

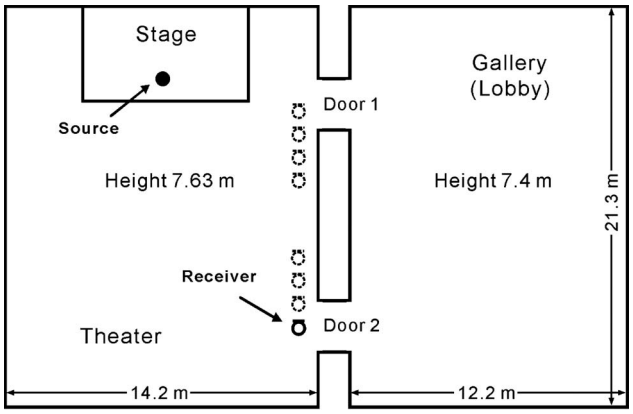


FIG. 4. Plane view of the Susan Howorth Theater in Powerhouse Arts Center, Oxford, MS. Two doors couple the theater and a reverberant gallery. Sound source is positioned on the stage, room impulse responses at receiver positions near two doors are measured, and the current paper focuses on one data set at the position marked by a solid-line symbol.

many strategic positions as indicated in sketch (Fig. 4). Room impulse responses are measured. In order to analyze decay characteristics over architectural acoustics-relevant frequency ranges, each room impulse response is first (octave) band-pass filtered. Schroeder integration is then applied to the room impulse responses for each octave band. A five-parameter model representing a double-slope decay associated with two decay times is used for energy decay analysis based on Schroeder integration results

$$F(t_k,\mathbf{T},\mathbf{A}) = A_0(t_K - t_k) + A_1\exp(-13.8t_k/T_1) + A_2\exp(-13.8t_k/T_2). \quad (19)$$

This model has exemplary illustrative purpose, as it is of both practical importance to architectural acousticians and sufficiently complex to demonstrate the benefits of the combined SSMC/FS algorithm in creating a method for estimating decay times with a minimum of user interaction. In the following the authors discuss the measurement results from a specific location being close to an opening door (door 2) to the gallery. The authors use exemplary data, which are a room impulse response band-pass filtered at 250 Hz octave band [see Fig. 5(a)] for the following discussion. Figure 5(b) illustrates the resulting Schroeder decay curve. The likelihood and posterior are determined as described in Sec. II, which results in a PPDF over a six-dimensional space when including variance σ^2 as a unknown parameter using Eq. (3), or with σ being removed by marginalization using Eq. (4) the PPDF is defined over a five-dimensional parameter space given by the decay times $\mathbf{T}$ and linear coefficients $\mathbf{A}$.

A. Initializing the SSMC/FS algorithm with experimental data

Applying the FS algorithm to the experimental data resulted in an initial estimate of both the decay time and linear parameters. The initial estimates $T_1=1.5$ s and $T_2=3.3$ s and $A_0=-4.0e-8$, $A_1=0.2486$, and $A_2=0.0613$, for a decay data segment starting −5 dB until the end of the decay trace [see Fig. 5(b)], are then used as the initial starting point of the SSMC algorithm. Decay time interval doubling values of

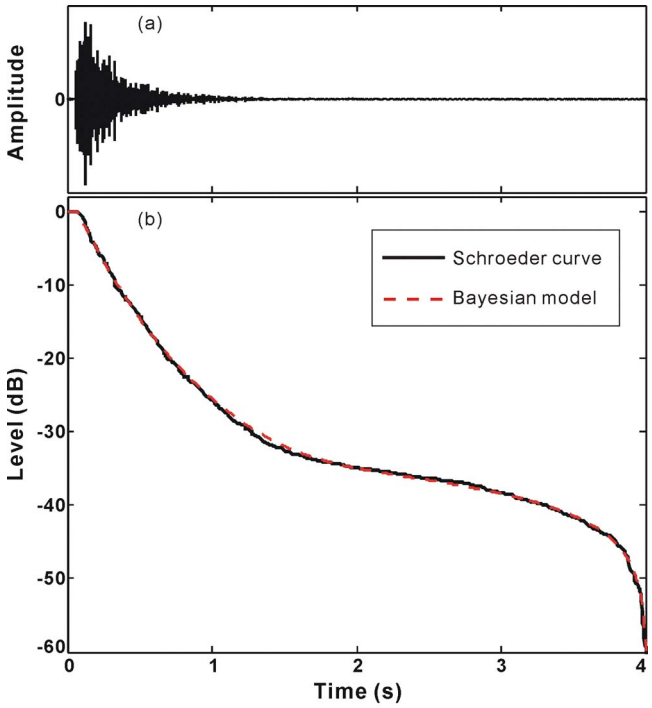


FIG. 5. (Color online) Room impulse response and corresponding decay curves experimentally measured in Howorth Theater. (a) Room impulse response after octave band-pass filtering. (b) Schroeder decay curve evaluated from the room impulse response compared with the Bayesian decay model curve determined by the SSMC/FS algorithm.

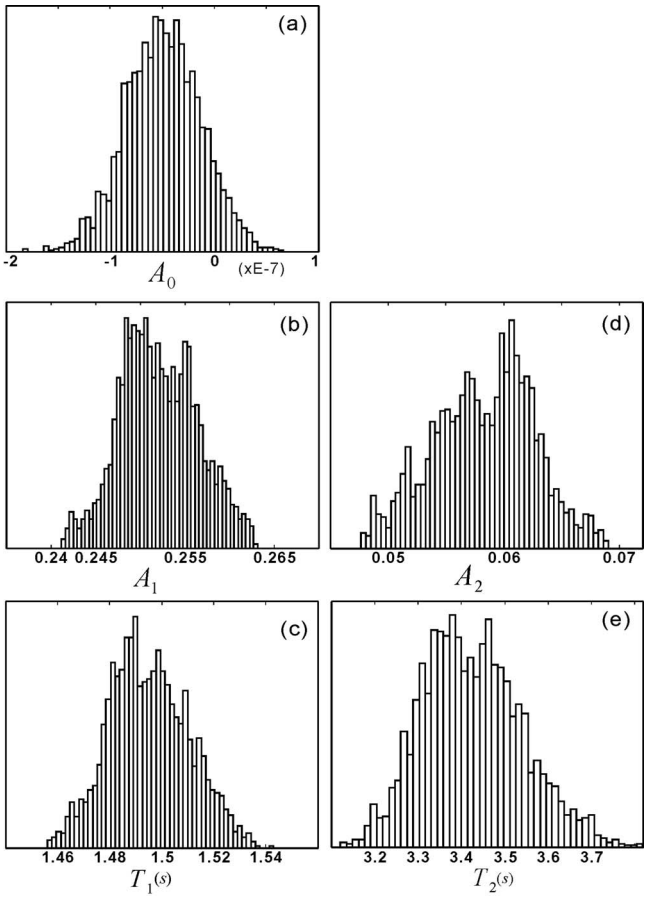


FIG. 6. Marginal histograms (MHs) of decay parameters from samples generated by SSMC/FS algorithm from experimental data measured in Susan Howorth Theater. (a) MH for A_0. (b) MH for A_1. (c) MH for T_1. (d) MH for A_2. (e) MH for T_2.

$w_{T_1}=0.03$ and $w_{T_2}=0.1$ are chosen based on the discussion given in Sec. IV C. Figures 6(c) and 6(e) show the marginal distributions of the decay time parameters T_1 and T_2 created from samples generated by the SSMC algorithm. The value of $w_{T_1}\approx 0.03$ is approximately three times that of the spread for T_1, while the value of $w_{T_2}\approx 0.1$ is a good guess to the spread of T_2. The linear coefficient doubling values are assigned values of $w_{A_0}=1$, $w_{A_1}=1$, and $w_{A_2}=1$. Figures 6(a), 6(b), and 6(d) show the marginal distributions of the linear parameters A_0, A_1, and A_2 created from samples generated by the SSMC algorithm. The values of w_{A_1} and w_{A_2} are approximately 30 times that of the spread for $A_1, A_2\approx 0.03$, while the value of w_{A_0} is approximately 3×10^6 times that of the spread for $A_0\approx 3\times 10^{-7}$. This poor choice of $w_{A_0}=1$, $w_{A_1}=1$, and $w_{A_2}=1$ represents a typical case in which an acoustician would have difficulty in assigning these parameters with good initial values. The SSMC algorithm, however, can compensate for this poor choice as will be shown in Sec. V B.

B. Convergence and decay parameter estimation for the experimental data

The SSMC algorithm is asymptotically guaranteed to converge and produce a sequence of M dependent samples $\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{M-1}$ from the PPDF $p(\mathbf{X}|\mathbf{D})$. There is, however, no indication as to how many samples are required to properly represent the PPDF in order to calculate the representation of Eq. (13). Convergence can be heuristically determined by finding the number of samples M such that all desired moments of Eq. (13) have converged within a predefined tolerance. For the experimental data the means and covariances of the decay times and linear parameters were used to assess convergence of the SSMC algorithm. Specifically the SSMC algorithm was deemed to converge when the quantities

$$\langle \hat{X}_j \rangle = \frac{1}{M}\sum_{r=0}^{M-1} X_{j,r}, \tag{20}$$

$$\langle \hat{X}_{jk} \rangle = \frac{1}{M}\sum_{r=0}^{M-2} (X_{j,r} - \langle \hat{X}_j \rangle)(X_{k,r} - \langle \hat{X}_k \rangle) \tag{21}$$

change less than 0.1%, where X is any decay time T, or linear parameter A, and $X_{j,r}$ denotes the jth component of the rth sample. For the experimental data in Fig. 5, the SSMC/FS algorithm converged within $\approx$12 000 samples. It is useful in this context to provide the acousticians with the histogram outputs given in Fig. 6. As the likelihood function $l(\mathbf{X}|\mathbf{D})$ given by Eq. (4) is a student-T type distribution and the PPDF is given by $p(\mathbf{X}|\mathbf{D}) \propto l(\mathbf{X}|\mathbf{D})\pi(\mathbf{X})$, histograms which follow the shape of a student-T distribution provide added evidence that the SSMC/FS algorithm has converged.

The moment estimates of Eqs. (20) and (21) were then used as the decay parameter estimates once convergence was determined. Specifically for decay times, the mean

TABLE I. Decay parameters estimated from one measurement in the Susan Howorth Theater. Decay times (T_1 and T_2) along with their standard deviations (Std_1 and Std_2) derived from covariance matrix of slice sampling. Level difference defined by $\Delta L = 10\log(A_1/A_2)$, A_0 is used to estimate the peak-to-noise ratio (PNR) (Ref. 1). Cross-correlation coefficients (CCCs) are listed in the last column.

Band (Hz)	T_1 (s)	Std_1 (s)	T_2 (s)	Std_2 (s)	ΔL (dB)	PNR (dB)	CCC
125	1.83	5.55×10^{-2}	3.96	5.04×10^{-1}	17.3	64.3	0.79
250	1.47	1.65×10^{-2}	3.27	1.93×10^{-1}	5.73	54.9	0.83
500	1.48	5.88×10^{-2}	4.46	4.86×10^{-1}	12.75	49.1	0.77
1000	1.49	5.24×10^{-2}	5.15	7.67×10^{-1}	15.84	50.1	0.85
2000	1.32	2.36×10^{-2}	2.97	5.96×10^{-2}	13.46	52.0	0.78
4000	0.94	2.47×10^{-2}	2.61	3.93×10^{-1}	16.54	52.2	0.81

$$\langle \hat{T}_j \rangle = \frac{1}{M}\sum_{r=0}^{M-1} T_{j,r} \tag{22}$$

and covariance

$$\langle \hat{C}_{jk} \rangle = \frac{1}{M}\sum_{r=0}^{M-1} (T_{j,r} - \langle \hat{T}_j \rangle)(T_{k,r} - \langle \hat{T}_k \rangle) \tag{23}$$

(where $T_{j,r}$ denotes the jth decay time component of the rth sample) given the M samples used in assessing convergence were used as estimates of the decay times $\mathbf{T}$. From the expected covariance matrix $\langle \hat{\mathbf{C}} \rangle = [\langle \hat{C}_{jk} \rangle]$, the individual variances τ_j^2 and the standard deviation τ_j of each decay time T_j were estimated as discussed in Ref. 11. The expected standard deviation τ_j serves a reliability estimate, since it is a measure of "error bar" of the estimated decay time $\langle \hat{T}_j \rangle$, while the inter-relationship between the decay times is measured by cross-correlation coefficient (CCC) $\hat{C}_{jk}/\sqrt{\hat{C}_{jj}\hat{C}_{kk}}$.[11] The error bars and CCCs for the experimental example are listed in Table I.

For the linear parameters, the mean

$$\langle \hat{A}_j \rangle = \frac{1}{M}\sum_{r=0}^{M-1} A_{j,r} \tag{24}$$

and standard deviation

$$\mathrm{Std}(\hat{A}_j) = \sqrt{\frac{1}{M}\sum_{r=0}^{M-1} (A_{j,r} - \langle \hat{A}_j \rangle)^2} \tag{25}$$

(where $A_{j,r}$ denotes the jth linear parameter component of the rth sample) given the M samples used in assessing convergence were used as estimates of the linear parameters $\mathbf{A}$. The means and standard deviations for the linear parameters for the experimental data taken at 250 Hz are shown in Table II. Figure 7 shows marginal posterior probability distributions (MPPDs) over two-dimensional (2D), zoomed-in parameter spaces from the experimental data. The MPPDs are generated by exhaustive sampling of the PPDF $p(\mathbf{X}|\mathbf{D})$ for all 2D MPPDs over $\{A_0,A_1\}$, $\{A_0,T_1\}$, $\{A_0,T_2\}$, $\{A_0,A_2\}$, $\{A_1,T_1\}$, $\{A_1,T_2\}$, $\{A_1,A_2\}$, $\{T_1,A_2\}$, $\{T_1,T_2\}$, and $\{A_2,T_2\}$, respectively. Parameters other than the pair shown were fixed to the mean values determined by Eq. (20) and given in Tables I and II. Estimated parameters listed in Tables I and II, when comparing with Fig. 7, indicate that the combined SSMC/FS algorithm successfully estimated the decay parameters and that the FS algorithm chose a good initial starting point for the SSMC algorithm. Figure 7 also shows that exhaustive sampling of the parameter space is not feasible without very good prior knowledge of the spread of the PPDF in all the dimensions. For example, exhaustive sampling over a five-dimensional space ranging between $-0.5e{-}7 \le A_0 \le 0.5e{-}7$, $0.1 \le A_1 \le 1.0$, $0.1 \le T_1 \le 5.0$, $0.001 \le A_2 \le 0.1$, and $1 \le T_2 \le 10$ (reasonable estimates of the parameter ranges for this acoustics problem), with each range partitioned into an appropriate number of cells to sample the marginal parameter distribution of Fig. 6, would require approximately 4×10^{12} samples compared to the 12 000 required for the SSMC/FS algorithm. Thus the SSMC/FS algorithm provides an efficient solution to the decay parameter estimation problem. Figure 7 also shows that choosing appropriate MCMC proposal or ISMC sampling distributions for the linear and decay time parameters can be difficult as there are large variations in sharpness (spreading) of the MPPDs, and in their orientations. While good MCMC proposal distributions could potentially be found by using initial runs or adaptive ISMC algorithms could be developed, in real experimental data, these variations also change from frequency band to frequency band, and from data to data, as such this task using ISMC or MCMC would be most likely require a significantly increased user tuning compared to the SSMC/FS algorithm.

TABLE II. Means (μ) and standard deviations (Std) of the linear parameters A_0, A_1, and A_2 estimated from the acoustical measurement in the Howorth Theater using the SSMC/FS algorithm, for 250 Hz octave-band evaluation.

Parameter	μ	Std
A_0	-2.91×10^{-8}	3.44×10^{-8}
A_1	0.2417	0.0044
A_2	0.0688	0.0044

It is unlikely that any type of initial processing used to define a MCMC proposal distribution and/or ISMC sampling distribution would not have an analog method to better choose the initial proposal distribution required by the SSMC algorithm as well, although this is a topic beyond the scope of the current work. Assessing SSMC convergence with the above heuristic scheme can be problematic for a PPDF with multiple modes or large areas with similar probability magnitude, where convergence of the representation will not give an indication that the SSMC algorithm has not

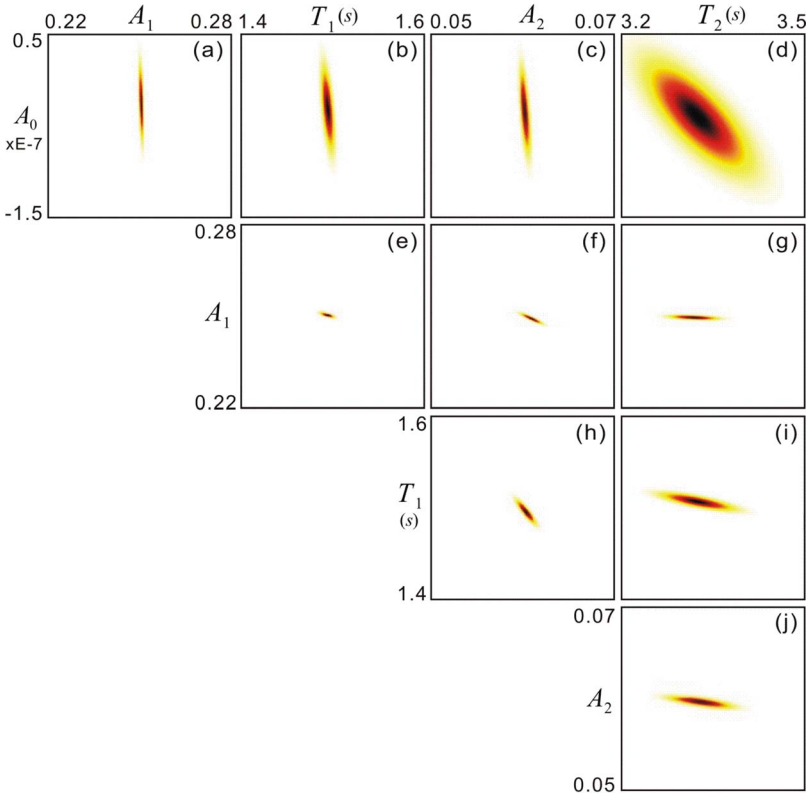


FIG. 7. (Color online) Marginal posterior probability distributions (MPPDs) over 2D (zoomed-in) parameter spaces from experimental data. Parameters other than the pair shown are fixed to the mean values (see Tables I and II). (a) MPPD over $\{A_0,A_1\}$. (b) MPPD over $\{A_0,T_1\}$. (c) MPPD over $\{A_0,A_2\}$. (d) MPPD over $\{A_0,T_2\}$. (e) MPPD over $\{A_1,T_1\}$. (f) MPPD over $\{A_1,T_2\}$. (f) MPPD over $\{A_1,T_2\}$. (g) MPPD over $\{T_1,A_2\}$. (h) MPPD over $\{T_1,A_2\}$. (i) MPPD over $\{T_1,T_2\}$. (j) $\{A_2,T_2\}$.

sufficiently explored the PPDF $p(\mathbf{X}|\mathbf{D})$. As discussed in Sec. IV, however, it is possible to focus on one mode of the PPDF; combining this fact with the use of the FS algorithm to initialize the SSMC algorithm in a region were the PPDF $p(\mathbf{X}|\mathbf{D})$ is significant allows for this heuristic scheme to be useful in practice. As assessing the convergence of MCMC and ISMC algorithms is also an open problem, the heuristic scheme discussed here is not considered as a drawback of the SSMC/FS algorithm in comparison to those algorithms.

VI. SUMMARY

This paper has shown that the SSMC algorithm is a suitable method for helping to automate the process of determining the decay parameter estimates in acoustically coupled-spaces with a minimum of user interaction and tuning. This paper has discussed potential problems with defining ISMC sampling distributions and MCMC proposal distributions when there was limited prior knowledge of the sharpness/position and orientation of the PPDF. In order to overcome this difficulty the SSMC algorithm was introduced in Sec. IV. The SSMC algorithm can overcome poor initialization of the proposal distribution through an adaptive process. In addition, Sec. IV also discussed how the FS algorithm could be combined with the SSMC algorithm to further improve SSMC performance by ignoring the burn-in phase and also improving the assessment of convergence for the SSMC algorithm. The SSMC/FS algorithm is applied to experimental data measured in Susan Howorth Theater in Sec. V. The plots of the MPPDs shown in Fig. 7 and the results in Tables I and II have demonstrated that the SSMC/FS algorithm is successful in estimating the decay parameters in an efficient manner as the number of samples required is on eight orders of fewer samples than possible through a deterministic exhaustive search algorithm, even with a relatively poor choice of initialization parameters. Figure 7 also illustrates the difficulty of defining ISMC sampling and MCMC proposal distributions for experimental data with a minimum of user tuning, which is especially important to acousticians who are unfamiliar with MCMC and ISMC algorithms, since large variations of posterior probability distributions in sharpness, orientation, position, and size can be encountered in the practice from data to data. A heuristic approach to assessing convergence of the SSMC/FS algorithm is also discussed. Choosing better heuristics specifically geared toward specific acoustics applications is an open problem not discussed in this paper.

In conclusion, the SSMC/FS algorithm is efficient in problems where good initialization of ISMC or MCMC algorithms is difficult, although it is possible that estimation of decay parameters in acoustically coupled-spaces (and other acoustics problems) could also be accomplished with a similar efficiency with better prior information about the nature of the PPDF and/or better expertise with MCMC and ISMC algorithms.

ACKNOWLEDGMENTS

The authors are very grateful to the reviewers for their insightful comments on the early version of the manuscript, which lead to a significant improvement of the paper. The authors would also like to thank David Woolworth for the data collection in the performance spaces. This paper is dedicated to Professor Jens Blauert on the occasion of his 70th birthday.

APPENDIX A

Marginalizing the likelihood given by Eq. (3)

$$l(\mathbf{T},\mathbf{A},\sigma|\mathbf{D},I) = (\sqrt{2\pi}\sigma)^{-K}\exp\left(\frac{-\mathbf{e}^{\mathrm{Tr}}\mathbf{e}}{2\sigma^2}\right) \tag{A1}$$

over the standard deviation σ can be accomplished by integrating the $l(\mathbf{T},\mathbf{A},\sigma|\mathbf{D},I)$ over σ using Jeffress' prior[18] $1/\sigma$, which results in

$$l(\mathbf{T},\mathbf{A}|\mathbf{D},I) = \int_0^\infty (\sqrt{2\pi}\sigma)^{-K}\exp\left(-\frac{Q}{\sigma^2}\right)\frac{1}{\sigma}d\sigma, \tag{A2}$$

where $Q=\mathbf{e}^{\mathrm{Tr}}\mathbf{e}/2$, then the identity[1]

$$\int_0^\infty x^{2n}\exp(-Qx^2)dx = \Gamma(n+1/2)\frac{Q^{-(n+1/2)}}{2} \tag{A3}$$

implies that

$$l(\mathbf{T},\mathbf{A}|\mathbf{D},I) = (2\pi)^{-K/2}\Gamma\left(\frac{K}{2}\right)\frac{Q^{-K/2}}{2}, \tag{A4}$$

which is Eq. (4).

APPENDIX B

Consider the PDF $g(\mathbf{X})$ and the cumulative density function (CDF) $G(\mathbf{X})$ related by

$$G(\mathbf{X}) = \int_{-\infty}^{\mathbf{X}} g(\mathbf{S})d\mathbf{S}, \tag{B1a}$$

$$g(\mathbf{X}) = \frac{d}{d\mathbf{X}}G(\mathbf{X}). \tag{B1b}$$

The representation of Eq. (9) can be determined using either the PDF or the CDF

$$L = \int f(\mathbf{X})w(\mathbf{X})g(\mathbf{X})d\mathbf{X} = \int f(\mathbf{X})w(\mathbf{X})dG(\mathbf{X}), \tag{B2}$$

where the CDF representation allows for both continuous and discrete distributions. The stepwise approximation

$$G(\mathbf{X}) \approx \frac{1}{M}\sum_{r=1}^{M} u(\mathbf{X}-\mathbf{X}_r), \tag{B3}$$

where $u(\mathbf{X})$ is a unit step function

$$u(\mathbf{X}) = \begin{cases} 1 & \text{if } \mathbf{X} \geq 0 \\ 0 & \text{otherwise}, \end{cases} \tag{B4}$$

is equivalent to creating a discrete or sampled approximation of the continuous CDF $G(\mathbf{X})$. Thus the representation of Eq. (9) is approximated by the sampled or discrete form of $G(\mathbf{X})$ as follows:

$$L \approx \frac{1}{M}\sum_{r=1}^{M}\int f(\mathbf{X})w(\mathbf{X})du(\mathbf{X}-\mathbf{X}_r). \tag{B5}$$

Using Eqs. (B1b) and (B3) and the property $\delta(\mathbf{X}-\mathbf{X}_r) = (d/d\mathbf{X})u(\mathbf{X}-\mathbf{X}_r)$ the approximated representation in Eq. (B5) can again be written with respect to the PDF $g(\mathbf{X})$. This results in Eq. (10) as follows:

$$g(\mathbf{X}) \approx \frac{1}{M}\sum_{r=1}^{M}\delta(\mathbf{X}-\mathbf{X}_r). \tag{B6}$$

TABLE III. Required number of samples M of the importance sampling for $3\sigma_w/\sqrt{M}<0.01$ given the absolute mean $|\mu|$ of the proposal distribution.

$\lvert\mu\rvert$	M
1	$1.55\times10^{+5}$
2	$4.83\times10^{+6}$
5	$6.49\times10^{+15}$

APPENDIX C

The following example illustrates the difficulties of choosing an importance sampling distribution (ISD) $g(X)$ in low dimensions. When generating independent samples from $g(X)$ is possible, the accuracy of the approximating representation is given by the central limit theorem[19]

$$|L_M - L| \to N\left(0,\frac{\sigma_w}{\sqrt{M}}\right) \quad \text{as } M\to\infty, \tag{C1}$$

where $N(\mu,\sigma)$ is a normal distribution with mean μ and standard deviation σ, with

$$\sigma_w^2 = \int \frac{(f(X)p(X))^2}{w(X)}dX - (L)^2. \tag{C2}$$

Consider the estimation of the normalizing constant

$$Z = \int \frac{p(X)}{g(X)}g(X)dX \tag{C3}$$

of a one-dimensional normal distribution with an unknown mean μ and identity standard deviation. The ISD is a zero-mean, normal distribution with an identity standard deviation. Using Eq. (C2), one can find the variance of the estimate as a function of μ given by

$$\sigma_w^2(\mu) = \exp(\mu^2) - 1. \tag{C4}$$

Table III shows the number of samples M required to achieve a standard statistical accuracy of $3\sigma_w/\sqrt{M}<0.01$ (a statistical 99.7% confidence interval that the result is correct) for different values of $|\mu|$. It is clear that the computational load required for the chosen ISD $g(X)$ becomes infeasible as $|\mu|$ increases.

APPENDIX D

This appendix presents a heuristic explanation of the acceptance probability given in Eq. (15) shown in Sec. III. For simplicity, the authors consider symmetric MCMC proposal distributions and a discrete parameter space. A MCMC transition kernel $K(\mathbf{X},\mathbf{U})=h(\mathbf{X}-\mathbf{U})$ defined by a symmetric proposal distribution $h(\mathbf{X}-\mathbf{U})=h(\mathbf{U}-\mathbf{X})$ [as is shown in Fig. 2, for example] represents the probability of moving from one point $\mathbf{X}$ in the parameter space to another $\mathbf{U}$. One must design the MCMC kernel $K(\mathbf{X},\mathbf{U})$ such that $p(\mathbf{X})$ (the distri-

bution of interest) is the unique invariant distribution of the kernel (see Ref. 19 for complete discussion of Markov chains and MCMC). The detailed balance property

$$h(\mathbf{X}-\mathbf{U})p(\mathbf{X}) = h(\mathbf{U}-\mathbf{X})p(\mathbf{U}) \tag{D1}$$

allows for $p(\mathbf{X})$ to be the desired invariance distribution. Thus detailed balance with a symmetric kernel implies that

$$p(\mathbf{X}) = p(\mathbf{U}) \tag{D2}$$

and so $p(\mathbf{X})$ is the uniform distribution. In order to generate samples from the desired invariant distribution $p(\mathbf{X})$, the MCMC transition kernel must be modified by the addition of an acceptance probability $A(\mathbf{X},\mathbf{U})$, which determines when a move from a point $\mathbf{X}$ to a proposed point $\mathbf{U}$ generated from the kernel $K(\mathbf{X},\mathbf{U})$ is accepted. Adding the acceptance probability, the detailed balance property becomes

$$A(\mathbf{X},\mathbf{U})h(\mathbf{X}-\mathbf{U})p(\mathbf{X}) = A(\mathbf{U},\mathbf{X})h(\mathbf{U}-\mathbf{X})p(\mathbf{U}) \tag{D3}$$

and so

$$A(\mathbf{X},\mathbf{U})p(\mathbf{U}) = A(\mathbf{U},\mathbf{X})p(\mathbf{X}). \tag{D4}$$

Choosing the acceptance probability $A(\mathbf{X},\mathbf{U}),A(\mathbf{U},\mathbf{X})$ so that $0 \le A(\mathbf{X},\mathbf{U}),A(\mathbf{U},\mathbf{X}) \le 1$ and satisfying the constraint given by Eq. (D4) result in $A(\mathbf{X},\mathbf{U})[p(\mathbf{X})/p(\mathbf{U})] \le 1$ and so $A(\mathbf{X},\mathbf{U}) \le [p(\mathbf{U})/p(\mathbf{X})]$. Thus setting $A(\mathbf{X},\mathbf{U}) = 1$ when $p(\mathbf{U}) \ge p(\mathbf{X})$ results in the proposed move $\mathbf{U}$ with higher probability always being accepted. Otherwise the proposed move $\mathbf{U}$ is accepted with probability $A(\mathbf{X},\mathbf{U}) = p(\mathbf{U})/p(\mathbf{X})$. This results in the acceptance probability for the MCMC algorithm given by

$$A(\mathbf{X},\mathbf{U}) = \min\left(1, \frac{p(\mathbf{U})}{p(\mathbf{X})}\right). \tag{D5}$$

It is important to note that the details of the symmetric proposal distribution $h(\mathbf{X}-\mathbf{U})$ do not effect the ability of the MCMC algorithm to correctly sample the desired distribution $p(\mathbf{X})$; however, the efficiency of the sampling is greatly dependent on the choice of the proposal distribution used (discussed in Sec. III).

[1]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian parameter estimation," J. Acoust. Soc. Am. **110**, 1415–1424 (2001).

[2]N. Xiang and P. M. Goggans, "Evaluation of decay times in coupled spaces: Bayesian decay model selection," J. Acoust. Soc. Am. **113**, 2685–2697 (2003).

[3]S. Dosso, "Quantifying uncertainty in geoacoustic inversion. I. A fast Gibbs sampler approach," J. Acoust. Soc. Am. **111**, 129–142 (2002).

[4]J. Dettmer, S. E. Dosso, and Ch. W. Holland, "Model selection and Bayesian inference for high-resolution seabed reflection inversion," J. Acoust. Soc. Am. **125**, 706–716 (2009).

[5]S. E. Dosso and M. J. Wilmut, "Comparison of focalization and marginalization for Bayesian tracking in an uncertain ocean environment," J. Acoust. Soc. Am. **125**, 717–722 (2009).

[6]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Statistical estimation of source location in presence of geoacoustic inversion uncertainty," J. Acoust. Soc. Am. **125**, EL171–176 (2009).

[7]Z. H. Michalopoulou and M. Picarelli, "Gibbs sampling for time-delay and amplitude estimation in underwater acoustics," J. Acoust. Soc. Am. **117**, 799–808 (2005).

[8]Z. H. Michalopoulou, "Multiple source localization using a maximum a posteriori Gibbs sampling approach," J. Acoust. Soc. Am. **120**, 2627–2634 (2006).

[9]Ch. Laplanche, "A Bayesian method to estimate the depth and the range of phonating sperm whales using a single hydrophone," J. Acoust. Soc. Am. **121**, 1519–1528 (2007).

[10]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).

[11]N. Xiang, P. M. Goggans, T. Jasa, and M. Kleiner, "Evaluation of decay times in coupled spaces: Reliability analysis of Bayesian decay time estimation," J. Acoust. Soc. Am. **117**, 3705–3715 (2005).

[12]N. Xiang and T. Jasa, "Evaluation of decay times in coupled spaces: An efficient search algorithm within the Bayesian framework," J. Acoust. Soc. Am. **120**, 3744–3749 (2006).

[13]R. M. Neal, "Slice sampling," Ann. Stat. **31**, 705–767 (2003).

[14]J. E. Summers, R. R. Torres, Y. Shimizu, and B.-I. L. Dalenbäck, "Adapting a randomized beam-axis-tracing algorithm to modeling of coupled rooms via late-part ray tracing," J. Acoust. Soc. Am. **118**, 1491–1502 (2005).

[15]D. T. Bradley and L. M. Wang, "The effects of simple coupled volume geometry on the objective and subjective results from nonexponential decay," J. Acoust. Soc. Am. **118**, 1480–1490 (2005).

[16]A. Billon, V. Valeau, A. Sakout, and J. Picaut, "On the use of a diffusion model for acoustically coupled rooms," J. Acoust. Soc. Am. **120**, 2043–2054 (2006).

[17]M. Meissner, "Computational studies of steady-state sound field and reverberant sound decay in a system of two coupled rooms," Cent. Eur. J. Phys. **5**, 293–312 (2007).

[18]E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).

[19]C. C. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer-Verlag, New York, 1999).

Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range

Wolfgang Kreuzer, Piotr Majdak, and Zhengsheng Chen
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

(Received 12 March 2009; revised 15 June 2009; accepted 16 June 2009)

Head-related transfer functions (HRTFs) play an important role in spatial sound localization. The boundary element method (BEM) can be applied to calculate HRTFs from non-contact visual scans. Because of high computational complexity, HRTF simulations with BEM for the whole head and pinnae have only been performed for frequencies below 10 kHz. In this study, the fast multipole method (FMM) is coupled with BEM to simulate HRTFs for a wide frequency range. The basic approach of the FMM and its implementation are described. A mesh with over 70 000 elements was used to calculate HRTFs for one subject. With this mesh, the method allowed to calculate HRTFs for frequencies up to 35 kHz. Comparison to acoustically-measured HRTFs has been performed for frequencies up to 16 kHz, showing a good congruence below 7 kHz. Simulations with an additional shoulder mesh improved the congruence in the vertical direction. Reduction in the mesh size by 5% resulted in a substantially-worse representation of spectral cues. The effects of temperature and mesh perturbation were negligible. The FMM appears to be a promising approach for HRTF simulations. Further limitations and potential advantages of the FMM-coupled BEM are discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3177264]

PACS number(s): 43.64.Ha, 43.20.Fn, 43.66.Qp [BLM] Pages: 1280–1290

I. INTRODUCTION

The shape of head, torso, and pinna plays an important role in localization of sounds in humans. Reflections, especially at the pinna, act as a filter, which can be described by the head-related transfer functions (HRTFs) (Blauert, 1974; Shaw, 1974; Møller *et al.*, 1995). These functions are dependent on sound source position (Makous and Middlebrooks, 1990) and they differ among listeners (Wightman and Kistler, 1989; Algazi *et al.*, 2001b). HRTFs can be applied to create virtual free-field sounds (Bronkhorst, 1995; Begault *et al.*, 2001). The required spatial resolution of HRTFs is given by the listeners' spatial localization accuracy, which is in the range of few degrees (Minnaar *et al.*, 2005). Thus, HRTFs must be measured for many directions, especially when virtual sounds in vertical planes are required. An acoustic measurement of one HRTF set including all positions in three-dimensional (3D)-space takes tens of minutes, even when sophisticated measurement methods are applied (Zotkin *et al.*, 2006; Majdak *et al.*, 2007). This may be uncomfortable for the subject, who has to keep still during the entire measurement procedure.

However, the data about subjects' morphology can also be collected using non-contact visual scans (Katz, 2001a; Kahana and Nelson, 2007). This procedure is very fast and compared to the acoustic measurements, it is much more comfortable for the subjects. From the visual data, it is possible to create surface meshes of the subject's head and then, in principle, numerically calculate HRTFs. Hence, a numerical method for accurate calculation of HRTFs from visual data is of great interest.

In the past years, the boundary element method (BEM) became more and more popular in the field of acoustic simulation. Katz (2001a) described results of HRTF simulations using BEM. From visually scanned data, he simulated HRTFs for frequencies up to 6 kHz and compared them to acoustically-measured data. However, for the median-plane localization, frequencies in the range of 3.5–16 kHz are essential (Middlebrooks and Green, 1991). Recently, Kahana and Nelson (2007) calculated HRTFs for frequencies up to 20 kHz; however, their method allowed only to use a baffled pinna without the head and torso.

One reason for frequency limitations in Katz, 2001a and Kahana and Nelson, 2007 was the size of the mesh they used. When applying BEM to solve acoustic problems, at least six elements per wavelength are required to ensure numerical accuracy (Marburg, 2002). Thus, a mesh resolution of few millimeters is required when applying BEM for higher frequencies. Katz (2001a) used a head mesh with 22 000 elements, which was valid for frequencies up to 5.4 kHz. Kahana and Nelson (2007) used baffled pinna meshes with 23 000 elements, which were valid for frequencies up to 20 kHz. To overcome the frequency limitation for combined pinna, head, and torso meshes, we used meshes with over 70 000 elements. However, simulation of such large meshes leads to a huge linear system of equations. The memory requirement for solving such a matrix equation is $\mathcal{O}(n^2)$, where n is the number of elements. Assuming 128-bit complex-valued entries, the memory requirement for just storing the matrix exceeds 70 Gbytes. Thus, with the memory and computation limitations of modern computer systems, an approach to reduce the required memory is essential to be able to calculate HRTFs for high frequencies.

In this study, the reduction in the required memory is achieved by coupling the BEM with the fast multipole method (FMM) (Greengard and Rokhlin, 1987). FMM was originally developed for the numerical computation of N-body problems and was later adapted to acoustic problems

(Greengard *et al.*, 1998). It reduces the computational complexity to $\mathcal{O}(n \log^2 n)$ (Fischer *et al.*, 2004). Thus, it appears reasonable to use the FMM-coupled BEM approach to simulate HRTFs within a wide frequency range.

In principle, the theory follows, Fischer and Gaul (2005), and Chen *et al.* (2008). Various additions and modifications were applied to adapt those algorithms to an efficient HRTF simulation. In Sec. II, we describe the resulting algorithm, including a brief overview of the BEM and the FMM.[1] Then, several computational issues are discussed and the validation of the code is presented. Finally, the results of HRTF simulation for one subject are presented and compared to the data from acoustic measurements. In addition, the effects of temperature, mesh quality, and mesh perturbation are presented. Finally, advantages and limitations of our approach are discussed.

II. THEORY

A. BEM

The general equation for exterior acoustic problems in a domain Ω is the Helmholtz equation:

$$\nabla^2\Phi(\boldsymbol{x}) + k^2\Phi(\boldsymbol{x}) = 0, \quad \boldsymbol{x} \in \Omega, \tag{1}$$

$$\alpha\frac{\partial\Phi(\boldsymbol{x})}{\partial n} + \beta\Phi(\boldsymbol{x}) = f, \quad \boldsymbol{x} \in \Gamma, \tag{2}$$

$$\left|\frac{\partial\Phi(\boldsymbol{x})}{\partial r} - ik\Phi(\boldsymbol{x})\right| \leqslant \frac{c}{\sqrt{r}}, \quad r = |\boldsymbol{x}| \to \infty. \tag{3}$$

Equations (2) and (3) define mixed boundary conditions in $\Gamma = \partial\Omega$ and the Sommerfeld radiation condition, respectively. $\Phi(\boldsymbol{x}) = p/(i\omega\rho)$ denotes the velocity potential at the point $\boldsymbol{x}$ and $k = \omega/c$ is the wavenumber with the frequency ω and the speed of sound c. p denotes the sound pressure, ρ the density of the fluid, and $i^2 = -1$. Γ is the boundary of the domain (in our case the surface of the head). α, β, and f are parameters and functions which determine appropriate boundary conditions.

To ensure uniqueness of the solution also at irregular frequencies the Burton–Miller approach is used (Burton and Miller, 1971).[2] From Eq. (1), the boundary integral equation is derived (Chen *et al.*, 2008):

$$\frac{1}{2}\left[\frac{i}{k}A(\boldsymbol{x}) - 1\right]\Phi(\boldsymbol{x}) + L[\Phi](\boldsymbol{x}) + \frac{i}{k}\frac{\partial}{\partial\boldsymbol{n}_x}L[\Phi](\boldsymbol{x}) = \frac{i}{k}\left[\frac{1}{2}v_0(\boldsymbol{x}) - v_i(\boldsymbol{x})\right] - \Phi_i(\boldsymbol{x}) + L[v_0](\boldsymbol{x}) + \frac{i}{k}\frac{\partial}{\partial\boldsymbol{n}_x}L[v_0](\boldsymbol{x}), \tag{4}$$

with

$$L[\Phi](\boldsymbol{x}) := \int_\Gamma [G(\boldsymbol{x},\boldsymbol{y})A(\boldsymbol{y}) + H(\boldsymbol{x},\boldsymbol{y})]\Phi(\boldsymbol{y})d\boldsymbol{y}, \tag{5}$$

$$L[v_0](\boldsymbol{x}) := \int_\Gamma G(\boldsymbol{x},\boldsymbol{y})v_0(\boldsymbol{y})d\boldsymbol{y}. \tag{6}$$

Φ_i and v_i are the potential field and the particle velocity of an incident sound wave, respectively. v_0 is the velocity at the surface node $\boldsymbol{x}$ in the normal direction $\boldsymbol{n}_x$. By including $A(\boldsymbol{x}) := i\omega\rho a(\boldsymbol{x}) := v(\boldsymbol{x})/\Phi(\boldsymbol{x})$, sound absorbing materials defined by admittance $a(\boldsymbol{x})$ can be modeled. $G(\boldsymbol{x},\boldsymbol{y})$ is the Green's function of the Helmholtz equation and $H(\boldsymbol{x},\boldsymbol{y})$ is its derivative with respect to the normal vector $\boldsymbol{n}_y$ to the surface Γ at a point $\boldsymbol{y}$:

$$G(\boldsymbol{x},\boldsymbol{y}) := \frac{e^{ikr}}{4\pi r}, \tag{7}$$

$$H(\boldsymbol{x},\boldsymbol{y}) := \frac{\partial G(\boldsymbol{x},\boldsymbol{y})}{\partial\boldsymbol{n}_y} = \frac{e^{ikr}}{4\pi r}\left(ik - \frac{1}{r}\right)\frac{\partial r}{\partial\boldsymbol{n}_y},$$

$$r := \|\boldsymbol{y} - \boldsymbol{x}\|. \tag{8}$$

The solutions for most of the integrals in Eqs. (5) and (6) are well-defined. However, the integrand of

$$\frac{\partial}{\partial\boldsymbol{n}_x}\int_\Gamma \frac{\partial G(\boldsymbol{x},\boldsymbol{y})}{\partial\boldsymbol{n}_y}\Phi(\boldsymbol{y})d\Gamma_y \tag{9}$$

becomes hypersingular when $\boldsymbol{x}$ lies on Γ. In our study, collocation and constant boundary elements are used. In that case, Eq. (9) can be converted to a sum of a line integral and a weakly singular integral. A numerical solution is given using appropriate quadrature schemes (for more details see Erichsen and Sauter, 1998; Chen *et al.*, 2008).

The numerical treatment of Eq. (4) yields an $n \times n$ system matrix. For large meshes, i.e., large n, the memory requirements are very high. Thus, simulation of meshes with tens of thousand of elements is not feasible without further modification of the collocation BEM.

B. Fast multipole BEM

In this paper, a short introduction of the FMM is given. For more details see, for example, Greengard and Rokhlin, 1987; Fischer and Gaul, 2005; Chen *et al.*, 2008; Gumerov and Duraiswami, 2009. The idea behind the FMM is the far-field expansion of kernels in Eqs. (7) and (8) (Greengard and Rokhlin, 1987):

$$F(\boldsymbol{x},\boldsymbol{y}) \approx \sum_{i,j=0}^{N} s_{ij}F_i^I(\boldsymbol{x})F_j^{II}(\boldsymbol{y}). \tag{10}$$

To separate the field in far field and near field, the mesh is divided into clusters. A cluster C_2 is in the far field $\mathcal{F}_1$ of cluster C_1 if the two clusters are well separated. Two clusters C_1 and C_2 are well separated, if $\|z_1 - z_2\| > \tau(r_1 + r_2)$, where r_1 and r_2 are the radii of the clusters and z_1 and z_2 are their midpoints (see Fig. 1). In our study, τ was $\sqrt{5}/2$. In the far field, the contribution of all combinations of $\boldsymbol{x} \in C_2$ and $\boldsymbol{y} \in C_1$ to the integrals in Eqs. (5) and (6) is reduced to that of the cluster centers z_2 and z_1. If a cluster C_2 is not in $\mathcal{F}_1$ then it is in the near field $\mathcal{N}_1$.

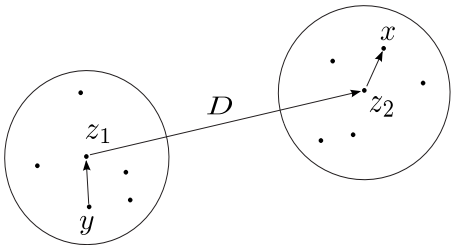


FIG. 1. Clusters around $\boldsymbol{x}$ and $\boldsymbol{y}$.

The kernel expansion of the Green's function [Eq. (7)] used in this study is given in Fischer and Gaul, 2005:

$$G(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{ik\|\boldsymbol{x}-\boldsymbol{y}\|}}{\|\boldsymbol{x}-\boldsymbol{y}\|} = \frac{e^{ik\|\boldsymbol{D}+\boldsymbol{d}\|}}{\|\boldsymbol{D}+\boldsymbol{d}\|}$$
$$= \frac{ik}{4\pi}\sum_{l=0}^{\infty}(2l+1)i^l h_l^{(1)}(k\|\boldsymbol{D}\|)\int_{\mathbb{S}} e^{ik\boldsymbol{s}\boldsymbol{d}}P_l(\boldsymbol{s}\hat{\boldsymbol{D}})ds$$
$$\approx \frac{ik}{4\pi}\int_{\mathbb{S}} e^{ik\boldsymbol{s}(\boldsymbol{x}-\boldsymbol{x}_2)}M_L(\boldsymbol{s},\boldsymbol{D})e^{-ik\boldsymbol{s}(\boldsymbol{y}-\boldsymbol{z}_1)}ds, \tag{11}$$

with

$$M_L(\boldsymbol{s},\boldsymbol{D}) := \sum_{l=0}^{L}(2l+1)i^l h_l^{(1)}(k\|\boldsymbol{D}\|)P_l(\boldsymbol{s}\hat{\boldsymbol{D}}), \tag{12}$$

where $\boldsymbol{d} := \boldsymbol{x}-\boldsymbol{z}_2-(\boldsymbol{y}-\boldsymbol{z}_1)$, $\boldsymbol{D} := \boldsymbol{z}_2-\boldsymbol{z}_1$, and $\hat{\boldsymbol{D}} := \boldsymbol{D}/\|\boldsymbol{D}\|$. It is assumed that $|\boldsymbol{D}|>|\boldsymbol{d}|$, i.e., $\boldsymbol{x}$ and $\boldsymbol{y}$ are well separated. $h_l^{(1)}(\cdot)$ denote the spherical Hankel functions of the first kind of order l, and $P_l(\cdot)$ are l-th order Legendre polynomials. $\mathbb{S}$ is the unit-sphere surface given by $\{(\cos\phi\sin\theta,\sin\phi\sin\theta,\cos\theta):0\leqslant\phi\leqslant 2\pi,0\leqslant\theta\leqslant\pi\}$. The truncation parameter L was $\max\{2kr_{\max}+1.8\log(2kr_{\max}+\pi),8\}$, where $r_{\max}$ is the radius of the largest cluster (Chen *et al.*, 2008). For all other kernels in Eq. (7), similar expansions can be found.

For the far field, the numeric treatment of Eq. (4) requires calculation of potentials of the form $\Phi(\boldsymbol{x})=\Sigma_{j=1}^{J}q_jG(\boldsymbol{x},\boldsymbol{y}_j)$, where $\boldsymbol{x}\in C_2$, $\boldsymbol{y}_j\in C_1$, q_j is the source strength at $\boldsymbol{y}_j$, and J is the number of nodes in the cluster C_1. The multipole approach to calculate $\Phi(\boldsymbol{x})$ consists of three steps. First, the far-field signature $F(s)$ is calculated:

$$F(s) = ik\sum_{j=1}^{J}e^{ik(\boldsymbol{z}_1-\boldsymbol{y}_j)s}q_j. \tag{13}$$

This step represents the local expansion of the cluster C_1 around $\boldsymbol{z}_1$. Second, the near-field signature $N(s)$ is calculated by applying the far-field signature $F(s)$ to the translation operator M_L for all combinations of clusters C_1 and C_2:

$$N(s) = M_L(s,\boldsymbol{D})F(s). \tag{14}$$

This step represents the translation of the far-field signature around $\boldsymbol{z}_1$ to the near-field signature around $\boldsymbol{z}_2$. It is numerically efficient because $M_L(s,\boldsymbol{D})$ only operates on the cluster centers. Finally, the potential $\Phi(\boldsymbol{x})$ is calculated:

$$\Phi(\boldsymbol{x}) = \frac{1}{4\pi}\int_{\mathbb{S}} e^{ik(\boldsymbol{x}-\boldsymbol{z}_2)s}N(s)ds. \tag{15}$$

This integral is calculated using Gauss–Legendre quadrature with L points in the θ-direction and a $2L$-point trapezoidal rule for the ϕ-direction (Rahola, 1996), where L is the length of the multipole expansion from Eq. (12). This step represents the local expansion of the near-field signature around $\boldsymbol{z}_2$.

For the near field, the multipole expansion can not be applied. Thus, for all nodes $\boldsymbol{y}$ in the near field of $\boldsymbol{x}$, collocation BEM is applied to set up the near-field matrix. Thus, the size of the near-field matrix depends on the number of nodes in the near field.

By applying the above steps to all clusters, the final system of equations is formally written as

$$(\boldsymbol{N}+\boldsymbol{SMT})\boldsymbol{u} = \boldsymbol{f}, \tag{16}$$

where $\boldsymbol{u}$ is the vector of the unknown potentials and $\boldsymbol{f}$ is the excitation force. $\boldsymbol{T}$ represents the far-field signatures for each cluster [Eq. (13)]. $\boldsymbol{M}$ represents translation operators M_L for each cluster pair [Eq. (14)]. $\boldsymbol{S}$ represents the local expansions of the near-field signatures. $\boldsymbol{N}$ is a block-diagonal matrix, which represents the near-field matrices from the collocation BEM for each cluster.

Because of the sparse form of Eq. (16), a substantial reduction in memory requirement is achieved. This makes the fast multipole BEM applicable for large meshes. In fact, the largest matrix, which has to be fully stored is the near-field matrix. Its size depends on the cluster size. Giebermann (1997) showed that the cluster size of $\sqrt{n}$ results in a maximum efficiency if only one level of clustering is used.

Further reduction in computational cost can be achieved with the multilevel FMM (see, for example, Fischer and Gaul, 2005). In the multilevel FMM, a binary tree of clusters is implemented. At its coarsest level ($\ell=0$), only one cluster is given by a parallelepiped which contains the whole mesh. Clusters at level ℓ are given by bisection of clusters from level $\ell-1$. The finest level $\ell_{\max}$ is reached when all clusters contain less than 20 elements. In our study, a modified multilevel FMM was used. At level $\ell=1$, instead of bisection, the cluster from level $\ell=0$ was divided into about $\sqrt{n}$ clusters. Clusters at finer levels were then constructed by bisection. This modification resulted in a smaller number of levels. In matrix form, this procedure yields

$$\left(\boldsymbol{N}_{\ell_{\max}}+\sum_{\ell=1}^{\ell_{\max}}\boldsymbol{S}_\ell\boldsymbol{M}_\ell\boldsymbol{T}_\ell\right)\boldsymbol{u} = \boldsymbol{f},$$

where $\boldsymbol{N}_{\ell_{\max}}$ is the near-field matrix at the finest level and $\boldsymbol{S}_\ell$ and $\boldsymbol{T}_\ell$ are the matrices $\boldsymbol{S}$ and $\boldsymbol{T}$, respectively, for the particular level ℓ.

The FMM shows stability problems for low frequencies (Darve, 2000). Thus, for frequencies below 1 kHz, only one level with $\sqrt{n}$ clusters was used ($\ell_{\max}=1$). This increased the stability for frequencies as low as 50 Hz. For frequencies above 1 kHz, the modified multilevel FMM was used. For our mesh sizes, the clustering procedure resulted in up to three levels ($\ell_{\max}\leqslant 3$).

Fischer and Gaul (2005) calculated the $\boldsymbol{S}_\ell$ and $\boldsymbol{T}_\ell$ only for the finest level. For coarser levels, they applied filtering and interpolation algorithms to obtain $\boldsymbol{S}_\ell$ and $\boldsymbol{T}_\ell$. This further reduced the memory requirement in return for higher computational cost. In our modified multilevel FMM, $\boldsymbol{S}_\ell$ and $\boldsymbol{T}_\ell$ were explicitly calculated and stored in memory for all levels. This was possible because for our mesh sizes, the number of levels was small enough to allow storage of all $\boldsymbol{S}_\ell$ and $\boldsymbol{T}_\ell$ for all levels in memory. Thus, compared to Fischer and Gaul (2005), a reduction in computational cost was achieved in return for a slightly higher memory requirement.

III. METHODS

A. Measurement of HRTFs

The HRTFs were measured for one subject in a semi-anechoic room. The subject had very short hair and was not wearing a cap. The *A*-weighted sound pressure level of the background noise in this room was 18 dB re 20 μPa on a typical testing day. The temperature was between 20 and 25 °C. Twenty-two loudspeakers (custom-made boxes with VIFA 10 BGS as drivers; the variation in the frequency response was ±4 dB in the range from 200 to 16 000 Hz) were mounted at fixed elevations from −30° to 80° with a spacing of 5°. The subject was seated in the center of the arc and had microphones (KE-4-211-2, Sennheiser) placed in his ear canals. The microphones were connected via preamplifiers (FP-MP1, RDL) to the digital audio interface. An exponential sweep with a duration of 1728.8 ms and a frequency from 50 Hz to 18 kHz was used to measure each HRTF. The sweep had a fade in/out of 20 ms. The multiple exponential sweep method was applied to measure HRTFs in an interleaved and overlapped order for one azimuth and all elevations at once (Majdak *et al.*, 2007). Then, the subject was rotated by 2.5° to measure HRTFs for the next azimuth. In total, 1550 HRTFs were measured with the positions distributed with a constant spherical angle on the sphere. The measurement procedure lasted for approximately 20 min.

Then, reference measurement was performed, in which in-ear microphones were placed in the center of the arc and the system identification procedure was performed for all loudspeakers. From the reference measurement, equipment transfer functions were derived. They were used to remove the effect of the equipment, which was done by dividing complex spectra of HRTFs by the complex spectra of the equipment transfer functions. In the next step, the directional transfer functions (DTFs) were calculated (Middlebrooks, 1999). The magnitude of the common transfer function (CTF) was calculated by averaging the log-amplitude spectra of all equalized HRTFs. The phase of the CTF was the minimum phase corresponding to the amplitude spectrum of the CTF. The DTFs were the result of filtering the HRTFs with the inverse complex CTF. Finally, all DTFs were temporally-windowed with a 5.33-ms long Tukey window. The DTFs had a valid frequency range from 300 Hz to 16 kHz.

B. Mesh generation

Visual scans of the subject's head were performed using a non-contact 3D scanner (Minolta VIVID-900). Hair was a major problem for the scanner, and, thus, a rubber cap was used to cover the hair (Katz, 2001a; Kahana and Nelson, 2007). The subject did not wear the in-ear-microphones from the acoustic measurements. The mesh was generated by combining six scans from different directions around the head using the "surface wrap" tool implemented in Geomagic Studio (Geomagic, Inc.). For the head, the resolution of the scanned data was higher then necessary; therefore, the node reduction mode has been used during the surface wrap. For the pinna, the surface wrap was applied on the full-resolution scanned data without node reduction. Then, the pinna and the head meshes were stitched together. The final mesh was manually edited to obtain almost regular triangles. This was done to enhance numerical performance and stability. Parts of the antihelix and concha have been coarsened using the "remove doubles" procedure from Blender (Roosendaal and Selleri, 2004). This could be safely done because these parts showed almost no curvatures and could be represented by less nodes without any structural changes. We assume that our post-processing procedure had not substantially affected the volume of the mesh.

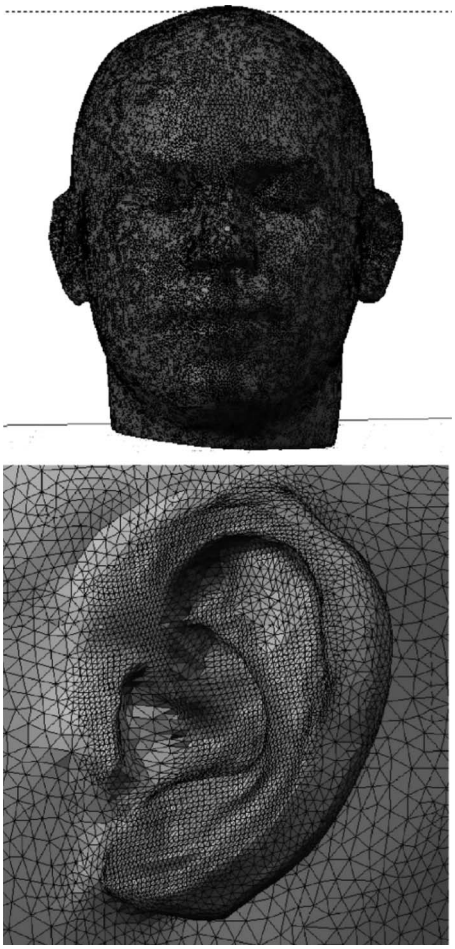


FIG. 2. Mesh used in the baseline condition.

The final mesh contained 70 785 elements with an average edge length of 2.1 mm, with a minimum of 0.14 mm and a maximum of 5.8 mm. Most of the small elements were located at the pinna. This mesh, shown in Fig. 2, represents the mesh for the baseline condition. Based on the 6-to-8-

elements-per-wavelength rule, this mesh can be used for BEM calculations for frequencies up to 35 kHz.

Our method allows to use impedance boundary conditions to simulate sound absorbing materials (for example hair, see Katz, 2001b). Preliminary experiments with different admittance boundary conditions for the hair area showed no substantial differences in the results. Hence, in this study, we present results for an acoustically-hard reflecting head only.

1550 nodes around the head at a distance of 1.2 m were chosen as point sources to represent the positions of the loudspeakers from the HRTF measurement. A receiver element was positioned at the entrance of the closed ear canal to represent the position of the microphone from the HRTF measurement. The receiver element was implemented by setting the velocity boundary condition at that specific element to a value different from 0. The position of the receiver element was chosen based on the photographs of subject's pinna (see also Fig. 6).

C. Reciprocity

The principle of reciprocity was implemented to further speed up calculations. The role of the receiver element was interchanged with role of the point sources. Let x_1 be the midpoint of the receiver element, and x_2 be the position of a point source outside the head. Using an analogon of Betti's reciprocal theorem for acoustics, the sound pressure $p_{x_1}(x_2)$ caused by an excitation at x_1 is related to the sound pressure $p_{x_2}(x_1)$ caused by an excitation at x_2:

$$p_{x_1}(x_2) = \frac{-qp_{x_2}(x_1)}{i\omega\rho A_0 v_{n_0}}, \tag{17}$$

where A_0 is the area of the vibrating element, v_{n_0} is the normal velocity at the midpoint of x_1, ρ is the density of the medium, and q is the intensity of the sound source positioned at x_2.

Hence, the receiver element at the entrance of the ear channel was defined to be an active vibrating element. The sound pressure was calculated at the nodes representing the position of the sound sources. This is a very efficient approach, because the solution of Eq. (4) for one active element results in the sound pressure information for all nodes. Thus, contrary to the direct method, with the reciprocity method, HRTFs for all sound source positions were calculated in one simulation at once.

D. Computational issues

Several pre-simulations were performed to evaluate our code. First, the numerical stability of the reciprocity method was tested. In a direct simulation, a point source was positioned in front of the head at a distance of 1.2 m. Sound pressure at each element of the baseline mesh was calculated for different frequencies. Figure 3 shows the amplitude spectra of calculated HRTFs as a comparison between the direct (lines) and reciprocity (symbols) methods for four positions (front, right, back, and left). All calculations were done for the right ear. In general, the results do not substantially differ

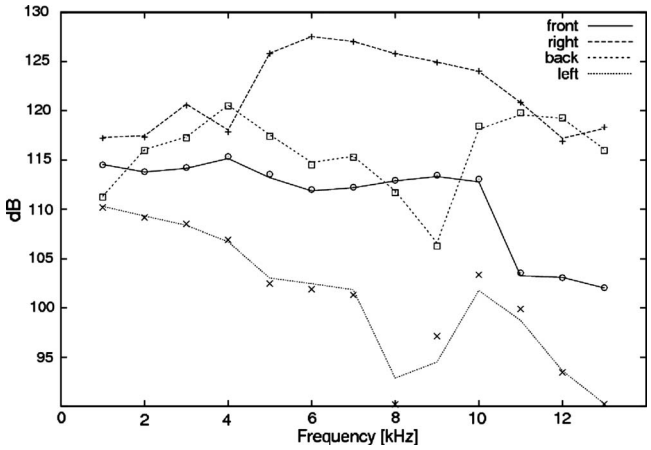


FIG. 3. Comparison of four different HRTFs using the direct approach (lines) and the results with the reciprocity method (symbols). All values show the sound pressure at the entrance of the right ear channel; sound sources are positioned in front, right, back and left of the head.

for both methods. However, for the contralateral position (left), the limitations of the reciprocity method are evident. This is because of round-off errors for positions and frequencies, which contain low energy. Nevertheless, such accuracy is sufficient for further simulations.

The evaluation of the FMM-coupling to the BEM was done by directly comparing the HRTFs calculated using collocation BEM (without FMM) and FMM-coupled BEM. Because of memory limitations, the calculations using collocation BEM could not be performed for the baseline mesh. To be able to still provide a fair comparison, a simplified mesh was used where the original pinnae were placed on a sphere representing an artificial head. This mesh consisted of a smaller number of nodes and still provided a high geometric complexity. The calculation results, i.e., HRTFs for the directions front, right, back, and left, are shown in Fig. 4. The lines represent the results for the collocation BEM and the symbols represent the results for the FMM-coupled BEM. The results show no differences between the BEM with and without FMM.

The computational complexity was evaluated by calculating sound pressure for a simple 3D cube. Figure 5 shows the time and memory requirements for the calculation of sound pressures for cubes with different number of elements. The memory requirement is represented by the number of

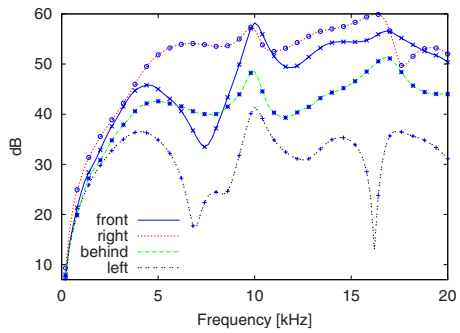


FIG. 4. (Color online) Comparison of four different HRTFs using the collocation BEM (without FMM, solid lines) and FMM-coupled BEM (symbols). A simplified head mesh was used (see text).

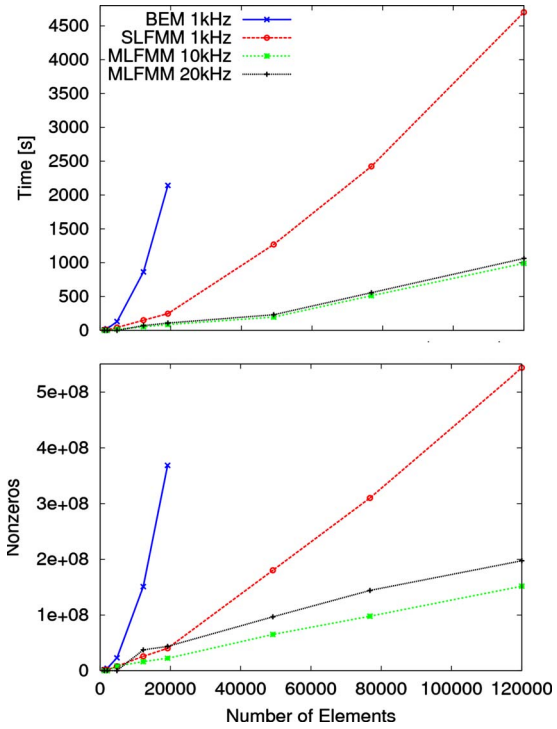


FIG. 5. (Color online) Computation time and memory requirement for different methods of calculation as functions of the mesh size. Compared are the collocation BEM without FMM, BEM with the single-level FMM ($\ell_{max}=1$), and BEM with the three-level FMM ($\ell_{max}=3$) for frequencies of 10 and 20 kHz. The calculations were performed for simple 3D cube meshes.

nonzeros in the system matrix. For the frequency of 1 kHz, the FMM with one level ($\ell_{max}=1$) was used. For frequencies of 10 and 20 kHz, FMM with three levels ($\ell_{max}=3$) was used. For BEM without FMM and meshes with more then 19 200 elements, results could not be calculated because of too high memory requirements. This comparison clearly shows the limitation of collocation BEM without FMM and advantages of coupling FMM to BEM in simulations.

In the main simulations, the HRTFs were calculated in the frequency range of 0.2 and 20 kHz in steps of 0.2 kHz. The overall computation time for 200 different frequencies was about 5 h on a Linux cluster containing five machines with dual Opteron processors (AMD) running with 2.0 GHz. Finally, based on the simulated HRTFs, DTFs were calculated in the same way as for the measured data. Even though the DTFs were calculated for frequencies up to 20 kHz, the comparisons to the measured DTFs (presented in Sec. IV) have been done only for frequencies up to 16 kHz because of the limited frequency range of the measured DTFs.

E. Parameters

The effects of the mesh quality, mesh perturbation, temperature, and shoulders were investigated by altering the mesh and simulation parameters with respect to the baseline condition. In the baseline condition, the mesh shown in Fig. 2 was used and the speed of sound and the density of air were set to simulate the temperature of 15 °C (see Table I). Figure 6 shows the subject's pinna (panel a) and its corresponding mesh (panel b) used in the baseline condition.

TABLE I. Speed of sound c and air density ρ for the tested temperatures.

Temp (°C)	c (m/s)	ρ (kg/m^3)
0	331.3	1.292
15	340.5	1.225
30	349.0	1.165

The mesh quality was tested by reducing the number of elements in the mesh by approximately 5%. The reduction in elements was performed in two steps. First, a smoothing algorithm was applied to the mesh in terms of moving each vertex in the mesh toward the barycenter of the linked vertices. Then, all nodes within a 0.43-mm distance to their neighbors were removed. The resulting low-quality mesh contained 67 428 elements with an average edge length of 2.19 mm. The modifications of the mesh were done using the software package BLENDER (Roosendaal and Selleri, 2004). Figure 6(c) shows the pinna from the low-quality mesh. As it can be seen the reduction of the nodes at the pinna had a major effect on the shape.

The stability of the mesh with respect to measurement errors was tested by applying a perturbation to the mesh. This was achieved by moving all nodes in the mesh in random directions. This procedure corresponds to degradation of the precision of the visual scans and tests the robustness and stability of the simulation to such changes. The perturbation level was represented by the length of the vectors with random direction added to each node. Two perturbation levels were used: 0.25 and 0.5 mm. Figure 6(d) shows an example of a pinna mesh, which was perturbed at the level of 0.5 mm.

The effect of the simulation temperature was investigated by varying the sound speed and the air density. In addition to the baseline condition, two temperatures were

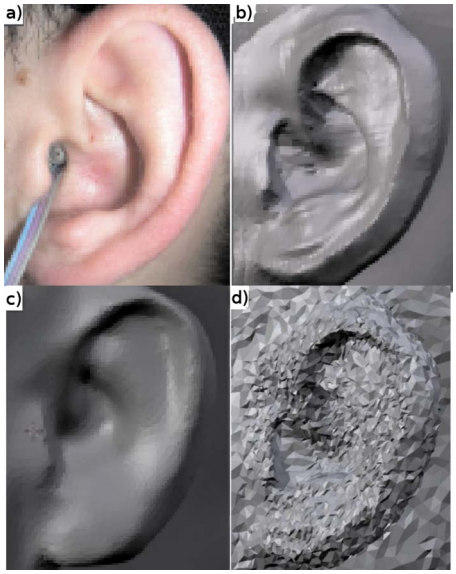


FIG. 6. (Color online) Panel a: Left pinna. Panel b: The mesh for the baseline condition. Panel c: Low-quality mesh. Panel d: Example of a mesh perturbed with random vectors of 0.5 mm length.

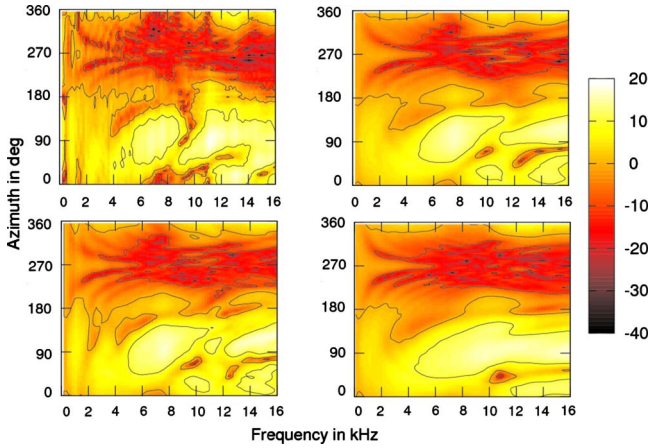


FIG. 7. (Color online) Comparison of the simulated and measured DTFs in the horizontal plane (elevation angle of 0°). Top-left: Measured data. Top-right: Simulation for the baseline condition. Bottom-left: Simulation for the baseline condition with shoulders. Bottom-right: Simulation for the low-quality mesh. The color represents the magnitude in dB.

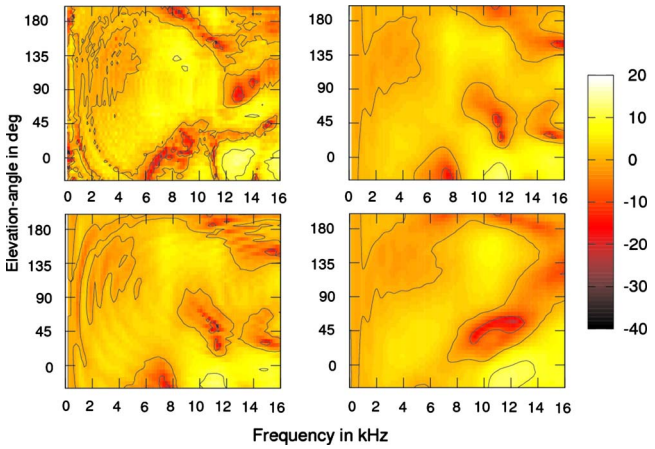


FIG. 8. (Color online) Comparison of the simulated and measured DTFs in the median plane (azimuth of 0°). All other conventions are as in Fig. 7.

simulated: 0 and 30 °C. The corresponding values for the sound speed and air density are given in Table I.

The effect of shoulders was investigated by including a shoulder mesh to the simulations, which was based on data from two-dimensional photographs of the subject from two different angles and did not require additional visual 3D scans. The shoulder mesh was combined with the head mesh from the baseline condition. The resulting mesh consisted of 3711 additional elements.

IV. RESULTS AND DISCUSSION

For the horizontal plane, the simulation and measurement results are shown in Fig. 7. The panels show the DTF amplitude spectra as functions of the azimuth of the sound source. The azimuths of 0°, 90°, 180°, and 270° represent the sound sources in the front, to the right, in the back, and to the left of the subject, respectively. The color represents the DTF amplitude in decibels. The top-left panel shows the measured data and the top-right panel shows the simulation results for the baseline condition. For the baseline condition, the general pattern is in agreement with that for the measured data. The broadband amplitude decreases when the sound source moves to the contralateral ear. This is because of the shadow caused by the head, which is consistent with the measurement results. The amplitude fluctuates along the azimuth more for the high frequencies than for the low frequencies. This is because for the high frequencies, the fine structure of the pinna leads to azimuth-dependent resonances and cancellations. For low frequencies, the pinna has little effect only. This is also in agreement with the measurement results. However, the spatial-spectral features are not exactly represented by the simulation. For example, for azimuth of 150°, the measured data show a notch between 9 and 10 kHz. This notch is not represented in the simulation results. Such discrepancies are not essential for the localization in the horizontal plane, where spectral cues play a minor role (Macpherson and Middlebrooks, 2002)

The effect of mesh quality on the simulation results for the horizontal plane is shown in the bottom-right panel of Fig. 7. The spatial-spectral features appear smeared and are not as prominent as in the baseline condition. The quality degradation had a substantial effect on particular features; however, the general pattern remained similar to that from the baseline condition and measured data.

The bottom-left panel of Fig. 7 shows the results of simulation with shoulder mesh. No substantial effects of the shoulder can be found for the horizontal plane.

For the median plane, the results are shown in Fig. 8. The panels show the DTF amplitude spectra as functions of the elevation angle. The elevation angles of 0°, 90°, and 180° represent the sound sources at eye-level in the front, at the top, and at eye-level in the back of a listener, respectively. Note that for the elevation angles between 80° and 100°, the data were linearly interpolated from 80° to 100° because they were not measured in that region. The top-left panel shows the measured data and the top-right panel shows the simulation results for the baseline condition.

For frequencies below 7 kHz, the DTFs show a striking congruence between simulation and measurement. However, the measured data show more modulations, which are most likely a result of comb filter effects caused by the shoulder reflections (Algazi *et al.*, 2001a). The bottom-left panel of Fig. 8 shows the results for simulation with the shoulder mesh. By including shoulders to the model, the modulations of the spatial-spectral patterns became clearly present. This may reduce the degradation of the vertical-plane localization ability, especially for sound source located away from the median-plane (Algazi *et al.*, 2001a). Thus, for the frequencies below 7 kHz, the mesh with shoulders provides the best congruence to the measured data.

For frequencies above 7 kHz, the differences between measurement and simulation are evident. The measurement results show a deep notch at 7 kHz for the eye-level positions. Such a notch is considered to be one of the main cues for encoding elevation (Carlile and Pralong, 1994; Middlebrooks, 1997; Iida *et al.*, 2007). The center frequency of this notch increases to 9 kHz with increasing elevation for angles up to 40°. This pattern is symmetric across the hemifields. In the simulation results, the elevation-dependent center frequency of the notch is not present. The only correspondence

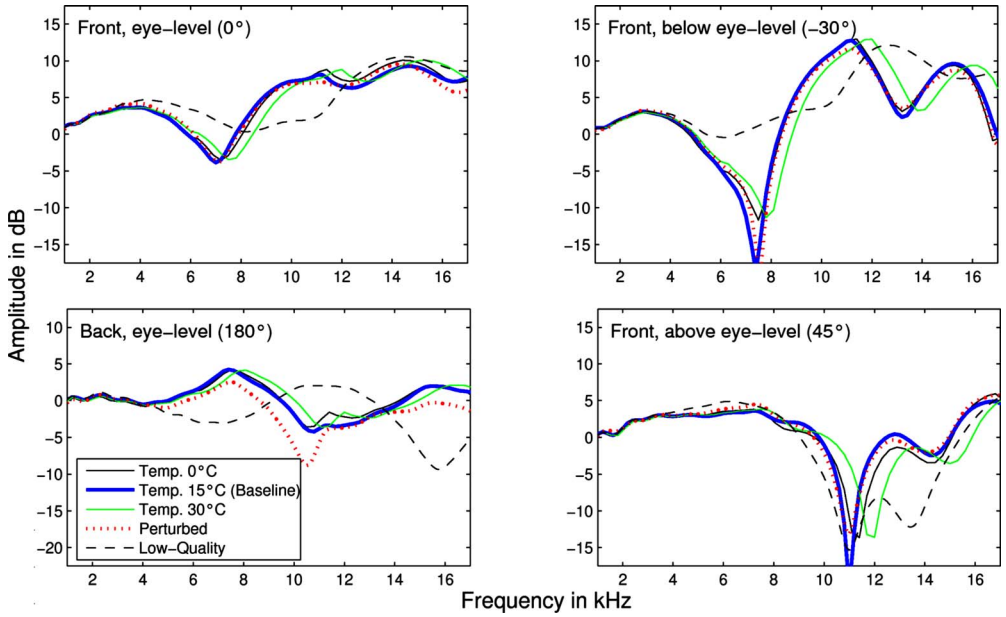


FIG. 9. (Color online) Amplitude spectra of simulated DTFs for sound sources in the median plane. The elevations are −30° (front, below eye-level), 0° (front, eye-level), 45° (front, above eye-level), and 180° (back, eye-level). The solid lines show the effect of the temperature. The dotted lines show the effect of perturbation at a level of 0.5 mm. The dashed lines show the effect of using low-quality mesh.

to the measured data can be observed for the frontal positions below eye-level, where the simulation results also show a notch at 7 kHz. However, this notch disappears for higher elevations.

In the front, the measurements show a large peak around 12 kHz for the eye-level positions. In the back, the height of this peak decreases. This is consistent with the effect of pinna, which forms an acoustic shadow for the high-frequency sound sources located in the back. The amplitude difference in this frequency band, relative to the notches at lower frequencies, is a potential candidate for the front-back cue in median-plane sound localization (Iida *et al.*, 2007). The higher peak for the frontal eye-level position can also be observed in the simulation results. However, other local spatial-spectral features found in the measurements are missing in the simulation results.

The simulation results for the low-quality mesh are presented in the bottom-right panel of Fig. 8. The results show similar effects to that found for the horizontal plane. The spatial-spectral features appear smeared and show less details compared to the baseline condition. This supports the previous findings that a high-quality mesh seems to be crucial for the simulation.

The effects of temperature and perturbation are shown in Fig. 9. Temperature changes, which imply changes in the propagation time of the sound waves, led to frequency shifts of the amplitude spectra. Interestingly, the shifts are small compared to the substantially different shapes of the measured DTFs. Thus, the choice of the correct temperature for the simulations seems to be negligible as long as it is in a range of a typical room temperature. The perturbation had also a small effect on the simulation results. The most changes can be observer for frequencies above 10 kHz, which appear as frequency shifts in the order of few hundred hertz. The small effect of perturbation is surprising, given that the reduction in the mesh quality had a substantial effect on the simulation results. The perturbation can be seen as adding noise to the mesh. Nevertheless, it preserved most of the details in the mesh. In contrast, the low-quality mesh was a result of smoothing and node reduction, which obviously removed important details about the fine structure from the mesh. This fine structure seems to be important. Thus, a mesh precision of 0.5 mm seems to be sufficient, as long as all details in the range of 0.5 mm are well represented by the mesh.

A more detailed analysis of the differences between the simulation and measurements is provided in Fig. 10. It shows the amplitude spectra of simulated and measured DTFs for four sound source positions located in the median plane. The positions are −30°, 0°, 45°, and 180°. For frequencies below 7 kHz, all four positions show a good congruence of simulations with the measurements. The baseline condition with shoulders resulted in more spectral modulations compared to the baseline condition without shoulders. This confirms the importance of shoulders for low frequencies. For high frequencies, the differences between the conditions with and without shoulders are negligible.

Katz (2001b) reported simulation results for frequencies up to 6 kHz and for positions comparable to that in Fig. 10.[3] His largest difference between measurements and simulation was 17 dB (for frequency of 6 kHz at elevation of 45°). Our largest difference between measurements and the simulation is 5 dB (for 6 kHz at 0°). Allowing a maximal difference of 5 dB, his simulation results show congruence for frequencies up to 4.5 kHz, while our simulation results show congruence for frequencies up to 7 kHz. An explanation for our improvements may be the higher number of elements in the mesh we used. Our mesh had 70 785 elements, while the mesh of Katz had only 22 000 elements. This is also supported by our effect of the mesh quality: the lower number of elements in the

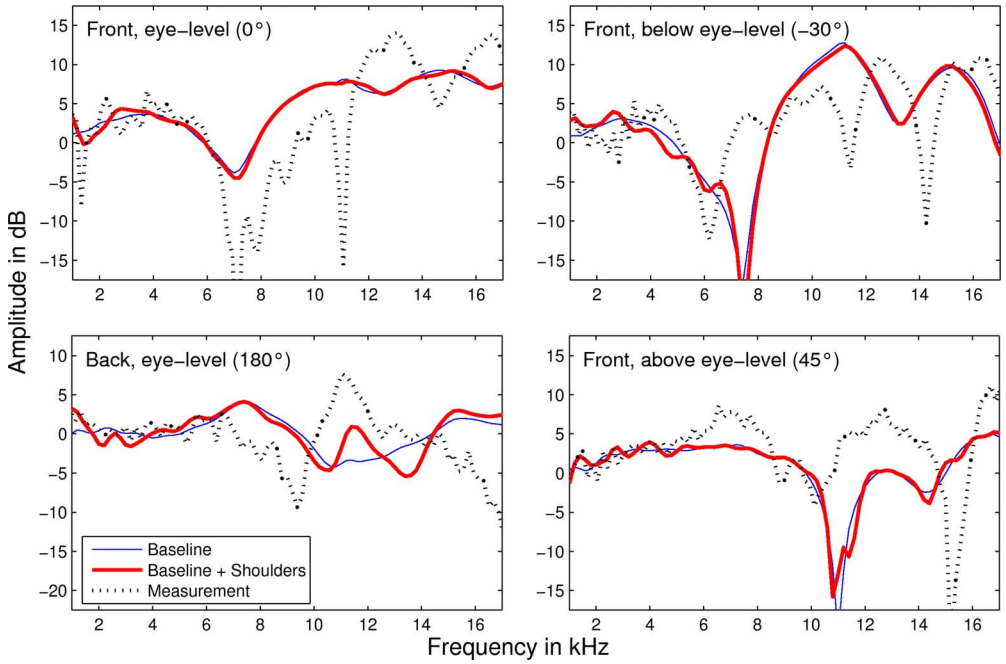


FIG. 10. (Color online) Amplitude spectra of simulated and measured DTFs for sound sources in the median plane. The elevations are −30° (front, below eye-level), 0° (front, eye-level), 45° (front, above eye-level), and 180° (back, eye-level). The solid lines show the simulation results for the baseline condition with and without shoulders. The dotted lines show the measured data.

low quality mesh resulted in a smearing of the spatial pinna details, which yielded a worse representation of DTFs' spectral features. Thus, a high-quality mesh seems to be essential for a good representation of spectral features in the simulation.

For frequencies above 7 kHz, the differences between the simulation and measurements are higher than for lower frequencies. The shapes of the simulated spectra still follow that of the measured spectra, as supported by a similar spectral tilt in both measurements and simulation. However, particular features like peaks and notches are not well represented in the simulations. For example, for the front above eye-level position (elevation angle of 45°), simulation results show a notch at 11 kHz, whereas measurements do not. However, the measurements show a notch at another frequency, namely, at 15 kHz. Unfortunately, the simulation results do not show the notch at this frequency. To address this issue, the sound pressure was analyzed in the surrounding of the receiver element. The left panel of Fig. 11 shows the sound pressure for the frequency of 11 kHz and sound source located in the median plane at the elevation angle of 45°. The position of the receiver element seems to be important because the pressure varies in the range of 20 dB within a few millimeters. If the receiver element does not represent the exact position of the microphone, then the propagation time of the reflections in the pinna is different than that in the measurements. Thus, moving the receiver element only a little could probably make the 11-kHz notch disappear. Thus, DTFs were calculated for a second receiver element in a distance of 1.2 mm from the first one. The right panel of Fig. 11 shows the difference between the amplitude spectra calculated for the two receiver elements. Local differences are evident for frequencies above 7 kHz. At 11 kHz, the dipole-like discontinuity at the elevation angle of 45°shows that the

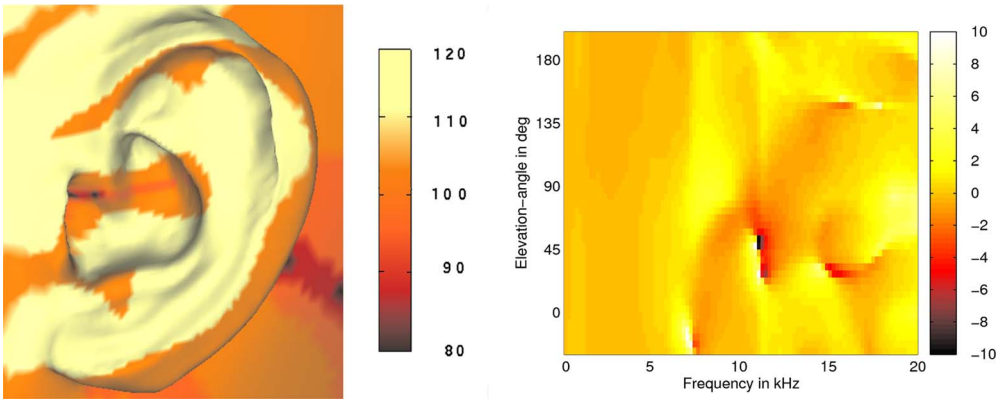


FIG. 11. (Color online) Left panel: Sound pressure at the pinna for a 11-kHz point source located in front at the elevation angle of 45°. Right panel: Difference between the amplitude spectra of DTFs calculated for two different receiver elements (see text). The color represents the magnitude in dB. Differences larger than 10 dB and smaller than −10 dB are shown in white and black, respectively.

11-kHz notch moved toward lower frequencies. This indicates that by moving the receiver element by just few millimeters some local high-frequency features of the DTFs substantially change. This may explain the good congruence of the global spectral shape and the poor congruence of the local spectral features in the DTFs. Thus, the accurate position of the receiver element seems to be crucial in the simulation of HRTFs.

The choice of the receiver element would have been easier when the pressure distribution were more homogeneous along the different elements. Preliminary simulation results for a mesh with a modeled ear canal showed that the pressure distribution inside the canal seems to be more homogeneous than outside of the canal. A systematic investigation of the role of the ear canal in HRTF simulations may allow to more easily choose the appropriate receiver element.

However, there are also other issues, which may be responsible for the differences between measurements and simulation. First, there are procedural differences between the visual scans and the acoustic measurements. For example, the visual scans were performed without the in-ear microphones, which probably complicated the choice of the receiver element. Also, the visual scans were performed with the rubber cap while the acoustic measurements were not. These procedural differences might have caused spectral differences in the results. Second, the mesh generation was a long-winded process. The mesh was stitched from six mesh parts; each mesh part was generated from a different perspective. The stitching process may have been inaccurate, leading to overlaps and shifts of the elements, and thus affecting simulation accuracy. As the mesh accuracy is a crucial factor for the simulations, a more accurate representation of the pinna geometry may further improve the simulation results.

V. CONCLUSIONS

In this study, a method for the calculation of HRTFs from visual scans is presented. The simulation is based on the BEM, which was coupled with the multilevel FMM in order to allow simulations for a wide frequency range. The upper frequency limit of this method does not depend on the computational limitations of modern computer systems but it depends only on the mesh size. The mesh from this study allows to calculate HRTFs for frequencies up to 35 kHz.

Comparison between the measured and simulated DTFs was performed for frequencies up to 16 kHz. It showed a good congruence of the spatial-spectral features for frequencies up to 7 kHz. For frequencies above 7 kHz, spectral shapes were in agreement with the measured data; however, local spatial-spectral features like peaks and notches were poorly represented by the simulation. Subsequent behavioral sound localization tests are required to show the actual localization ability using the simulated HRTFs.

An additional simple model of shoulders was included to the head model. Including shoulders to the simulation improved the representation of the elevation-dependent reflections for low frequencies and should be considered in further studies. The simulation temperature had a minor effect on the results. Also, the effect of mesh perturbation was very small, showing that the precision of visual scans is not crucial as long as the spatial pinna features are well represented. However, the reduction in the mesh size by only 5% had a substantial effect on the results, showing the importance of using high-quality meshes with large number of elements in HRTF simulations.

The differences between measurements and simulations may have several origins like procedural differences in the visual and acoustic data acquisition, imperfect representation of the pinna's geometry, and mismatch in the choice of the receiver element. To address these issues, improvements in the procedures and the numerical model are required. Fast acquisition of the geometrical data and easy mesh modifications make further research on the approach worthwhile.

ACKNOWLEDGMENTS

We would like to thank Michael Hofer from the TU Vienna for providing 3D scans and meshes. We are grateful to Bernhard Laback and Holger Waubke for helping comments. Also we would like to thank Alexander Haider for his support. This work was funded by the Austrian Science Fund (Project No. P18401-B15) and the Austrian Academy of Sciences.

[1]Recently, a detailed introduction to FMM-coupled BEM has been provided in Gumerov and Duraiswami, 2009.

[2]The Burton–Miller approach was preferred over the CHIEF-point method because the selection of appropriate CHIEF points is not trivial (Schenck, 1968; Ciskowski and Brebbia, 1991).

[3]Katz (2001b) reported data for the elevation of −45°. This elevation is outside of our measured range and thus, in our study, results for elevation of −30° are provided.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001a**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001b**). "The CIPIC HRTF database," in Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, New Paltz, NY.

Begault, D. R., Wenzel, E. M., and Anderson, M. R. (**2001**). "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc. **49**, 904–916.

Blauert, J. (**1974**). *Räumliches Hören (Spatial Hearing)* (S. Hirzel-Verlag, Stuttgart).

Bronkhorst, A. W. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.

Burton, A. J., and Miller, G. F. (**1971**). "The application of integral equation methods to the solution of exterior boundary-value problems," Proc. R. Soc. London, Ser. A **323**, 201–210.

Carlile, S., and Pralong, D. (**1994**). "The location-dependent nature of perceptually salient features of the human head-related transfer functions," J. Acoust. Soc. Am. **95**, 3445–3459.

Chen, Z.-S., Waubke, H., and Kreuzer, W. (**2008**). "A formulation of the fast multipole boundary element method (FMBEM) for acoustic radiation and scattering from three-dimensional structures," J. Comput. Acoust. **16**, 1–18.

Ciskowski, R. D., and Brebbia, C. A. (**1991**). *Boundary Element Methods in Acoustics* (Springer, Heidelberg).

Darve, E. (**2000**). "The fast multipole method I: Error analysis and asymptotic complexity," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **38**, 98–128.

Erichsen, S., and Sauter, S. (**1998**). "Efficient automatic quadrature in 3D Galerkin BEM," Comput. Methods Appl. Mech. Eng. **157**, 215–224.

Fischer, M., and Gaul, L. (**2005**). "Application of the fast multipole BEM

for structural-acoustic simulations," J. Comput. Acoust. **13**, 97–98.
Fischer, M., Gauger, U., and Gaul, L. (**2004**). "A multipole Galerkin boundary element method for acoustics," Eng. Anal. Boundary Elem. **28**, 155–162.
Giebermann, K. (**1997**). "Schnelle summationsverfahren zur numerischen lösung von integralgleichungen für streuprobleme im R^3 (Fast algorithms to numerically calculate the integral equation for the scattering problem in R^3)," Ph.D. thesis, Universität Karlsruhe.
Greengard, L., and Rokhlin, V. (**1987**). "A fast algorithm for particle simulations," J. Comput. Phys. **73**, 325–348.
Greengard, L., Huang, J., Rokhlin, V., and Wandzura, S. (**1998**). "Accelerating fast multipole methods for the Helmholtz equation at low frequencies," IEEE Comput. Sci. Eng. **5**, 32–38.
Gumerov, N. A., and Duraiswami, R. (**2009**). "A broadband fast multipole accelerated boundary element method for the three dimensional Helmholtz equation," J. Acoust. Soc. Am. **125**, 191–205.
Iida, K., Motokuni, I., Itagaki, A., and Morimoto, M. (**2007**). "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," Appl. Acoust. **68**, 835–850.
Kahana, Y., and Nelson, P. A. (**2007**). "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," J. Sound Vib. **300**, 552–579.
Katz, B. F. G. (**2001a**). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448.
Katz, B. F. G. (**2001b**). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects," J. Acoust. Soc. Am. **110**, 2449–2455.
Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.
Majdak, P., Balazs, P., and Laback, B. (**2007**). "Multiple exponential sweep method for fast measurement of head-related transfer functions," J. Audio Eng. Soc. **55**, 623–637.
Makous, J. C., and Middlebrooks, J. C. (**1990**). "Two-dimensional sound localization by human listeners," J. Acoust. Soc. Am. **87**, 2188–2200.
Marburg, S. (**2002**). "Six boundary elements per wavelength. Is that enough?," J. Comput. Acoust. **10**, 25–51.
Middlebrooks, J. C. (**1997**). *Spectral Shape Cues for Sound Localization* (Lawrence Erlbaum Associates, Mahwah, NJ).
Middlebrooks, J. C. (**1999**). "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**, 1480–1492.
Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.
Minnaar, P., Plogsties, J., and Christensen, F. (**2005**). "Directional resolution of head-related transfer functions required in binaural synthesis," J. Audio Eng. Soc. **53**, 919–929.
Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.
Rahola, J. (**1996**). "Diagonal forms of the translation operators in the fast multipole algorithm for scattering problems," BIT **36**, 333–358.
Roosendaal, T., and Selleri, S. (**2004**). *The Official Blender 2.3 Guide: Free 3D Creation Suite for Modeling, Animation, and Rendering* (No Starch, San Francisco).
Schenck, H. A. (**1968**). "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58.
Shaw, E. A. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.
Wightman, F. L., and Kistler, D. J. (**1989**). "Headphone simulation of free-field listening, I: Stimulus synthesis," J. Acoust. Soc. Am. **85**, 858–867.
Zotkin, D. N., Duraiswami, R., Grassi, E., and Gumerov, N. A. (**2006**). "Fast head-related transfer function measurement via reciprocity," J. Acoust. Soc. Am. **120**, 2202–2215.

Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses

James M. Harte,[a] Gilles Pigasse, and Torsten Dau
Department of Electrical Engineering, Centre for Applied Hearing Research, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

(Received 21 November 2008; revised 10 June 2009; accepted 14 June 2009)

Different attempts have been made to directly measure frequency specific basilar membrane (BM) delays in animals, e.g., laser velocimetry of BM vibrations and auditory nerve fiber recordings. The present study uses otoacoustic emissions (OAEs) and auditory brainstem responses (ABRs) to estimate BM delay non-invasively in normal-hearing humans. Tone bursts at nine frequencies from 0.5 to 8 kHz served as stimuli, with care taken to quantify possible bias due to the use of tone bursts with different rise times. BM delays are estimated from the ABR latency estimates by subtracting the neural and synaptic delays. This allows a comparison between individual OAE and BM delays over a large frequency range in the same subjects, and offers support to the theory that OAEs are reflected from a tonotopic place and carried back to the cochlear base via a reverse traveling wave. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3168508]

PACS number(s): 43.64.Jb, 43.64.Ri, 43.64.Kc [BLM] Pages: 1291–1301

I. INTRODUCTION

The human auditory system is known to possess high frequency selectivity and to compress a wide range of sound levels into an audible range. These characteristics can be explained by the mechanical properties of the inner ear. The tonotopic organization of the cochlea is such that highest frequencies are processed at the base of the basilar membrane (BM) and the lowest frequencies are processed at the apex. This is due to a stiffness gradient along the BM. The inward traveling wave has a longer path to travel to reach the low-frequency region of the cochlea compared to the high-frequency region (von Békésy, 1960). There is an intrinsic relation between frequency and travel time in the cochlea, defined as the cochlear delay, $\tau_{BM}(f)$. Different attempts have been made to directly measure BM vibration, e.g., by measuring the reflections of a laser Doppler velocimeter on the BM of animals and human cadavers (e.g., Recio *et al.*, 1998; Goodman *et al.*, 2004). Opening the cochlea, however, decreases its sensitivity at low frequencies (Dong and Cooper, 2006), and this only provides an approximation of the normally functioning human cochlea. Various measurement techniques have been developed to non-invasively and indirectly record BM delay such as otoacoustic emissions (OAEs) (Norton and Neely, 1987; Şerbetçioğlu and Parker, 1999) and auditory brainstem responses (ABRs); (Gorga *et al.*, 1988; Murray *et al.*, 1998). Comparing cochlear delay estimated from both these measures, in the same subject group, can provide valuable knowledge about the physical generation mechanisms of both. Such experimental data are, so far, limited in normal-hearing listeners, only extending across the mid-frequency region of the cochlea.

[a]Author to whom correspondence should be addressed. Electronic mail: jha@elektro.dtu.dk

A. OAEs

OAEs are low-level signals originating in the cochlea, traveling through the middle ear and recorded in the ear canal (EC). This epiphenomenon of a normally functioning auditory system was first recorded in humans by Kemp (1978). Classically, OAEs have been classified in terms of the stimulus with which they are evoked, i.e., tones, tonal complexes, transients, or no stimulus at all. When transient signals like clicks or tone-bursts (TBs) are used, the recorded OAEs are called transient evoked OAEs (TEOAEs). Studies have also shown that OAEs have a spectrum similar to that of the evoking tone burst (Wit and Ritsma, 1980; Kemp *et al.*, 1986). However, TEOAE spectra are often dominated by a few resonant peaks or dominant frequencies, often associated with measurable spontaneous OAEs (Jedrzejczak *et al.*, 2008). If present, these may confound estimates of latency for the OAE recordings. Previous studies have investigated the latencies of TEOAEs in the time domain (Norton and Neely, 1987; Şerbetçioğlu and Parker, 1999; Kapadia and Lutman, 2000; Hoth and Weber, 2001; Goodman *et al.*, 2004; Thornton *et al.*, 2006) and in the time-frequency domain (Elberling *et al.*, 1985; Probst *et al.*, 1986; Tognola *et al.*, 1997; Lucertini *et al.*, 2002; Sisto and Moleti, 2002; Jedrzejczak *et al.*, 2005). Unlike clicks which are broadband stimuli, TBs offer the advantage of being limited in frequency and can therefore be used to investigate limited regions of the cochlea. However, they are therefore less precisely defined in time making latency estimation more difficult. A number of studies have investigated TB evoked OAE (TBOAE) latencies in the time domain, using TBs over a limited frequency range (Wilson, 1980a; Grandori, 1985; Şerbetçioğlu and Parker, 1999; Norton and Neely, 1987).

Kalluri and Shera (2007) demonstrated the near equivalence of TEOAEs and stimulus frequency OAEs (SFOAEs), when the stimulus intensity is given in a bandwidth-compensated sound pressure level (SPL). The most accepted

hypothesis for the generation of SFOAEs at low to moderate excitation levels is the coherent reflection filtering (CRF) theory (Zweig and Shera, 1995). The theory states that randomly distributed inhomogeneities in the BM local impedance reflect microscopic wavelets from the forward traveling wave (Shera and Guinan, 1999; Kalluri and Shera, 2007). These wavelets sum up in-phase around the peak of the traveling wave and form a so-called retrograde or backwards traveling wave.

B. ABRs

Cochlear activity and therefore the cochlear delay, τ_{BM}, is also reflected at stages higher than the cochlea in the human auditory pathway. Auditory evoked potentials can be used to obtain an indirect estimate of cochlear delay in humans (Neely *et al.*, 1988). Specifically, ABRs are generated above the cochlea and are the result of the simultaneous activation of nerve cells in the brainstem (Møller, 1994). A typical ABR is made up of a series of so-called waves, whose latency is linked with the distance between the cochlea and the source of this wave. Waves I, III, and V are typically the most pronounced waves. The exact physiological sources of these waves are not fully known in humans. However, it is generally accepted that wave I stems from the distal portion of the afferent cochleo-vestibular nerve (VIIIth nerve), as demonstrated by Jewett and Williston (1971) and Møller and Jannetta (1983). It is assumed that wave III arises from an area near or inside the cochlear nucleus, situated at the bottom of the brainstem, and that wave V is attributed to neuronal activities between the lateral lemniscus and the inferior colliculus, contralateral to the side of the stimulus (Møller, 1994). Wave V is the wave with the largest amplitude and hence the most easily detectable. In order to obtain an estimate, $\hat{\tau}_{BM}^{(ABR)}$, of the cochlear delay, all of the component latencies must be considered. Wave-V latency is often considered to be composed of the sum of the synaptic delay, $\tau_{synaptic}$, the neural delay, τ_{neural}, as well as the cochlear delay τ_{BM} (Neely *et al.*, 1988). Thus, an estimate of the cochlear delay can be given by

$$\hat{\tau}_{BM}^{(ABR)} = \tau_{wave\ V} - \tau_{neural} - \tau_{synaptic}. \quad (1)$$

The synaptic delay is the time to be estimated between the inner hair-cells activity and the auditory-nerve fibers firing. It is typically around 1 ms (Kiang, 1975; Kim and Molnar, 1979; Møller and Jannetta, 1983; Burkard and Secor, 2002) and assumed to be independent of frequency and level (Don *et al.*, 1998). The neural conduction time (neural delay) is the time between the auditory-nerve activity and the place generating the ABR wave. It can be estimated from the interpeak delays of the ABR, i.e., by the latency difference of wave I and wave V: $\tau_{neural}=\Delta_{I\text{-}V}$. However, unlike wave V, the detection of wave I is rather difficult due to its small amplitude. It was therefore decided in this study to detect wave III instead and use the assumption $\Delta_{I\text{-}V}=2\Delta_{III\text{-}V}$ (Don and Eggermont, 1978; Eggermont and Don, 1980; Don and Kwong, 2002). The cochlear delay estimate, $\hat{\tau}_{BM}^{(ABR)}$, can then be calculated for each subject, using individual estimates of the wave-V latency and the interpeak delay:

$$\hat{\tau}_{BM}^{(ABR)} = \tau_{wave\ V} - 2\Delta_{III-V} - 1 \quad \text{ms}. \quad (2)$$

When measuring cochlear delays, the propagation time in the outer- and middle-ear is typically neglected because it is very short and not frequency dependent, unlike the cochlear delay. This assumption will be revisited in Sec. IV.

Historical experimental studies have demonstrated that τ_{BM} decreases as a power law of frequency. Neely *et al.* (1988) used the formula $\tau_{BM} \propto bc^{-i}f^{-d}$ to describe the latency of ABR as a function of frequency normalized to 1 kHz, f in kHz, with i representing the tone burst intensity in dB SPL divided by 100, and $b=12.9$ ms, $c=5.0$ ms, and $d=0.413$ representing fitting constants.

C. Goals and scope of the study

The main objectives of this study were to estimate the cochlear delay over a broader frequency range than historically carried out, and to determine the most appropriate method to do so. OAEs and ABRs were measured for the same subjects, using the same stimuli in the same laboratory, in an attempt to fill the dearth of experimental data pointed out by Moleti and Sisto (2008). In the present study, TBs ranging from 0.5 to 8 kHz were used and the implicit assumptions required to convert ABR latencies to cochlear delay are discussed. This study also compares OAE and ABR latency accumulation and tests the consistency with current theories about the OAE generation mechanisms. The CRF theory for OAE generation allows prediction upon the build up of delay in the final OAE recorded in the EC. At its simplest, the CRF theory suggests that the OAE delay is twice the delay between the stimulus onset and the peak of the (forward) traveling wave (Zweig and Shera, 1995). From this simple prediction, some debate has appeared in the literature to challenge the CRF theory, on the basis that the observed delay is too short. This will be discussed in the light of the experimental results from the present study. However, it should be noted that in this paper physical transmission delay is measured and not the relation between phase and frequency predicted by CRF theory. If the relation $\tau_{OAE}=2\tau_{BM}$ is obtained, then this implies that the OAE is generated at the tonotopic place, and propagates backwards to the base as a reverse traveling wave, with the same speed as that of the forward path. If this relation is not found, then this would imply that either the backward transmission is faster, as recently suggested by Ren (2004), or the backward wave is generated at a more basal cochlear place.

II. METHODS

A. Subjects

The subjects participating in this experiment were 11 normal-hearing adults: 2 females, 9 males, aged between 22 and 30 years. The different numbers of male and female subjects might bias the result. However, the present study compares OAE and ABR latencies on an individual basis. The higher number of male subject might lead to longer latencies compared to other studies (Don *et al.*, 1993). All subjects had pure-tone thresholds better than 15 dB HL in the range 0.25–8 kHz.

TABLE I. TB stimuli used, with length in ms and number of cycles.

Frequency (kHz)	Total length	
	ms	cycles
0.5	10	5
0.75	7	5.25
1	5	5
1.5	5	7.5
2	5	10
3	3.4	10.2
4	2.5	10
6	1.7	10.2
8	1.25	10

B. Apparatus and stimuli

1. OAEs

For the OAE recordings, the subjects were seated in a comfortable chair in an IEC 268-13 compliant sound insulated booth. Each recording session lasted around 45 min and the responses were recorded during three different sessions. The stimuli used for the OAE experiment were clicks and TBs repeated at a rate of 25/s. Clicks were approximately 113 μs long and were presented 4000 times at 56 and 66 dB pe (peak equivalent) SPL. TBs at 0.5, 0.75, 1, 2, 3, 4, 6, and 8 kHz were presented 4000 times at 66 dB pe SPL, chosen to be comparable with historical studies. A lower level would compromise the comparison with ABR, whose waves are difficult to detect. At higher levels, the cochlear amplifier is not as active and the relative size of the OAEs is therefore lower (Kemp, 2002). It should also be noted that 66 dB pe SPL is relatively high so there will be some spread of excitation. Therefore, the OAE and ABR may originate from a larger area on the BM. However, this was a necessary trade-off in the experimental design. TB frequencies ranging from 0.5 to 8 kHz were used and their durations ranged between 10 and 1.25 ms (see Table I). The experiment was repeated three times for each subject to get a measure of intra-subject variability.

2. Brainstem responses

For the TB evoked ABR (TBABR) recordings, the subjects were laid down on a clinical couch in the same booth as for the OAE recordings. Subjects were encouraged to sleep during the experiment to reduce electrophysiological background noise. The ABR recordings were collected during three sessions, lasting about 2 h each. The same TB stimuli as in the OAE experiment were used in the ABR experiment. However, the number of averages varied with center frequencies for ABR. A preliminary experiment showed that wave V could be detected using fewer averages for high frequencies, due to the stronger signal strength at these frequencies. The lower frequencies (0.5, 0.75, and 1 kHz) were repeated 8000 times, the middle frequencies (1.5, 2, and 3 kHz) 4000 times, and the higher frequencies (4, 6, and 8 kHz) 3000 times for each run.

The choice of the stimuli was inspired by the experiments from Norton and Neely (1987) and Şerbetçioğlu and Parker (1999). They were generated following the standard IEC 645-3 on short-duration test signals. These durations represent a trade-off between having an equal number of cycles for all frequencies and a relative narrow spread in their spectrum. The organization of frequency along the cochlear partition is roughly logarithmic and TBs with a fixed number of cycles result in uniform energy splatter in log-frequency. The stimulus rise time is responsible for the simultaneous neural activation leading to the brainstem responses (Suzuki and Horiuchi, 1981) and to obtain a detectable ABR. A sharp stimulus onset (i.e., a short rise time) produces a large amount of synchronized neural activity, but also decreases the frequency specificity of the stimulus. Rise times for frequencies of 2 kHz and above include approximately 5 cycles and therefore ranged from 2.5 to 1.25 ms. Below 2 kHz it was felt that the reduced energy spread, by keeping a fixed number of cycles, would make it almost impossible to record a wave-V response. Therefore, a compromise was struck, similar to Gorga *et al.* (1988), between the need for rapid stimulus onsets and reduced energy spread in the choice of rise time. The rise times were reduced to 3.25 at 1.5 kHz and approximately 2.5 in the 0.5–1.0 kHz range.

C. Procedure

The stimuli were produced in MATLAB, and then sent to a D/A converter (RME ADI8-Pro). The analog signal was transmitted to a programmable attenuator (TDT PA5) and a headphone driver (TDT HB7) and finally to the insert earphone ER-2. The stimuli were calibrated using an ear mould simulator (B&K DB 0370) connected to an IEC 711 coupler (B&K 4157) and a B&K 2607 sound level meter.

1. OAEs

The TBOAEs were recorded in the right ear for most of the subjects but in the case of a blocked EC, the left ear was chosen (two subjects only). Studies have shown that the two ears of a subject presented very similar TEOAEs (Jedrzejczak *et al.*, 2005; Moleti *et al.*, 2008); therefore, the choice of ear should have no influence. The TBOAEs were recorded with an ER-10B low-noise microphone (Etymotic Research), band-pass filtered 0.15–16 kHz with an analog filter (Krohn-Hite 3750), digitalized with the RME A/D converter and, finally, stored on a PC for off-line analysis. Both stimulus generation and response measurement were controlled via MATLAB.

2. Brainstem responses

For the TBABR, the responses were recorded using the Synamps I EEG amplifier and the data saved for off-line analysis. Despite attempts to shield the transducers, some electrical artifact signal due to the stimulus could be recorded in the response. However, an alternating polarity procedure (Gorga *et al.*, 1988) was used to reduce this. It has previously been shown that this also leads to higher amplitude of the wave V, making its detection easier (Foxe and Stapells, 1993; Schönweiler *et al.*, 2005). A repetition rate of 24.5 Hz was used. Studies have shown that the amplitude of

TABLE II. Table of individual mean EC impulse response lengths and standard deviations.

Subject	3	4	5	6	7	9	10	11	12	15	16
$\mu_{\tau_{EC}}$ (ms)	3.1	4.2	4.1	3.6	2.2	2.9	4.3	4.3	2.1	4.0	3.9
$\sigma_{\tau_{EC}}$ (ms)	0.1	0.1	0.2	0.3	1.6	0.4	0.2	0.2	0.2	0.3	0.1

the responses remains unaffected by repetition rates up to 35/s (Stapells and Picton, 1981) and the latency is stable for a large range of repetition rates (Burkard and Secor, 2002). The ground electrode was placed on the subjects' forehead, the reference electrode was placed at the vertex [Cz in the 10–20 electrode system, Jasper (1958)], and the remaining electrode was placed at the ipsilateral mastoid. The impedance of each electrode was maintained below 5 kΩ. The order of presentation of the stimuli was randomized for each subject.

D. Off-line data analysis

1. OAE artefact rejection

In order to reject epochs containing a great amount of noise due to subject movement or swallowing, artifact rejection was applied. Averages contaminated by external noise can be detected due to the presence of sudden high amplitude signals, the averages were ranked according to their maximum amplitude, and 10% of them were discarded.

2. OAE detection method

The detection of the OAE onset is difficult due to early components in the recorded pressure attributable to the transducer response and the EC (Stover and Norton, 1993). In previous studies, the separation between linear and nonlinear reflections has been identical for all subjects (Kemp, 1978; Norton and Neely, 1987; Keefe, 1998; Şerbetçigğlu and Parker, 1999; Jedrzejczak *et al.*, 2005), by assuming that all have identical idealized ECs and middle ears. The paradigm used in the present study tries to determine the OAE onset in each subject separately. This method is based on the separation, in time, between the reflections occurring in the EC and the signal originating in the cochlea. This separation is made, on the one hand, by calculating the impulse response of the subject's EC and, on the other hand, by detecting the peak attributed to the OAE.

In order to calculate the impulse response of the subject's EC, two clicks are presented at two levels, 56 and 66 dB pe SPL, both sufficiently high to be in the compressive regions of the OAE input-output function. These transient responses can be considered as being composed of two components; the first is the EC and transducer response linearly scaled with input level; the second is the OAE compressively scaled with input level. The recorded transient responses are then normalized so that the peak absolute amplitude is set to unity. By normalizing the responses, the linear EC component in each should be identical within the noise floor of the recording. The OAE components on the other hand will differ due to the compressive input-output function. After normalization, the OAE component for the 66 dB pe SPL input level will be smaller relative to the OAE component for the 56 dB pe SPL excitation level. Thus, the region where these two transient response curves diverge indicates where the OAE dominates the time series. The EC impulse response estimate, $\tilde{h}_{EC}(\tau)$, is thus defined as the normalized transient response up to the point the OAE response dominates, obtained via visual inspection for each subject. Table II gives the mean (across the three repeat runs) and standard deviation of the $\tilde{h}_{EC}(\tau)$ durations for each subject.

Similarly, for the TB stimuli, it is assumed that the response recorded is composed of EC and OAE components, $p(t)=p_{EC}(t)+p_{OAE}(t)$. However, the low-level tail end of the $p_{EC}(t)$ component overlaps with the start of the $p_{OAE}(t)$. This OAE onset ambiguity is worse for TB stimuli as they are less well defined in time than a delta function or impulse. In order to obtain an estimate for the EC component, the estimated impulse response obtained from the transient analysis can be used and convolved with the TB stimuli to yield

$$\tilde{p}_{EC}(t) = \int \tilde{h}_{EC}(\tau) x_{stim}(t-\tau) d\tau, \tag{3}$$

where $\tilde{p}_{EC}(t)$ is the estimated EC pressure component due to the TB stimulus, $x_{stim}(t)$. Thus, to obtain the OAE latency a comparison is made between the stimulus onset and the peak of the OAE response to the TB, using the two time series, the recorded pressure due to the TB stimuli, $p(t)$, and the estimated pressure component due to the EC and transducers, $\tilde{p}_{EC}(t)$. It is beneficial to restrict the analysis to a narrow frequency range around the TB center frequency, to remove any ambiguous components not directly evoked by the stimulus. One approach to narrowband analysis of the signals is via non-parametric time-frequency analysis, i.e., a simple linear spectrogram or a quadratic time-frequency representation such as the Wigner–Ville or related distributions (Cheng, 1995; Hatzopoulos *et al.*, 2000; Konrad-Martin and Keefe 2003) or even a wavelet based method (Wit *et al.*, 1994; Tognola *et al.*, 1997; Moleti *et al.*, 2005; Sisto and Moleti, 2007). These methods each have their advantages and disadvantages which will not be discussed here. In the present study, a parametric method of signal analysis was adopted from Long and Talmadge (1997), where an underlying model of the data is assumed. A least-squares fit was made to the following signal model:

$$y = a\cos((\omega+\Delta\omega)t) + b\sin((\omega+\Delta\omega)t), \tag{4}$$

where ω is the expected frequency of the signal (TB center frequency), and $\Delta\omega$ is a small offset around ω (where typically $\Delta\omega \ll 1$), allowing detection despite small changes in frequency. This method uses a least-squares fitting procedure to obtain the unknown parameters a, b, and $\Delta\omega$, and allows the calculations of the instantaneous amplitude, $A(t) = \sqrt{a^2(t)+b^2(t)}$, or envelope of the signal (see Long and Talmadge, 1997).

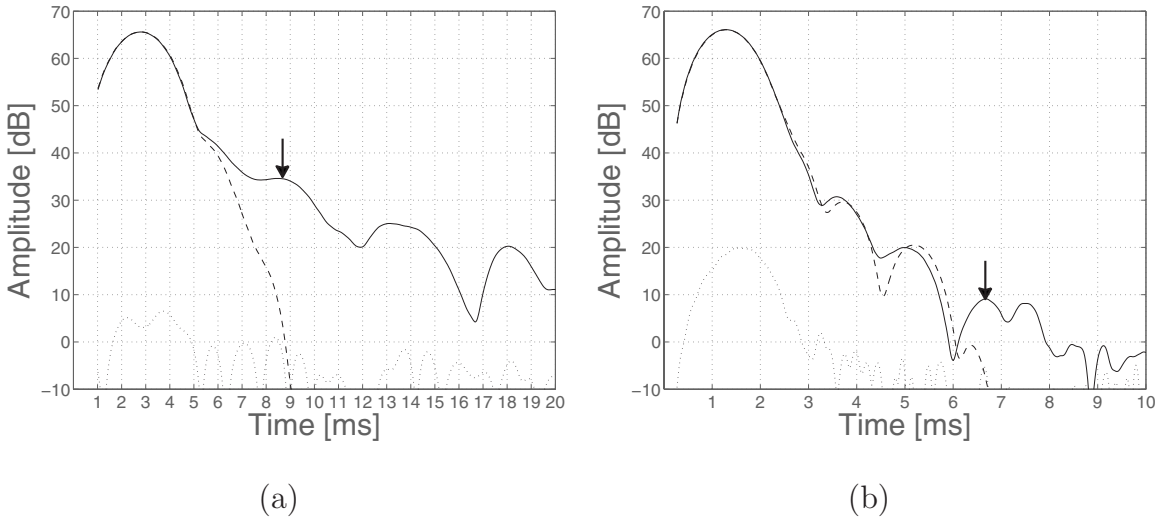


FIG. 1. Envelopes of the estimated EC and transducer signals (dashed curve), noise residue (dotted curve), and of the OAE response (solid curve) for (a) a 1 kHz TB and (b) a 4 kHz TB. The identified first peak attributed to the TBOAE is indicated by the arrow. Both these examples are from the same illustrative subject.

The least mean square fitting procedure therefore obtains the recorded signals envelope variation centered around the frequency of interest. Both the recorded pressure, $p(t)$, and the EC and transducer estimate, $\tilde{p}_{\mathrm{EC}}(t)$, are processed through this algorithm, thus allowing detection of the first peak in $p(t)$ after the EC and transducer component has died away. An example is shown in Fig. 1(a) for a 1 kHz TB and in Fig. 1(b) for a 4 kHz TB. The dashed curves indicate the estimated EC component and the solid curve the recorded pressure. An arrow indicates the location of the OAE peak used to define latency. In order to label the OAE onset, the burst of energy envelope was required to be 6 dB above both the noise floor and the estimated EC component. In Fig. 1(a), the OAE delay, τ_{OAE}, was found to be 8.7 ms where the EC component had sufficiently decayed away from the recorded TB response. The noise floor in this case was around 35 dB below the OAE. The 4 kHz delay in Fig. 1(b) was labeled as 6.6 ms. In this case, the OAE component was only 12 dB above the noise floor. Typically, for the higher frequency TB stimuli, the reduced SNR made labeling more difficult and occasionally impossible for some subjects.

There are a number of peaks and dips appearing in the recorded signal, each with different levels and latencies. The expected level of the OAEs is about 45 dB below the peak stimulus level (Wilson, 1980a; Wit and Ritsma, 1980). The OAE latency was defined, in the present study, as the time between the onset of the stimulus and the peak local maximum detected as an OAE, following Şerbetçioğlu and Parker (1999). The method used in the present study does not allow to detect the OAE burst with a 100% certainty. Although best efforts were made to provide a reliable tool, the OAE onset ambiguity is not fully solved. This ambiguity seems to be inherent to any TEOAE recordings.

3. ABR off-line analysis

The ABR recordings were first epoched and averaged using an iterative weighted-averaging algorithm (Riedel *et al.*, 2001). The responses were then filtered, between 0.1 and 1.5 kHz. Wave V peaks and latencies were determined assuming a sufficiently high signal to noise ratio and good repeatability across the three repeated runs.

III. RESULTS

A. Experimental data

The individual OAE delay, τ_{OAE}, and BM delay estimated from ABR, $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$, are shown in the left and right panel of Fig. 2, respectively. As expected, these two delays show exponentially decreasing delays as a function of frequency. The detection of wave V at low frequencies proved to be impossible for some subjects, despite the high number of averages (8000). This problem at low frequencies has been pointed out in many previous studies (e.g., Stapells, 1994; Stürzebecher *et al.*, 2006) and could be due to the lower speed of the traveling wave in the low-frequency region of the cochlea compared to the basal part. With a lower

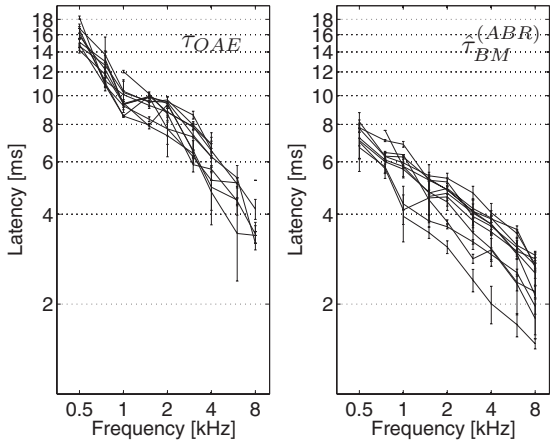


FIG. 2. The OAE and BM latency estimates for the 11 subjects are shown in the left and right panels, respectively. The OAE latency is defined as the time between stimulus onset and the peak of the OAE burst. The BM latency estimate, $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$, is calculated following Eq. (2). The error bars represent ± 1 std.

TABLE III. Inter-subject variability for τ_{OAE} and $\hat{\tau}_{BM}^{(ABR)}$ across frequency and given in percentage. The percentage is calculated as the ratio between the standard deviation and the mean value.

Freq. (kHz)	0.5	0.75	1	1.5	2	3	4	6	8
τ_{OAE} var. (%)	8.2	8.9	10.1	9.0	9.4	12.3	13.3	14.7	18.8
$\hat{\tau}_{BM}^{(ABR)}$ var. (%)	10.2	10.8	16.9	10.9	10.9	12.3	12.4	14.4	14.0

velocity, adjacent nerves fire with a certain delay, leading to asynchronous neural activity, and thus a lower amplitude of the ABR. An increase in the stimulus level would produce a more detectable wave V at 0.5 kHz, due to a greater discharge of the neurons (Gorga *et al.*, 1988). However, the stimulus level of the TBABR measurements (66 dB pe SPL) was chosen according to the OAE experiment for better comparability. A level higher than 66 dB pe SPL in the OAE experiment would have resulted in a relatively lower response and to a reduced frequency specificity due to broader filter activation (Ruggero *et al.*, 1997).

B. Data analysis

1. Intra- and inter-subject variabilities

For both measurement methods, the *intra*-subject variability of the latency is relatively small (error bars in Fig. 2). For OAEs (left panel), the maximum standard deviation is 2.01 ms for subject 10 at 0.75 kHz, which is rather small in comparison to the latency at that frequency (13.7 ms). However, this variability is only determined using three (at most) repeat measures and should therefore only be considered as a rough estimate. Some of the intra-subject variability could be due to the difficulty to assessing the true OAE onset, and not a component of the EC or middle ear response. Similarly, for ABR (right panel), the maximum standard deviation observed is 1.13 ms at 0.5 kHz for subject 3. The reproducibility of both these data sets is high and indicates that both techniques are reliable.

Figure 2 also demonstrates the inter-subject variability, which appears similar across frequency for both OAEs and ABRs. At 0.5 kHz, it is 1.29 and 0.72 ms for OAEs and ABR, respectively, and at 8 kHz it is 0.71 and 0.33 ms. It is necessary to consider the relative standard deviation with respect to the magnitude of the latency estimates at each frequency. The inter-subject variability, in percentage, is calculated as the ratio between the standard deviation and the mean value. Percentages are used in order to compare with the results from Norton and Neely (1987). The relative variabilities found in this study, listed in Table III, are of the same order as the ones found by Norton and Neely (1987). The relative variability of $\hat{\tau}_{BM}^{(ABR)}$ varies between 10.2% and 16.9%, with little systematic dependence on frequency. For τ_{OAE}, the range from 8.2% to 18.8% is slightly larger. However, the variability of τ_{OAE} is larger at higher frequencies probably due to the difficulties in correctly labeling the OAE onset. It is seen that the latency estimates differ slightly more between subjects for OAE than for ABR. On the cochlear level, subjects can indeed reflect different cochlear filtering properties which may variably affect the traveling wave. They may also have different hearing thresholds across the audible frequency range (Don *et al.*, 1994). These differences affect, however, both the OAE and ABR inter-subject variability. In the case of ABR, variabilities also occur at a neuronal level, due to different head sizes or gender (Hall *et al.*, 1988; Don *et al.*, 1993; Burkard and Secor, 2002).

2. Ratio of delay estimates

Figure 3(a) shows, the OAE delay, τ_{OAE}, and the derived BM delay from ABR estimates, $\hat{\tau}_{BM}^{(ABR)}$, plotted together for an illustrative subject. Figure 3(b) shows the same for the mean across subjects. The delay-frequency relations for both τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ can be modeled as a power law similar to Neely *et al.* (1988), where

$$\tau \propto \beta f^{-\alpha}. \tag{5}$$

Plotting this on a double-log axis will result in a straight line with slope $-\alpha$ and ordinate intercept β. The parameters $-\alpha$ and β were found for both τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ for each subject and for the across subject mean using an unconstrained nonlinear optimization routine in MATLAB. The model fitting parameters are given in Table IV. For comparison, equivalent of Neely *et al.* (1988) to β was 4.46, which is around half of the values obtained here. This is as expected as Neely *et al.* (1988) fitted this function to $\hat{\tau}_{BM}^{(ABR)}$ instead of $2\hat{\tau}_{BM}^{(ABR)}$ as used in the present study and β represents the ordinate intercept on a double-log plot. Multiplying $\hat{\tau}_{BM}^{(ABR)}$ by 2 will have no effect on the slope $-\alpha$, where Neely *et al.* (1988) found a value of 0.413 which is comparable to those values found here in Table IV. The solid and dashed lines in Fig. 3(a) show the model best fit to τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ for one illustrative subject; Fig. 3(b) shows the comparison for the group mean values. If τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ would align then we could argue that the OAE is generated at the tonotopic resonant

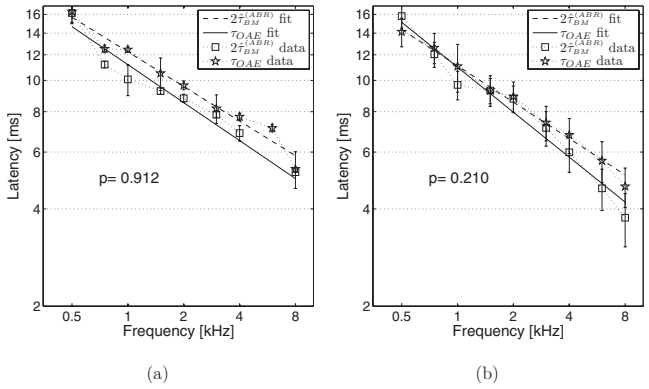


FIG. 3. Comparison between τ_{OAE} (solid curve) and $2\hat{\tau}_{BM}^{(ABR)}$ (dashed curve) for (a) subject 10 and (b) mean across subjects. The data points are also plotted (symbols ⋆ and □) and are connected by dotted lines. The solid and dashed lines represent the best fit to the data. The value of p resulting from the two-way ANOVA test is indicated. For both the illustrative subject and the mean across subjects $p<0.05$, it is therefore considered that OAE and ABR are significantly different.

TABLE IV. Delay power-law function fitting parameters for Eq. (5) across subject. Note the mean refers to the model fitted to the mean latency data and not the mean of the parameters shown.

Subject	$\alpha_{2\hat{\tau}_{BM}^{(ABR)}}$ (ms^2)	$\beta_{2\hat{\tau}_{BM}^{(ABR)}}$ (ms^2)	$\alpha_{\tau_{OAE}}$ (ms^2)	$\beta_{\tau_{OAE}}$
3	0.42	9.76	0.57	10.85
4	0.28	10.41	0.31	11.13
5	0.48	8.16	0.42	11.04
6	0.54	11.51	0.44	10.22
7	0.45	12.77	0.57	10.40
9	0.41	10.89	0.47	10.82
10	0.36	12.25	0.39	11.19
11	0.31	12.19	0.46	11.89
12	0.34	12.55	0.59	12.35
15	0.31	11.45	0.36	10.71
16	0.31	10.43	0.42	11.08
Mean	0.37	11.09	0.46	10.98

place and propagates backwards via a reverse traveling wave. If they are parallel to each other, this would support a factor different from 2 since, on a log-log axis, a multiplication factor just shifts the curves up and down. To consolidate this visual analysis, a two-way analysis of variance (ANOVA) was carried out. It examines the effect of independent factors on the BM latency estimate. The independent factors are the frequency ($n=9$) and the measurement technique ($n=2$, ABR or OAE). The null hypothesis is as follows: The estimate of τ_{BM} does not differ between techniques. Results are declared significant if the p-value is less than 0.05 and this would cast doubt on the null hypothesis.

The results of the ANOVA test (p-values) for one exemplary subject and the mean across subjects are also presented in Figs. 3(a) and 3(b), respectively. For the exemplary subject, the slopes of the OAE delay and two times the ABR estimated BM delay, $\hat{\tau}_{BM}^{(ABR)}$, do not differ significantly ($p=0.912$). In other words, the two techniques estimate the same rate of change of latency, i.e., the delay relation for τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ are statistically similar. Likewise, the slopes for the latencies averaged across subject [Fig. 3(b)] are the same, with the ANOVA test yielding a p-value of 0.210. The ANOVA test was run on the 11 subjects and the p-values given in Table V. For 8 of the 11 subjects, it was found that $p>0.05$, implying that the slopes of τ_{OAE} and $2\hat{\tau}_{BM}^{(ABR)}$ are not significantly different in most cases. This suggests that the OAE being generated by a tonotopically resonant place and propagating back via a reverse traveling wave would be supported by these results, since the data obtained in this study verify that $\tau_{OAE}\approx 2\hat{\tau}_{BM}^{(ABR)}$ for most of the subjects.

It has been suggested (Narayan, 1991) that the return trip for OAEs is not the same as the forward traveling wave, i.e., that there is no retrograde wave and that OAEs might travel faster on their way back. Therefore, the factor 2 between τ_{OAE} and $\hat{\tau}_{BM}^{(ABR)}$ might not be correct. Shera and Guinan (2003) found factors relating τ_{BM} to τ_{OAE} of 1.7 and 1.6 for cats and guinea pigs using SFOAEs, respectively, and Moleti and Sisto (2008) found a factor of 2.08 ± 0.19 in humans using TEOAEs. In the present study, this factor was calculated from all τ_{OAE} and $\hat{\tau}_{BM}^{(ABR)}$ pairs (a total of 81 points) and found to be 1.92 ± 0.42 ms. This standard deviation is relatively high, and is probably dependent on outliers due to incorrect identification of the OAE onset. Figure 4 shows a histogram for all of the possible calculated ratios across the whole frequency range and all subjects. The light gray bars could be considered outliers from a more Gaussian distribution of ratios (darker gray bars). If the ratio were recalculated only on the dark gray subset of results, then it is found to be 1.85 ± 0.21 ms. However, as mentioned earlier, the factor between τ_{OAE} and $\hat{\tau}_{BM}^{(ABR)}$ just shifts the curves and their slopes are not affected. The present analysis therefore remains unaffected by the exact constant of proportionality between these two delays.

TABLE V. Statistical analysis results. Column 2 gives the subject dependent interpeak and synaptic delay means ± 1 standard deviation, with the difference from that used by Moleti and Sisto (2008) given in parentheses. Column 3 gives the p-values from the ANOVA test with any significant results given in bold. The p-values in parentheses were obtained when using the fixed Moleti and Sisto (2008) interpeak and synaptic delay. Note the mean refers to the results for the mean latency data and not the mean of the parameters shown.

Subject	$2\cdot\Delta_{III-V}+\tau_{synaptic}$ ms ± 1 s.d. (ms)	ANOVA p-value
3	4.40 ± 0.12 (0.6)	0.168 (0.450)
4	4.16 ± 0.08 (0.84)	0.203 (0.661)
5	5.16 ± 0.24 (−0.16)	0.495 (0.358)
6	4.96 ± 0.08 (0.04)	0.971 (0.923)
7	4.56 ± 0.14 (0.44)	0.251 (0.575)
9	4.40 ± 0.18 (0.60)	0.947 (0.386)
10	4.64 ± 0.30 (0.36)	0.912 (0.633)
11	4.14 ± 0.22 (0.85)	**0.016** (0.194)
12	4.20 ± 0.18 (0.80)	0.286 (0.974)
15	4.36 ± 0.08 (0.65)	**0.018 (0.044)**
16	4.56 ± 0.26 (0.44)	**0.006 (0.022)**
Mean	4.52 ± 0.36 (0.49)	0.210 (0.640)

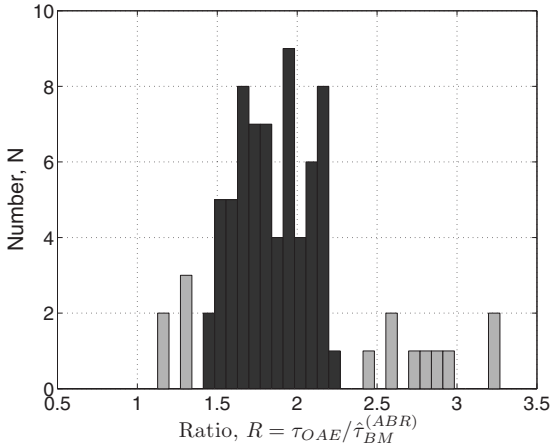


FIG. 4. Histogram of all calculated τ_{OAE} to $\hat{\tau}_{BM}^{(ABR)}$ ratios. Group shown in light gray represent potential outliers.

IV. DISCUSSION

A. Revisiting the assumptions for the delay estimates

1. Role of middle-ear delay

As discussed by Moleti and Sisto (2008) and Abdala and Keefe (2006), neglecting the middle-ear delay could lead to misinterpretation of the cochlear delay estimates based on OAE and ABR recordings. Puria (2003) made measurements of human middle ear forward and reverse acoustics, and discussed their implications for OAEs. In particular, he recorded the frequency dependent forward, backward, and round trip pressure gain and phase. This showed a significant bandpass type characteristic with frequency. Both delay estimates obtained in this study, τ_{OAE} and $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$ would be affected by the forward transmission. Puria (2003) showed this to be maximum around 0.9 kHz with a slope of −10.4 dB/oct below it and −7.2 dB/oct above it. This implies a difference in effective stimulus level at the stapes. Cochlear tuning is lower at higher stimulus levels, thus introducing an additional dependence on frequency for the stimuli used in this experiment. It is difficult to quantify this difference; however, it should be the same for both delay estimates τ_{OAE} and $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$. ABR forward latency of Neely *et al.* (1988) was well represented by the function, also used here, $\tau_{\mathrm{BM}} \propto \beta f^{-\alpha}$ across 0.25–8 kHz. Therefore, this power-law relation appears independent of middle-ear filtering.

Additional to the frequency-dependent delay associated with different excitation levels, Puria (2003) recorded the phase of the forward and reverse transmission paths, leading to an accumulated group delay. Within the range 2–6 kHz, the phase angle decreases with increasing frequency for the forward transmission path with a slope of −73°/oct, with the reverse transmission path having a slope of −105°/oct. If this is converted to group delay, this would imply approximately 0.08 ms in forward and 0.12 ms in the reverse directions. Thus, τ_{OAE} would have a delay of 0.2 ms attributed to middle-ear transmission (forward and reverse).

Multiplying the ABR delay estimate, $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$, by two would result in a middle-ear round trip delay of twice the forward delay, 0.16 ms. Thus, in the frequency region 2–6 kHz there is an additional transmission delay expected in the OAE over the ABR delay estimates due to the round trip through the middle ear of 0.04 ms. However, this is considered here to be small relative to the existing experimental error, and is therefore thought to be negligible.

2. Effects of neural and synaptic delays

In the present study, the cochlear delay estimate, $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$, was made using individual estimates of interpeak or neural delays, i.e., it was assumed that $\Delta_{\mathrm{I-V}} = 2 \cdot \Delta_{\mathrm{III-V}}$, as also used by Don and Eggermont (1978); Eggermont and Don (1980) as well as Don and Kwong (2002). The wave-III to wave-V delays were recorded in all subjects tested for frequencies in the range 2–8 kHz. Below 2 kHz it was felt that wave-III could not be consistently recorded. The averaged neural and synaptic delays for each subject, as well as a grand mean, are shown in Table V alongside the standard deviation. The synaptic delay was assumed here to be constant and 1 ms.

Another approach for dealing with the unknown neural delay is to assume a constant delay across subjects. For example, Don *et al.* (1993) used a fixed interpeak delay of 4.0 ms for males and 3.8 ms for females. Moleti and Sisto (2008) subtracted a constant offset of 4.2 ms for the neural delay and assumed a 0.8 ms synaptic delay. The grand average delay used in this study was 4.52 ms, which would under-predict the estimate obtained from 5.0 ms by 0.48 ms of Moleti and Sisto (2008). However, this offset or difference across subjects ranges from 0.86 ms under- to 0.16 ms over-prediction relative to Moleti and Sisto (2008)'s fixed delay. The results reported here, using a subject-dependent delay, would therefore tend to under-predict the estimate $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$ relative to method of Moleti and Sisto (2008).

The statistical analysis used in this study was based on comparing the evolution of delay with frequency. A constant offset of delay would thus change the slope of $\hat{\tau}_{\mathrm{BM}}^{(\mathrm{ABR})}$ when plotted on a double-log plot. This could potentially alter the conclusions, as the assumptions on Eq. (2) would more significantly affect the high-frequency delay estimates, due to the shorter delays in this region relative to the offset. Table V presents in round brackets the results of the ANOVA test applied when using the fixed neural delay from Moleti and Sisto (2008). It can be seen that 9 of the 11 subjects demonstrated no significant difference in their slopes. It is apparent that this ANOVA test is sensitive to the assumption of how to model neural delay. However, both the ABR and OAE methods tend to give the same evolution of latency with frequency.

B. Effect of TB rise time on wave-V latency

The TBs used to measure OAEs and ABRs were inspired by the experiment of Neely *et al.* (1988), who used different rise times across frequency in order to have the same width of BM excitation for each stimulus. The choice for different absolute rise times for different frequencies has recently been criticized by Ruggero and Temchin (2007) who claimed that this difference adds a supplementary delay to $\tau_{\mathrm{wave\ V}}$, which would yield a change in the latency slope function. They argued that identical TB rise times are necessary to have the synchronous neural firing occurring at the same time for all the stimuli. Neely *et al.* (1988) used TBs with onset-ramp durations that decreased as a function of increasing frequency (e.g., 4 ms for 0.25 and 0.5 kHz, 2 ms for 1–2 kHz, 1.4 ms for 3 kHz, 1 ms for 4 kHz, and 0.5 ms for 8 kHz). Ruggero and Temchin (2007) argued that this should artificially produce delays that increase as stimulus frequency decreases, based on the study on auditory nerve fiber first spike firing by Heil and Irvine (1997). Ruggero and Temchin (2007) cited Heil and Irvine's (1997) Figs. 2E and 3E, that for a 30 dB SPL tone, increasing the ramp duration from 1.7 to 4.2 ms can increase the first-spike latency by 1.5–3 ms. This calls into question some of the findings for Neely *et al.* (1988) at low excitation levels. However, citing again Heil and Irvine's (1997) Figs. 2E and 3E, this time for a 70 dB SPL tone, the first-spike latency increase is reduced to the order of 0.3 ms. Therefore, at the moderate excitation levels used in the present study, the role of TB rise time

should be minimal and below the range of experimental precision. It is thus argued here that this aspect does not alter the conclusions of the current study.

C. Implications for OAE generation mechanisms

A discussion on the OAE generation mechanisms is made as these are still under debate and the subject of controversy in the literature (Ren and Nuttall, 2006; Shera *et al.*, 2007; Ruggero and Temchin, 2007; Dong and Olson, 2008; Shera *et al.*, 2008). The present study suggests that OAE delay, τ_{OAE}, fits relatively well with the idea that OAEs are generated at the tonotopic site and propagate backwards to the base as a traveling wave. The approximate ratio of OAE delay, τ_{OAE}, to the BM delay estimated from ABRs, $\hat{\tau}_{BM}^{(ABR)}$, was found to be 1.92 ± 0.42 ms or 1.85 ± 0.21 ms if one could justify the removal of the outliers. This is also consistent with the approximate CRF theory prediction of a factor of 2. Moleti and Sisto (2008) found a factor of 2.08 ± 0.19 in humans; however, the study used TEOAE results compared with historical ABR recordings evoked from both TBs and derived band techniques. In the present studies, the same subjects were tested with identical stimulus types for both ABR and OAE recordings. The mean ratios were similar for both studies; however, the present study demonstrated a rather large variability of 0.42 ms. This is possibly due to the methods employed here to identify the OAE onset burst, where mislabeling would lead to large outliers.

The present results are in agreement with previous animal studies that showed reasonable agreement with the CRF prediction at high frequencies (Shera and Guinan, 2003; Siegel *et al.*, 2005). Shera and Guinan (2003) compared SFOAE group delays measured in cats and guinea pigs with BM mechanical transfer functions taken from the literature. They found that the relation $\tau_{OAE} \approx 2\tau_{BM}$ holds for the basal part of the cochlea only (i.e., above 4 kHz). Siegel *et al.* (2005) tested the hypothesis that SFOAE group delay was twice that of BM delay. They estimated the BM delay by using Wiener-kernel analysis of responses to noise of auditory nerve-fibers in chinchillas. Siegel *et al.* (2005) demonstrated that SFOAE group delays were slightly below that predicted with the simplified assumption of coherent reflection at frequencies above 4 kHz, i.e., $\tau_{SFOAE} \neq 2\tau_{BMgroup}$. However, they claimed that the data are indeed compatible with the hypothesis of SFOAE propagation to the stapes via fluid coupling (compression-wave) or via reverse BM traveling wave with speeds corresponding to the signal-front delays rather than group delays of forward waves. These results contradict the CRF prediction, but not the reflection mechanism itself. Shera *et al.* (2008) offered a different physical argument why the factor between τ_{OAE} and τ_{BM} is below 2 for frequencies above 4 kHz. They used a semi-analytic form (developed by Shera *et al.*, 2005) of the coherent reflection model, and tested model predictions tailored for chinchilla cochleae against recorded SFOAE delays. The model predicts SFOAE delays, $\hat{\tau}_{SFOAE}$, corresponding to the round trip delays for *pressure difference* waves, rather than BM velocity traveling waves (or alternatively "signal-front" delay). Revised prediction of Shera *et al.* (2005) for CRF theory is summarized as $\hat{\tau}_{SFOAE} \approx 2(1 - \hat{\tau}_{BM}/\hat{\tau}_k)$, where $\hat{\tau}_k$ is a positive delay, empirically determined by Shera *et al.* (2008) to be approximately 10-20% of $\hat{\tau}_{BM}$. This delay arises from the complex relation between the transportation pressure across the cochlear partition and the BM velocity or travelling wave pattern. As $\hat{\tau}_k$ is positive, the predicted delay ratio between BM velocity delay and SFOAE delay is typically less than 2. The data shown in the present study would be consistent with this prediction. Shera and Guinan (2003), Siegel *et al.* (2005), and Shera *et al.* (2008, 2005) made use of SFOAEs and obtained delay estimates using phase-gradient delays. As pointed out by Moleti and Sisto (2008) and Sisto *et al.* (2007), the interpretation of the SFOAE phase-gradient delays in terms of cochlear transmission delays is model dependent, due to the necessary simplifying assumptions to obtain a tractable model. Therefore, it is not possible to enter the debate of OAE generation mechanisms in the present study, other than to say that the data seem to align well with the general linear CRF theory model predictions at higher frequencies.

For frequencies below 4 kHz, Siegel *et al.* (2005) argued that the errors of CRF-theory prediction are much greater. They argued that CRF could not account for their data and that a different model of reverse energy propagating backwards outside the cochlea via a compression wave was necessary, as first suggested by Wilson (1980b). This was also suggested by Ren (2004), who used scanning laser interferometry to detect forward-traveling waves in gerbil cochleae, but failed to provide any evidence for backward-traveling waves. Differences in experimental paradigms and species could be the cause of the present study contradicting the conclusions of Siegel *et al.* (2005), where the present study would tend to support CRF theory throughout the range of frequencies tested. However, it should be noted that even though there is good experimental agreement between SFOAE and TEOAE, when stimulus intensity is in bandwidth-compensated SPL (cSPL) (Kalluri and Shera, 2007), the levels used in this study ranged from 42–58 dB cSPL as TB frequency drops from 8 to 0.5 kHz systematically. It is not clear what this level variation and the effects of spread of excitation may do for the OAE model predictions on delay, so one should be careful making a direct comparison to the historical studies reported in this section. Overall, these data from the present study seem to support the coherent reflection filtering theory within the experimental limitations.

V. SUMMARY AND CONCLUSION

This study estimated cochlear delays in humans using both OAEs and ABRs, evoked from TBs in the same subjects. It was shown that latency estimates can be reliably obtained from both methods. This study demonstrated that the ratio of OAE delay, τ_{OAE}, to the BM delay estimated from ABRs, $\hat{\tau}_{BM}^{(ABR)}$, to be 1.92 ± 0.42 ms, which is in line with the argument that OAEs are generated at a tonotopic place and propagate backward to the base as a reverse traveling wave, with the same speed as that of the forward wave.

This is further supported by the similar variation of delay with frequency between both the OAE delay, τ_{OAE}, and the BM delay estimated from ABRs, $\hat{\tau}_{BM}^{(ABR)}$.

ACKNOWLEDGMENTS

The OAE recording software was developed in collaboration with Manfred Mauermann at the Carl von Ossietzky Universität in Oldenburg, Germany. This project was jointly funded by GN Resound A/S, Oticon A/S, and Widex A/S and carried out by Gilles Pigasse as part of his doctoral studies.

Abdala, C., and Keefe, D. (**2006**). "Effects of middle-ear immaturity on distortion product otoacoustic emission suppression tuning in infant ears," J. Acoust. Soc. Am. **120**, 3832–3842.

Burkard, R., and Secor, C. (**2002**). "Overview of auditory evoked potential," in *Handbook of Clinical Audiology*, edited by J. Katz (Lippincott, Williams, and Wilkins, Philadelphia, PA), Chap. 14, pp. 233–248.

Cheng, J. (**1995**). "Time-frequency analysis of transient evoked otoacoustic emissions via smoothed pseudo Wigner distribution," Scand. Audiol. **24**, 91–96.

Don, M., and Eggermont, J. J. (**1978**). "Analysis of the click-evoked brainstem potentials in man using high-pass noise masking," J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., and Kwong, B. (**2002**). "Auditory brainstem response: Differential diagnosis," in *Handbook of Clinical Audiology*, edited by J. Katz (Lippincott, Williams, and Wilkins, Philadelphia, PA), Chap. 16, pp. 274–297.

Don, M., Ponton, C. W., Eggermont, J. J., and Kwong, B. (**1998**). "The effects of sensory hearing loss on cochlear filter times estimated from auditory brainstem response latencies," J. Acoust. Soc. Am. **104**, 2280–2289.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response," J. Acoust. Soc. Am. **94**, 2135–2148.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1994**). "Auditory brainstem response peak amplitude variability reflects individual differences in cochlear response times," J. Acoust. Soc. Am. **96**, 3476–3491.

Dong, W., and Cooper, N. P. (**2006**). "An experimental study into the acousto-mechanical effects of invading the cochlea," J. R. Soc., Interface **3**, 561–571.

Dong, W., and Olson, E. S. (**2008**). "Supporting evidence for reverse cochlear traveling waves," J. Acoust. Soc. Am. **123**, 222–240.

Eggermont, 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.

Elberling, C., Parbo, J., Johnsen, N., and Bagi, P. (**1985**). "Evoked acoustic emission: Clinical application," Acta Oto-Laryngol. **421**, 77–85.

Foxe, J. J., and Stapells, D. R. (**1993**). "Normal infant and adult auditory brain-stem responses to bone-conducted tones," Audiology **32**, 95–109.

Goodman, S., Withnell, R., de Boer, E., Lilly, D., and Nuttall, A. (**2004**). "Cochlear delays measured with amplitude-modulated tone-burst-evoked OAEs," Hear. Res. **188**, 57–69.

Gorga, M., Kaminski, J., Beauchaine, K., and Jesteadt, W. (**1988**). "Auditory brainstem responses to tone bursts in normally hearing subjects," J. Speech Hear. Res. **31**, 87–97.

Grandori, F. (**1985**). "Nonlinear phenomena in click- and tone-burst-evoked otoacoustic emissions from human ears," Audiology **24**, 71–80.

Hall, J. W., Bull, J. M., and Cronau, L. H. (**1988**). "Hypo- and hyperthermia in clinical auditory brain stem response measurement: Two case reports," Ear Hear. **9**, 137–143.

Hatzopoulos, S., Cheng, J., Grzanka, A., and Martini, A. (**2000**). "Time-frequency analysis of TEOAE recordings from normals and SNHL patients," Audiology **39**, 1–12.

Heil, P., and Irvine, D. R. (**1997**). "First-spike timing of auditory-nerve fibers and comparison with auditory cortex," J. Neurophysiol. **78**, 2438–2454.

Hoth, S., and Weber, F. (**2001**). "The latency of evoked otoacoustic emissions: Its relation to hearing loss and auditory evoked potentials," Scand. Audiol. **30**, 173–183.

Jasper, H. (**1958**). "The ten-twenty electrode system of the international federation," Electroencephalogr. Clin. Neurophysiol. **10**, 371–375.

Jedrzejczak, W., Blinowska, K., and Konopka, W. (**2005**). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," Hear. Res. **205**, 249–255.

Jedrzejczak, W., Blinowska, K., Kochanek, K., and Skarzynski, H. (**2008**). "Synchronized spontaneous otoacoustic emissions analyzed in a time-frequency domain," J. Acoust. Soc. Am. **124**, 3720–3729.

Jewett, D. L., and Williston, J. (**1971**). "Auditory-evoked far fields averaged from scalp of humans," Brain **94**, 681–696.

Kalluri, R., and Shera, C. A. (**2007**). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," J. Acoust. Soc. Am. **121**, 2097–2110.

Kapadia, S., and Lutman, M. E. (**2000**). "Nonlinear temporal interactions in click-evoked otoacoustic emissions. II. Experimental data," Hear. Res. **146**, 101–120.

Keefe, D. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T. (**2002**). "Exploring cochlear status with otoacoustic emissions: The potential for new clinical applications," in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattike (Thieme, New York), Chap. 1, pp. 1–47.

Kemp, D., Bray, P., Alexander, L., and Brown, A. (**1986**). "Acoustic emission cochleography-practical aspects," Scand. Audiol. Suppl. **1**, 71–95.

Kiang, N. Y. (**1975**). "Stimulus representation in the discharge patterns of auditory neurons," in *The Nervous System. Volume 3: Human Communication and Its Disorders*, edited by E. L. Eagles (Raven, New York), pp. 81–96.

Kim, D. O., and Molnar, C. E. (**1979**). "A population study of cochlear nerve fibers: Comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones," J. Neurophysiol. **42**, 16–30.

Konrad-Martin, D., and Keefe, D. H. (**2003**). "Time-frequency analysis of transient-evoked stimulus-frequency and distortion-product otoacoustic emissions: Testing cochlear model predictions," J. Acoust. Soc. Am. **114**, 2021–2043.

Long, G., and Talmadge, C. (**1997**). "Spontaneous otoacoustic emission frequency is modulated by heartbeat," J. Acoust. Soc. Am. **102**, 2831–2848.

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.

Moleti, A., and Sisto, R. (**2008**). "Comparison between otoacoustic and auditory brainstem response latencies supports slow backward propagation of otoacoustic emissions," J. Acoust. Soc. Am. **123**, 1495–1503.

Moleti, A., Sisto, R., and Paglialonga, A. (**2008**). "Transient evoked otoacoustic emission latency and estimates of cochlear tuning in preterm neonates," J. Acoust. Soc. Am. **124**, 2984–2994.

Moleti, A., Sisto, R., Tognoloa, G., Parazzini, M., and Ravazzani, P. (**2005**). "Otoacoustic emission latency, cochlear tuning, and hearing functionality in neonates," J. Acoust. Soc. Am. **118**, 1576–1584.

Møller, A. (**1994**). "Neural generators of auditory evoked potentials," in *Principles and Applications in Auditory Evoked Potentials*, edited by J. T. Jacobson (Allyn and Bacon, Massachusetts).

Møller, A. R., and Jannetta, P. J. (**1983**). "Interpretation of brainstem auditory evoked-potentials: Results from intracranial recordings in humans," Scand. Audiol. **12**, 125–133.

Murray, J. G., Cohn, E. S., Harker, L. A., and Gorga, M. P. (**1998**). "Tone burst auditory brain stem response latency estimates of cochlear travel time in Meniere's disease, cochlear hearing loss, and normal ears," Am. J. Otol. **19**, 854–859.

Narayan, S. S. (**1991**). "Comparison of latencies of N1 and transient evoked otoacoustic emissions: An evaluation of reverse travel in the cochlea," Ph.D. thesis, Purdue University, West Lafayette, IN.

Neely, S., Norton, S., Gorga, M., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.

Norton, S., and Neely, S. (**1987**). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1860–1872.

Probst, R., Coats, A., Martin, G., and Lonsbury-Martin, B. (**1986**). "Spontaneous, click-, and toneburst-evoked otoacoustic emissions from normal ears," Hear. Res. **21**, 261–275.

Puria, S. (**2003**). "Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions," J. Acoust. Soc. Am. **113**, 2773–2789.

Recio, A., Rich, N. C., Narayan, S. S., and Ruggero, M. A. (**1998**). "Basilar-membrane responses to clicks at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **103**, 1972–1989.
Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.
Ren, T., and Nuttall, A. L. (**2006**). "Cochlear compression wave: an implication of the Allen–Fahey experiment," J. Acoust. Soc. Am. **119**, 1940–1942.
Riedel, H., Granzow, M., and Kollmeier, B. (**2001**). "Single-sweep-based methods to improve the quality of auditory brain stem responses Part II: Averaging methods," Z. Fuer Audiologie, Audiological Acoust. **40**, 62–85.
Ruggero, M. A., and Temchin, A. N. (**2007**). "Similarity of traveling-wave delays in the hearing organs of humans and other tetrapods," J. Assoc. Res. Otolaryngol. **8**, 153–166.
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.
Schönweiler, R., Neumann, A., and Ptok, M. (**2005**). "Frequency specific auditory evoked responses. Experiments on stimulus polarity, sweep frequency, stimulus duration, notched-noise masking level, and threshold estimation in volunteers with normal hearing," HNO **53**, 983–994.
Şerbetçioğlu, M. B., and Parker, D. J. (**1999**). "Measures of cochlear travelling wave delay in humans: I. Comparison of three techniques in subjects with normal hearing," Acta Oto-Laryngol. **119**, 537–543.
Shera, C. A., and Guinan, J. J. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, J. J. (**2003**). "Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning," J. Acoust. Soc. Am. **113**, 2762–2772.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2005**). "Coherent reflection in a two-dimensional cochlea: Short-wave versus long-wave scattering in the generation of reflection-source otoacoustic emissions," J. Acoust. Soc. Am. **118**, 287–313.
Shera, C. A., Tubis, A., and Talmadge, C. L. (**2008**). "Testing coherent reflection in chinchilla: Auditory-nerve responses predict stimulus-frequency emissions," J. Acoust. Soc. Am. **124**, 381–395.
Shera, C. A., Tubis, A., Talmadge, C. L., de Boer, E., Fahey, P. F., and Guinan, J. J. (**2007**). "Allen–Fahey and related experiments support the predominance of cochlear slow-wave otoacoustic emissions," J. Acoust. Soc. Am. **121**, 1564–1575.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., Van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
Sisto, R., and Moleti, A. (**2007**). "Transient-evoked otoacoustic emission latency and cochlear tuning at different stimulus levels," J. Acoust. Soc. Am. **122**, 2183–2190.
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 Shera, C. A. (**2007**). "Cochlear reflectivity in transmission-line models and otoacoustic emission characteristic time delays," J. Acoust. Soc. Am. **122**, 3554–3561.
Stapells, D. R. (**1994**). "Low-frequency hearing and the auditory brainstem response," Am. J. Audiol. **7**, 11–13.
Stapells, D. R., and Picton, T. W. (**1981**). "Technical aspects of brainstem evoked potential audiometry using tones," Ear Hear. **2**, 20–29.
Stover, L., and Norton, S. (**1993**). "The effects of aging on otoacoustic emissions," J. Acoust. Soc. Am. **94**, 2670–2681.
Stürzebecher, E., Cebulla, M., Elberling, C., and Berger, T. (**2006**). "New efficient stimuli for evoking frequency-specific auditory steady-state responses," J. Am. Acad. Audiol **17**, 448–461.
Suzuki, T., and Horiuchi, K. (**1981**). "Rise time of pure-tone stimuli in brain stem response audiometry," Audiology **20**, 101–112.
Thornton, A., Lineton, B., Baker, V., and Slaven, A. (**2006**). "Nonlinear properties of otoacoustic emissions in normal and impaired hearing," Hear. Res. **219**, 56–65.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
von Békésy, G. (**1960**). *Experiments in Hearing* (McGraw Hill, New York).
Wilson, J. P. (**1980a**). "Evidence for a cochlear origin for acoustic re-emission, threshold fine structure and tonal tinnitus," Hear. Res. **2**, 233–252.
Wilson, J. P. (**1980b**). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," Hear. Res. **2**, 527–532.
Wit, H., and Ritsma, R. (**1980**). "Evoked acoustical responses from the human ear: Some experimental results," Hear. Res. **2**, 253–261.
Wit, H., van Dijk, P., and Avan, P. (**1994**). "Wavelet analysis of real ear and synthesized click evoked otoacoustic emissions," Hear. Res. **73**, 141–147.
Zweig, G., and Shera, C. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Estimation of cochlear response times using lateralization of frequency-mismatched tones

Olaf Strelcyk and Torsten Dau
Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Building 352, Ørsteds Plads, 2800 Kongens Lyngby, Denmark

(Received 3 April 2009; revised 1 July 2009; accepted 1 July 2009)

Behavioral and objective estimates of cochlear response times (CRTs) and traveling-wave (TW) velocity were compared for three normal-hearing listeners. Differences between frequency-specific CRTs were estimated via lateralization of pulsed tones that were interaurally mismatched in frequency, similar to a paradigm proposed by Zerlin [(1969). J. Acoust. Soc. Am. **46**, 1011–1015]. In addition, derived-band auditory brainstem responses were obtained as a function of derived-band center frequency. The latencies extracted from these responses served as objective estimates of CRTs. Estimates of TW velocity were calculated from the obtained CRTs. The correspondence between behavioral and objective estimates of CRT and TW velocity was examined. For frequencies up to 1.5 kHz, the behavioral method yielded reproducible results, which were consistent with the objective estimates. For higher frequencies, CRT differences could not be estimated with the behavioral method due to limitations of the lateralization paradigm. The method might be useful for studying the spatiotemporal cochlear response pattern in human listeners.
 [DOI: 10.1121/1.3192220]

PACS number(s): 43.64.Ri, 43.64.Kc, 43.66.Pn, 43.66.Lj [BLM] Pages: 1302–1311

I. INTRODUCTION

The cochlea separates a sound into its constituent tonal components and distributes their responses spatially along its length by the distinctive spatial and temporal vibration patterns of its basilar membrane (BM). For example, the vibration pattern evoked by a single tone appears as a traveling wave (TW) (e.g., Ruggero, 1994; Robles and Ruggero, 2001). This wave propagates down the cochlea and reaches maximum amplitude at a particular point, before slowing down and decaying rapidly. The lower the frequency of the tone, the further its wave propagates down the cochlea. Hence, each point along the cochlea has a characteristic frequency (CF) to which it is most responsive. This tonotopic map is an important organizational principle of the primary auditory pathway and is preserved all the way to the auditory cortex (Clarey *et al.*, 1992).

At the level of the auditory nerve, the frequency of a tone is encoded both spatially, by its CF location, and temporally, by the periodicity of the responses in the nerve fibers that innervate the CF (cf. Ruggero, 1992). Several studies have suggested that the extraction of spatiotemporal information, i.e., the combination of phase-locked responses and systematic frequency-dependent delays along the cochlea (associated with the TW), may be important in the context of pitch perception (e.g., Loeb *et al.*, 1983; Shamma and Klein, 2000), loudness perception (Carney, 1994), localization (e.g., Shamma *et al.*, 1989; Joris *et al.*, 2006), speech formant extraction (e.g., Deng and Geisler, 1987), and tone-in-noise detection (e.g., Carney *et al.*, 2002). It has been proposed that a distorted spatiotemporal response might be, at least partly, responsible for the problems of hearing-impaired listeners to process temporal-fine-structure information (e.g., Moore, 1996; Moore and Skrodzka, 2002; Buss *et al.*, 2004). This may be one of the reasons for their difficulties to understand speech in noise. However, so far, empirical evidence for spatiotemporal information processing in humans is lacking since BM response patterns are difficult to monitor.

This study focused on one important component of the spatiotemporal BM response pattern: the cochlear response time (CRT) (e.g., Don *et al.*, 1993), which reflects the propagation delay of the TW. Consistent estimates of frequency-specific CRTs in humans have been obtained using different objective noninvasive methods, such as measurements of compound action potentials (e.g., Eggermont, 1976), stimulus-evoked otoacoustic emissions (e.g., Norton and Neely, 1987; Tognola *et al.*, 1997), tone-burst-evoked auditory brainstem responses (ABRs) (e.g., Gorga *et al.*, 1988), and derived-band click-evoked ABRs (e.g., Don and Eggermont, 1978; Parker and Thornton, 1978a; Eggermont and Don, 1980; Donaldson and Ruth, 1993; Don *et al.*, 1993).

Early psychoacoustic attempts to estimate CRTs or TW velocity were motivated by von Békésy's (1933) observation that the perceived position of clicks, presented to both ears, varied systematically when low-frequency masking tones were presented to one ear. Elaborating on this, Schubert and Elpern (1959) presented clicks in the presence of high-pass filtered noise with cutoff frequencies differing by half an octave between the two ears. The interaural time difference (ITD) that centered the unified percept at the midline was taken as an estimate of the difference in CRTs between the BM places corresponding to the noise cutoff frequencies in the two ears. However, the TW velocity derived from these CRT disparities was substantially larger than the TW velocity

estimates obtained by means of the above mentioned objective methods (e.g., Donaldson and Ruth, 1993). As mentioned by Deatherage and Hirsh (1959) and Zerlin (1969), interaural loudness differences of the clicks might have influenced lateralization in the paradigms used by von Békésy (1933) and Schubert and Elpern (1959).

Instead of using click stimuli, von Békésy (1963b) and later Zerlin (1969) used pulsed tones that were interaurally mismatched in frequency. Both, von Békésy and Zerlin reported that listeners perceived the tones as fused, lateralized toward the ear receiving the higher-frequency tone. Zerlin measured the ITD needed to center the percept of the tones and took this as an estimate of the difference in CRTs between the BM places corresponding to the different tone frequencies in the two ears. The derived TW velocities were in good agreement with objective estimates of TW velocity (cf. Donaldson and Ruth, 1993). However, as noted by Neely *et al.* (1988), the reliability of Zerlin's estimates may be limited considering the difficulty of the psychoacoustic task and the fact that no further reports have been published since the original study in 1969.

If the lateralization of the interaurally mismatched tones reflected differences in CRTs, the paradigm would present a direct link between early cochlear disparities and spatial perception. Hence, particularly in view of the high temporal acuity of binaural auditory processing, which resolves ITD changes of less than 10 μs (Yost, 1974), this behavioral paradigm might serve as a complement to the objective measures of CRT mentioned above. Furthermore, Zerlin's (1969) paradigm bears a close relation to the concept of (across-ear) spatiotemporal processing. In both concepts, lateralization is supposed to be based on the comparison of information from mismatched frequency channels in the two ears. However, it is not clear if the lateralization in Zerlin's paradigm is based on interaural level differences (in the envelope at onset/offset), interaural time differences (in the fine structure), or a combination of both. Buus *et al.* (1984) suggested that temporal-fine-structure information during the first tone cycles might play a role in the lateralization of mismatched tones at low frequencies. This was supported by Magezi and Krumbholz (2008), who provided evidence that the binaural system can extract fine-structure information from interaurally mismatched frequency channels.

In the present study, behavioral estimates of CRT disparities and TW velocity were obtained for three normal-hearing listeners, using a similar paradigm to the one used by Zerlin (1969). In order to minimize measurement variability due to subjective listener criteria, an adaptive procedure was used to determine the ITD that centered the unified percept. The influences of loudness balancing, tone presentation level, and potential between-ear asymmetries on the CRT and TW velocity estimates were examined. For direct comparison, estimates of CRTs and TW velocities for the same listeners were obtained from derived-band ABRs. Since these estimates provide an objective "reference," they are presented first.

II. ABRs

A. Method

1. Listeners

The three female listeners were aged between 23 and 24 years and had audiometric thresholds better than 20 dB hearing level (ISO 389-8, 2004) at all octave frequencies from 125 to 8000 Hz and from 750 to 6000 Hz.

2. Stimuli

Rarefaction clicks were produced by applying 83-μs rectangular pulses (generated in MATLAB®) to an Etymotic Research ER-2 insert earphone. The clicks were presented monaurally at a level of 93-dB peak-to-peak equivalent sound pressure level (ppe SPL), with a repetition rate of 45 Hz. The acoustic clicks were calibrated using an occluded-ear simulator [IEC 60711, 1981; Brüel & Kjær (B&K) 4157] mounted with an ear-canal extension (B&K DP0370). Response latencies were corrected for a constant 1-ms delay introduced by the tubing of the ER-2 earphone.

Ipsilateral pink-noise masking was used to obtain derived-band ABRs (Don and Eggermont, 1978). High-pass noise maskers with cutoff frequencies of 0.5, 1, 2, 4, and 8 kHz were generated in the spectral domain as random-phase noise (with components outside the passband set to zero) and played back via a second ER-2 insert earphone, which was coupled to the first ER-2 earphone via an ER-10B+ transducer (without using the microphone). The spectrum level of the high-pass noise maskers was identical to that of the broadband pink noise, for which a level of 91 dB SPL was found to be sufficient to mask the ABR to the 93-dB ppe SPL clicks.

Perceptual click thresholds were measured for 500-ms click trains using a three-interval, three-alternative, forced-choice (3I-3AFC) task, tracking the 71%-correct point (one up, two down) on the psychometric function. The final threshold was estimated as the arithmetic mean over three runs. The average click threshold for the three listeners was 33.7 (31.5, 36.4) dB ppe SPL, with the values in parentheses representing the range of the individual results. These thresholds are lower than the corresponding reference threshold of 43.2 dB ppe SPL given by Richter and Fedtke (2005), which can be attributed to differences in click repetition rate and the different ear tips used. The ER1-14A used by Richter and Fedtke and the ER10-14 used in the present study differ in the diameter of the ear-tip tubes.

3. ABR recordings

Listeners lay on a couch in an acoustically and electrically shielded booth. The ABRs were measured differentially between electrodes applied to the vertex (C_z in the 10/20 system) and the ipsilateral mastoid (M_1 or M_2). Another electrode applied to the forehead (F_{pz}) served as ground. The electrode signals were acquired using a Neuroscan SynAmps 2 system, at a sampling rate of 20 kHz. Off-line bandpass-filtering between 0.1 and 2 kHz (forward-backward filtering) was applied. Weighted averaging, as discussed in Elberling and Wahlgreen (1985) and in Don and

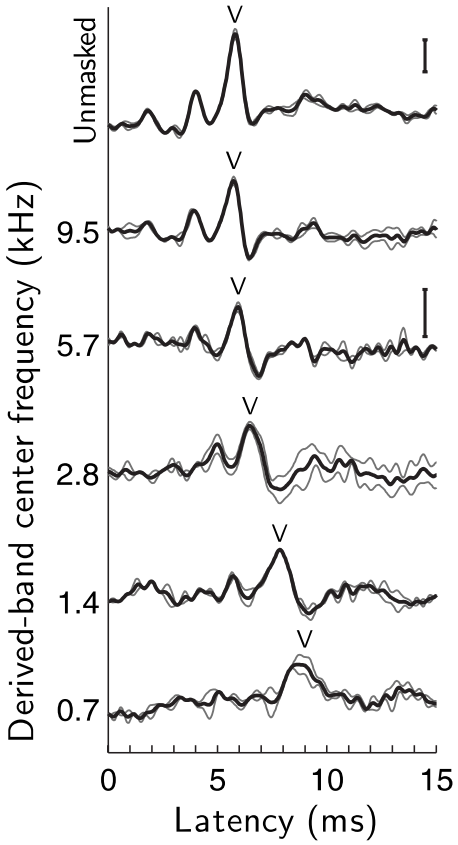


FIG. 1. Examples of unmasked and derived-band ABRs to 93-dB ppe SPL clicks from one listener. Two replications (gray) and their average (black) are shown. Wave *V*s are indicated by the corresponding symbols. The bars to the right represent 200 nV. If no bar is shown, the nearest bar above holds.

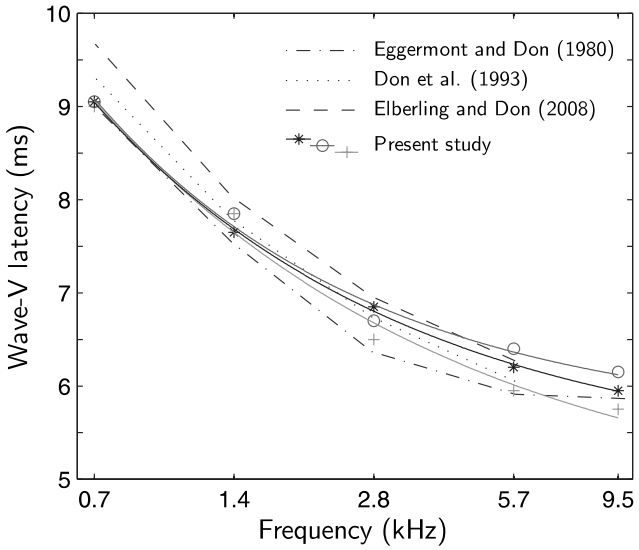


FIG. 2. Measured derived-band ABR wave-*V* latencies (symbols) for three listeners in response to 93-dB ppe SPL clicks, as a function of the derived-band center frequency. The solid curves show individual model fits according to Eq. (1). For comparison, the dash-dotted, dotted, and dashed curves show latency results from Eggermont and Don (1980), Don *et al.* (1993), and Elberling and Don (2008), respectively. The same center frequencies as in Elberling and Don (2008) were assigned to the derived-band latencies of Eggermont and Don (1980), since the same high-pass masking noise stimuli were used in both studies.

Elberling (1994), was used for estimation of the auditory evoked potentials. Two replications, each consisting of 4096 sweeps, were recorded. The 4096 sweeps were subdivided into 16 equally sized blocks and averaged. Each block was weighted in inverse proportion to its amount of background noise, which was estimated as the sweep-to-sweep variance at a single point in time (Elberling and Don, 1984). The residual background noise level in the final evoked potential estimates was 23 nV, averaged across listeners and conditions.

4. Analysis

Narrow-band cochlear contributions to the ABR were derived by means of the derived-band technique (e.g., Don and Eggermont, 1978; Parker and Thornton, 1978b, 1978a). Derived-band ABRs, i.e., differences between the ABRs to clicks presented in adjacent high-pass maskers, were obtained and the corresponding wave-*V* latencies were extracted. The center frequencies of the derived bands were computed as the geometric means of the two corresponding high-pass cut-off frequencies (Parker and Thornton, 1978a). The frequency of 11.3 kHz, where the acoustic-click power was attenuated by 30 dB, was chosen as the upper frequency limit of the highest derived band. Hence, the following frequencies were assigned to the derived bands: 0.7, 1.4, 2.8, 5.7, and 9.5 kHz.

Figure 1 illustrates a series of derived-band ABRs from one listener. Wave *V*s are indicated. As can be seen, wave-*V* latencies increased with decreasing derived-band center frequency. For the further analysis of the wave-*V* latencies, the following latency model was adapted from Neely *et al.* (1988):

$$\tau(f) = a + bf^{-d}, \tag{1}$$

where f represents the derived-band center frequency, normalized to 1 kHz, and a, b, and d are fitting constants. The model parameter a represents an asymptotic delay. It reflects the post-cochlear contributions, i.e., synapse and neural conduction delays, to the wave-*V* latency, which are independent of frequency (cf. Don and Eggermont, 1978; Ponton *et al.*, 1992; Ruggero, 1992).

B. Results

Figure 2 shows the measured (symbols) and fitted (solid curves) wave-*V* latencies. The results of all three listeners were similar. Latencies decreased with increasing frequency from about 9 ms at 0.7 kHz to about 6 ms at 9.5 kHz. For comparison, previously reported latencies from Eggermont and Don, 1980 (dash-dotted curve), Don *et al.*, 1993 (dotted curve), and Elberling and Don, 2008 (dashed curve) are shown. The results of the present study agree well with those from the earlier studies. The latency model specified in Eq. (1) provided a good description of the individual latency data, with a residual root-mean-square (rms) fitting error of 0.09 (0.03, 0.13) ms, averaged across listeners (values in parentheses represent the range of individual results). The mean estimated parameters were a=5.1 ms, b=3.2 ms, and d=0.6.

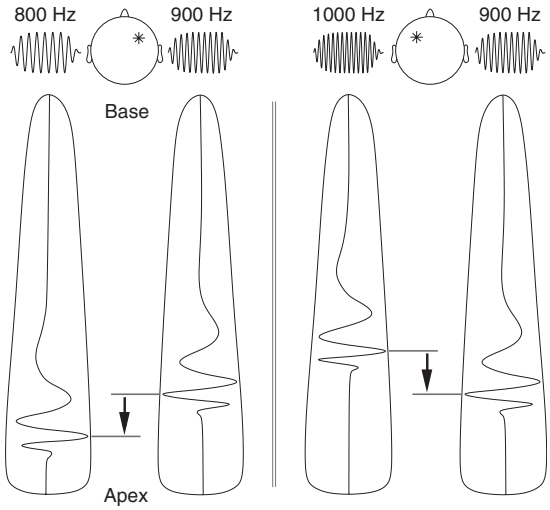


FIG. 3. Sketch of the stimuli used in the lateralization task, for the 800|900-Hz (top left) and 1000|900-Hz (top right) conditions. In the depicted configuration, the left ear corresponds to the ABR test-ear. Basilar membrane traveling waves are indicated at the bottom. It is assumed that the CRT disparities, indicated by the arrows, can be measured in terms of the ITDs that center the percepts at the midline.

III. LATERALIZATION OF MISMATCHED TONES

A. Method

1. Listeners

The lateralization measurements were performed by the same listeners who participated in the ABR measurements.

2. Stimuli and procedures

Short trains of tone bursts with interaurally mismatched frequencies f_1 and f_2 were presented to the two ears, as illustrated in Fig. 3. In the following, the notation $f_1|f_2$ is used where f_1 represents the frequency of the tone presented in the ABR test-ear and f_2 the frequency of the tone presented in the other ear. The considered tone frequencies were 400|480, 800|900, 1000|900, and 1400|1550 Hz. Each tone burst had a total duration of 40 ms, including an exponential onset with a rise time of 10 ms and a 10-ms raised-cosine shaped offset-ramp. In contrast to Scharf *et al.* (1976) and Buus *et al.* (1984), who used exponential ramps at onset and offset, a cosine offset-ramp was used here in order to minimize spectral splatter. The tones were presented in sine phase, i.e., the onset-ramp started with a positive-going zero crossing of the sinusoid. Each train consisted of six tone bursts, separated by 40-ms silent gaps. Its lateralization was varied by introducing a waveform delay to one of the ears, giving rise to an ITD. The ITD that produced a unified percept centered at the midline was measured.

A two-interval, two-alternative, forced-choice task was used. The first interval always contained the diotic reference tone-burst train, consisting of both tones (with frequencies f_1 and f_2) in both ears, while the second interval contained the $f_1|f_2$ target train. Listeners were instructed to indicate if the latter was lateralized to the left or right side relative to the reference train. In order to ease the task, the whole trial consisting of reference and target train was repeated once before the listener made a response. If the target train was lateralized to the right, the ITD was adjusted such that the percept would move further to the left in the next presentation, and vice versa. Following the adaptive procedure for subjective judgments introduced by Jesteadt (1980), two sequences of trials were interleaved, tracking 71% (one up, two down) and 29% (two up, one down) lateralization to the right. Each of these sequences was terminated after ten reversals, and the tracked ITDs were estimated as the arithmetic means of all ITD values following the sixth reversals. Subsequently, the ITD yielding a centered percept was estimated by calculating the mean of the two ITDs leading to 71% and 29% lateralization judgments to the right.

ITDs were measured for tone levels of 50 and 75 dB SPL. In addition to the ITDs in quiet, for the 800|900-Hz tones at 75 dB, ITDs were measured in the presence of a diotic notched-noise background (flat-spectrum noise bands of 100–700 and 1000–9000 Hz), which limited spread of excitation. The noise was presented continuously during the whole run, with a spectrum level of 16 dB SPL. For higher levels, a fused position of the tones could no longer be perceived.

Prior to actual data collection, listeners received up to ten runs of training until consistent ITD results were obtained. The final ITD was estimated as the arithmetic mean over four interleaved runs. If the standard deviation (SD) over these runs, relative to the mean ITD, exceeded a factor of 0.1, additional runs were taken and the average of all was used. The final relative standard error of the ITD estimate, averaged across listeners and conditions, was 0.05.

3. Loudness balancing

In addition to the conditions where the tones were presented at equal SPLs, ITDs were measured with the tones balanced in loudness between the two ears. Loudness balancing was also applied by Zerlin (1969). The adaptive procedure introduced by Jesteadt (1980) was used for the loudness balancing of the frequency-mismatched tones. The first interval contained the f_1-tone, presented to the ABR test-ear, and the second interval contained the f_2-tone, presented to the other ear. Listeners were instructed to indicate if the second tone was perceived as softer or louder than the first tone. As in the lateralization task, the whole trial was repeated once before the listener made a response. The interaural level balance was adjusted to yield both 71% and 29% judgments of the second tone to be the louder one. The point of equal loudness was estimated as the arithmetic mean of these two loudness adjustments (in decibels). An equal number of runs were performed with the opposite order of presentation, i.e., with the f_2-tone presented in the first interval and the f_1-tone presented in the second interval.

The final level adjustment for loudness balancing was estimated as the arithmetic mean over at least six interleaved runs. The final standard error of the level adjustment was 0.4 (SD 0.2) dB, averaged across listeners and conditions. There were no significant differences between listeners and conditions [$p > 0.1$].

TABLE I. The ITDs yielding centered percepts of the tones with interaurally mismatched frequencies f_1 and f_2, for three listeners (the numbers in parentheses represent standard errors). LB denotes loudness balancing. The ABR wave-V latency differences $\Delta\tau_{ABR}$ between the frequencies f_1 and f_2 are also given for the individual listeners. The values in square brackets are based on extrapolations beyond the range of measured frequencies. Conditions for which the listener could not perform the lateralization task are indicated by "NM" (not measurable). Dots indicate combinations that were not measured.

Tone level	$f_1 \mid f_2$ (kHz)	ITD (μs) for NH_1		ITD (μs) for NH_2		ITD (μs) for NH_3		$\Delta\tau_{ABR}$ (μs)		
		With LB	Without LB	With LB	Without LB	With LB	Without LB	NH_1	NH_2	NH_3
50 dB	0.4|0.48	442(38)	340(40)	404(6)	396(8)	357(8)	335(5)	[580	648	597]
	0.8|0.9	205(12)	264(13)	186(6)	232(10)	184(7)	207(3)	261	260	255
	1.0|0.9	187(18)	185(13)	224(2)	262(3)	173(3)	180(9)	220	215	212
	1.4|1.55	110(5)	138(9)	NM	129(22)	NM	NM	167	151	155
	$\frac{0.8}{1.0}$ |	392(21)	449(18)	410(7)	494(11)	356(8)	387(9)	481	475	467
75 dB	0.8|0.9	99(3)	⋯	79(5)	⋯	61(7)	⋯			
	0.8|0.9 in noise	190(6)	⋯	183(2)	⋯	192(6)	⋯			

4. Apparatus

The stimuli were generated in MATLAB® and converted to analog signals using a 24-bit digital-to-analog converter (RME DIGI96/8) with a sampling rate of 96 kHz. The stimuli were presented in a double-walled sound-attenuating booth via Sennheiser HD580 headphones. Calibrations were done using an ear simulator (IEC 60318-1and -2, 1998; B&K 4153 with flat plate) and, prior to playing, 128-tap linear-phase FIR equalization filters were applied to the stimuli, rendering the headphone frequency response flat.

B. Results and discussion

1. Response-time differences

The results of the lateralization measurements for the three listeners are presented in Table I. It shows the ITDs that led to centered percepts of the 50- and 75-dB tones with interaurally mismatched frequencies f_1 and f_2. The ITDs are given for the conditions with and without interaural loudness balancing. As illustrated in Fig. 3, the frequency-mismatched tones with zero ITD were always lateralized toward the ear receiving the higher-frequency tone, consistent with previous reports in literature (e.g., von Békésy, 1963b; Zerlin, 1969). Hence, the sound presented to this ear required a delay in order to center the percept (for this reason, ITDs are stated only in absolute terms in the following). The centering ITDs were generally consistent and well reproducible. Therefore, the standard errors of the ITD estimates were relatively small. For comparison, the objective ABR wave-V latency differences $\Delta\tau_{ABR}$ are also represented in Table I (rightmost column). They were calculated on the basis of the individual latency fits to the derived-band ABR data, which followed the model in Eq. (1) and were shown in Fig. 2. The lowest derived-band frequency was 700 Hz. Therefore, the extrapolation to lower frequencies (400|480 Hz) should be regarded with caution. At the remaining frequencies of 800|900, 1000|900, and 1400|1550 Hz (second, third, and fourth rows in Table I, respectively), the perceptual ITD-based measure and the objective ABR-based measure yielded very similar results. The average rms deviation between the ITDs (without loudness balancing) and the latency differences $\Delta\tau_{ABR}$ was 39 μs. The correspondence between the behavioral and the objective data is remarkably good, given the different experimental paradigms. It strongly supports the hypothesis that the ITDs that produced centered sound images reflected differences in CRTs between remote places on the BM.

The ITDs reflect interaural time differences whereas the ABR latency differences $\Delta\tau_{ABR}$ reflect monaural time differences. Hence, part of the remaining deviations between these two could be due to differences in CRTs between the left and right cochleae (e.g., differences in the cochlear frequency-place maps). Therefore, the ITDs for the 800|900-Hz and 1000|900-Hz tone pairs were added (see fifth row in Table I). Since these tone pairs shared the common reference frequency of 900 Hz (cf. Fig. 3), the sum estimates the time difference between 800 and 1000 Hz in the ABR test-ear alone. Still, similar deviations from the ABR latencies as for the single-tone-pair ITDs were observed for these "monaural" time differences. Hence, the remaining deviations did not seem to be attributable to asymmetries between the left and right cochleae.

In addition to the measurements at 50 dB, for the 800|900-Hz tones, measurements were also performed at the higher tone level of 75 dB. For all listeners, ITDs were shorter at 75 dB than at 50 dB, by an average factor of 2.5. However, in the presence of the notched-noise masker, the ITDs obtained with the 75-dB tones were essentially identical to those obtained with 50-dB tones presented in quiet. This is consistent with the following interpretation in terms of excitation spread on the BM. The higher the tone level, the larger is the spread of excitation toward places with higher CFs than the nominal tone frequencies f_1 and f_2. The disparities in CRT between these places are smaller than at the nominal places due to the exponentially decreasing latency-frequency dependence (cf. Fig. 2). Therefore, smaller centering ITDs would be expected for the higher tone level of 75 dB than for the lower level of 50 dB. The notched noise limits excitation spread. This may explain why similar ITDs were obtained for the 75-dB tones in noise as for the 50-dB tones in quiet, for which spread of excitation plays a minor role. Hence, the observed effects of tone level and

background noise further indicate that the perceived lateralization of the mismatched tones reflected cochlear disparities.

Different stimuli, clicks versus tones, were used for the ABR recordings and the lateralization measurements, respectively. It seems reasonable to assume that stimulation at equal sensation levels results in similar levels of neural excitation, summed across the BM. The sensation level of the 93-dB ppe SPL clicks was 59 dB, averaged across listeners. The average sensation level of the mismatched 50-dB SPL tones with center frequencies of 890 and 1470 Hz was 49 dB (the same 3I-3AFC task was used for estimation of the click and tone thresholds). However, the tones excited only a limited part of the cochlea, while the broadband clicks excited most of the cochlea partition. Hence, the "effective" click levels in the one-octave-wide derived bands were lower than the nominal click level. In order to estimate these levels, the portion of the click power falling within the derived bands was calculated based on the acoustic-click power spectra. For the 0.7- and 1.4-kHz derived bands, this yielded values of −11 and −9 dB relative to the broadband click level, respectively. Hence, within these derived bands, the effective click level was about 83 dB ppe SPL, corresponding to a sensation level of 49 dB, which matches the sensation level of the mismatched tones. Also, remaining level differences should be of minor importance, since the 75-dB tones yielded very similar ITDs to the 50-dB tones when notched-noise masking was applied.

All three listeners had more difficulties with the lateralization task for the mid-frequency tones (1400|1550 Hz) than for the low-frequency tones. At 1400|1550 Hz, listener NH_2 could not consistently lateralize the mismatched tones when loudness balancing was applied, while listener NH_3 could not consistently lateralize the tones whether loudness balancing was applied or not. None of the listeners could perform the task reliably for frequencies above 1.5 kHz. Here, the sound image could not be lateralized with reasonable precision. It was perceived as rather diffuse and often did not cross the midline.

2. Loudness balancing

For all tone pairs, ITDs changed systematically when loudness balancing was applied: The ITD increased (decreased) when the level of the higher-frequency tone was increased (decreased). The level adjustment was 0.7 (SD 0.4) dB, averaged across listeners and conditions, without showing a systematic pattern across listeners and conditions. The ITDs obtained without loudness balancing seemed to match the objective latency differences $\Delta\tau_{ABR}$ slightly better than the ones obtained with loudness balancing. The average rms deviations were 39 and 66 μs, respectively, excluding the 400|480-Hz data.

Depending on the mechanism underlying the lateralization of the mismatched tones, loudness or level imbalances could influence the results of the lateralization measurements. While a temporal (phase-locking-based) mechanism should hardly be affected, a mechanism based on interaural level differences should be sensitive to level/loudness imbalances. The observed systematic change in ITDs with loudness balancing may indicate that interaural level cues contributed to the lateralization of the mismatched tones, although the small ITD changes might simply reflect changes in CRT with tone level. In any case, the centering ITDs obtained with and without loudness balancing were fairly comparable. This is consistent with the hypothesis that the lateralization was, at least to some extent, based on a temporal mechanism. This hypothesis is corroborated by the finding that tone-onset phase influences the lateralization of mismatched tones for frequencies below about 2 kHz (Scharf *et al.*, 1976; Buus *et al.*, 1984). With respect to the estimation of CRT disparities, the observed invariance of the results to loudness balancing is crucial. If the lateralization depended strongly on interaural level (or loudness) imbalances, it would be impossible to assess disparities in CRTs with this method. The trade-off between timing and level would give rise to unresolvable ambiguities.

As mentioned above, the behavioral results obtained without loudness balancing matched the objective data better than the ones obtained with loudness balancing. The observed loudness imbalances might have been due to different amounts of excitation or specific loudness (Moore *et al.*, 1997) at the two tone frequencies f_1 and f_2 as well as due to frequency-independent *between-ear* differences in excitation or specific loudness. Depending on which of these two factors was dominant, either equal loudness or equal SPLs at the two ears would be more appropriate for the lateralization paradigm. Frequency-independent between-ear differences in excitation/specific loudness would affect the lateralization of the diotic reference stimulus and mismatched target stimulus in the same way. Therefore, the ITD necessary for matching their positions would not be affected, as long as loudness balancing was applied neither to the reference stimulus nor to the target stimulus.[1] Hence, the better match between the objective results and the behavioral results obtained without loudness balancing and the lack of a systematic effect of tone frequency on loudness balancing suggest that the observed loudness imbalances may have reflected between-ear differences rather than frequency-dependent variations in excitation/specific loudness.

3. Traveling–wave velocity

Assuming that the centering ITDs and latency differences $\Delta\tau_{ABR}$ reflected travel times on the BM, the corresponding TW velocities were estimated using the cochlear frequency-place map supplied by Greenwood (1961).[2] The ratio of the distance between the CF places (corresponding to the mismatched-tone frequencies) and the centering ITD was taken as behavioral estimate of TW velocity at the geometric mean frequency of the two tones. Objective ABR-based estimates were derived by substituting the frequency f in Eq. (1) by the CF place x on the BM [Eq. (2) in Greenwood, 1961] and taking the derivative $-dx/d\tau$ as estimate of velocity.

Figure 4 shows the TW velocity estimates, based on the ABR latencies (curves) and the centering ITDs (bullets) obtained for the 50-dB tones without loudness balancing. The behavioral velocity estimate at 890 Hz (geometric means of 800 and 1000 Hz) is based on the "monaural" time difference

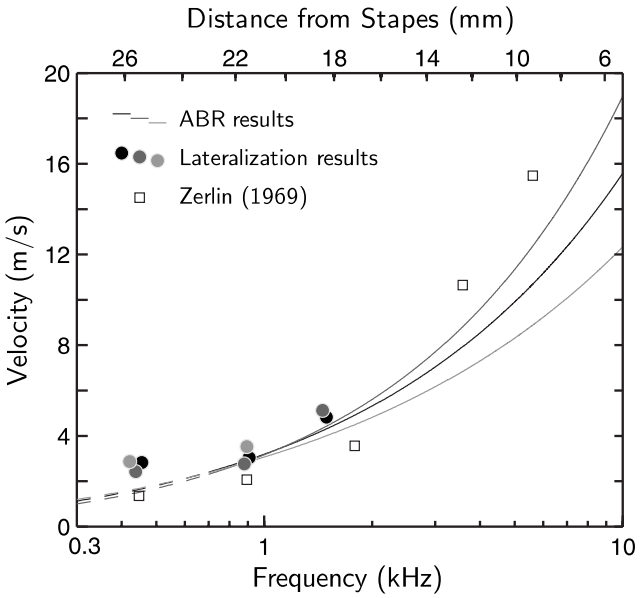


FIG. 4. TW velocity as a function of frequency/distance from stapes for three listeners. The solid curves represent the individual velocity estimates derived from the derived-band ABR latencies. At low frequencies, the curves are dashed since they are extrapolated beyond the actual measurement range. The bullets denote the estimates based on the mismatched-tone ITDs. For better visibility, they are slightly horizontally displaced for the individual listeners. The squares are corresponding estimates based on the ITDs reported by Zerlin (1969).

obtained by summing the 800|900 and 1000|900-Hz ITDs. For direct comparison, the open squares indicate velocities that were derived from Zerlin's (1969) ITDs.[3]

The ITD-based velocity estimates were consistent with the ABR-based velocity estimates. In both measures, velocities increased with increasing frequency. In order to compare the ITD-based estimates at 440 Hz with the ABR-based estimates, the ABR data were extrapolated beyond the actual measurement range (dashed part of the curves). Here, the deviations between the two measures were larger than at the higher frequencies of 890 and 1470 Hz, reflecting the corresponding deviations of the CRT estimates (compare ITDs and $\Delta\tau_{\mathrm{ABR}}$ values in Table I). The larger behavioral velocity estimates at 440 Hz might indicate that the actual latency-frequency functions were less steep at the low frequencies (below about 700 Hz) than the predictions based on the extrapolation of the ABR latencies (Fig. 2). This would be consistent with the latency-frequency curves in Fig. 1 of Neely *et al.* (1988), obtained from tone-burst-evoked ABRs, which showed shallower slopes for frequencies below about 500 Hz than for the higher frequencies. However, Ruggero and Temchin (2007) noted that the use of different tone-burst rise times for the different frequencies in the study by Neely *et al.* (1988) could have affected the observed ABR latencies.

Only small inter-individual differences were observed for frequencies up to 2 kHz, consistent with Donaldson and Ruth (1993). For frequencies above 1.5 kHz, no centering ITDs and thus no behavioral velocity estimates could be obtained in this study. At low frequencies, the velocity estimates were higher than the ones based on Zerlin's (1969) ITDs (open squares).[4] Zerlin (1969) also estimated TW velocities at high frequencies. These velocities were larger than the velocities at low frequencies and roughly consistent with the present ABR-based estimates.

IV. LIMITATIONS OF THE LATERALIZATION PARADIGM

A. Critical band and lateralization threshold

Despite the encouraging results of the lateralization paradigm for tone frequencies up to 1.5 kHz, no behavioral estimates of CRT could be obtained at higher frequencies. This was due to fundamental limitations in the lateralization paradigm, which are discussed in the following. In principle, a large frequency mismatch $|f_2-f_1|$ between the tones would be desirable to increase the accuracy of the ITD estimate. However, with increasing frequency mismatch, it becomes increasingly difficult to attribute a fused position (Scharf, 1972). More importantly, the lateralization threshold, i.e., the ITD for which the position of a non-centered sound object can just be distinguished from that of a centered object, increases strongly as soon as the interaural frequency mismatch exceeds a value that corresponds to the critical bandwidth for that frequency (Scharf *et al.*, 1976; Buus *et al.*, 1984). Scharf *et al.* (1976) found this bandwidth to be roughly independent of tone level and tone duration. The centering ITD, reflecting CRT disparity, needs to be larger than the corresponding lateralization threshold in order to be measurable. Therefore, in the present study, each tone pair was chosen such that the frequency mismatch between the tones did not exceed the critical bandwidth at the corresponding center frequency. The tone level of 50 dB SPL should have been comparable to the levels used by Zerlin (1969), which corresponded to an approximate loudness level of 50 phon. It was chosen as a compromise between decreasing lateralization thresholds and increasing spread of excitation with increasing tone level.

The feasibility of the measurements can, in principle, be predicted by comparing expected CRT disparities for maximally mismatched tones (tones that fall just within the same critical band) with the corresponding lateralization thresholds. As mentioned above, the CRT disparity for the mismatched tones can only be measured in terms of the ITD that is required to center the percept, if this ITD is larger than the lateralization threshold (which determines ITD sensitivity). This is discussed in the following.

B. Predicted CRT disparities from objective data

Critical bandwidths at 500, 1000, 2000, 4000, and 6000 Hz were extracted by digitizing the figures in Scharf *et al.* (1976). The obtained values were 115, 163, 310, 702, and 1080 Hz, respectively. In the next step, the frequencies of maximally mismatched tones were calculated such that the geometric means of the two frequencies were equal to 500, 1000, 2000, 4000, and 6000 Hz. At 2000 Hz, for example, the tone frequencies were 1850 and 2160 Hz. Distances between the corresponding CF places on the BM were calculated according to the Greenwood (1961) frequency-place map. Next, objective TW velocity estimates from different studies, as given in Fig. 10 of Donaldson and Ruth (1993),

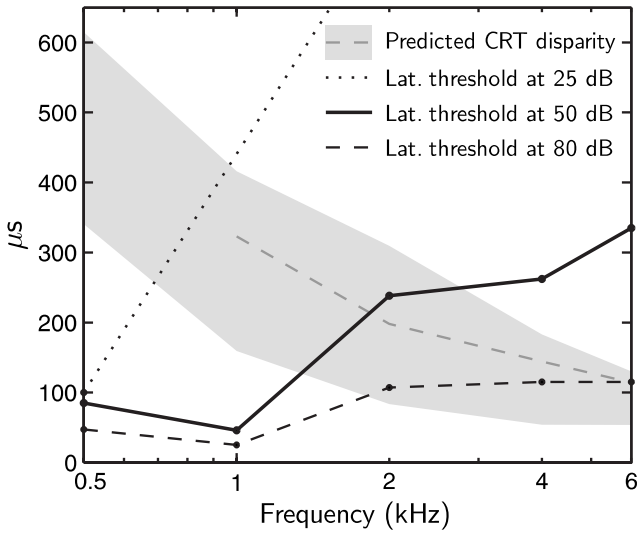


FIG. 5. Predicted CRT disparities (gray) for maximally mismatched tones as a function of the center frequency of the tones. The gray shaded area indicates CRT disparities based on the TW velocity estimates given in Donaldson and Ruth (1993). The gray dashed curve shows disparity estimates based on the TW velocity estimates obtained in the present study (curves in Fig. 4). The black curves indicate the lateralization thresholds at 25 (dotted curve), 50 (solid curve), and 80 dB SPL (black dashed curve), obtained by Scharf *et al.* (1976) and Buus *et al.* (1984).

were used to predict CRT disparities ("travel times") corresponding to these distances along the BM. The different velocity estimates yielded a range of CRT disparities, which are shown as gray shaded area in Fig. 5. The gray dashed curve indicates disparity estimates that are based on the average TW velocities obtained for the three listeners of the present study (curves in Fig. 4). As can be seen, CRT disparities for the maximally mismatched tones decrease with increasing center frequency of the tones. Furthermore, the estimates based on TW velocities obtained in the present study are consistent with those based on the TW velocities in Donaldson and Ruth (1993). Figure 5 also shows the lateralization thresholds at the different tone levels of 25 (dotted curve), 50 (solid curve), and 80 dB SPL (black dashed curve), obtained by Scharf *et al.* (1976) and Buus *et al.* (1984).[5] Up to frequencies of about 1.5 kHz, the predicted CRT disparities are larger than the corresponding lateralization thresholds for mismatched 50-dB tones (solid curve) and are therefore measurable. However, with increasing frequency, the CRT disparities fall below the lateralization thresholds and are not measurable at a tone level of 50 dB. In theory, they are measurable using tone levels of about 80 dB and higher, since lateralization thresholds are smaller at these higher levels (black dashed curve). However, this assumes that spread of excitation can be adequately limited, for example, by means of notched-noise masking. For high frequencies of about 4 kHz or higher, the predicted CRT disparities are too small to be measurable, even for tone levels of 80 dB. These predictions are consistent with the finding from the present study that ITDs could not be obtained for 50-dB tones at frequencies above 1.5 kHz.

C. Comparison with Zerlin's study

The frequency mismatches for all tone pairs used by Zerlin (1969) exceeded the critical bandwidths given by Scharf *et al.* (1976) and Buus *et al.* (1984). For the 3200|4000-Hz and 5000|6300-Hz tone pairs, the reported centering ITDs clearly fall below the corresponding lateralization thresholds in those studies. Furthermore, Scharf *et al.* (1976) emphasized the importance of controlled tone-onset phases for tone frequencies below about 2 kHz: Without controlling the onset phase, their ITD data were inconsistent and the observed lateralization thresholds became substantially larger. Zerlin, however, did not control onset phases. Hence, the validity of his results appears questionable both at low and high frequencies.

One might argue that part of the discrepancies could be due to different ramp durations. Zerlin (1969) used 2.5-ms ramps, whereas 10-ms ramps were used in the present study as well as in Scharf *et al.* (1976) and Buus *et al.* (1984). However, even with such short ramp durations (tested in pilot measurements), it was not possible to obtain consistent ITD data at high frequencies. Apart from this, the percept gained a click-like character indicating a loss of frequency specificity.

V. CONCLUSIONS

For frequencies up to 1.5 kHz, the lateralization of mismatched tones yielded estimates of CRT disparities (across remote BM places) and TW velocities that were reasonably accurate and consistent with objective estimates based on ABR measurements. However, due to intrinsic limitations of the lateralization paradigm, it was impossible to obtain behavioral estimates of CRT disparities at high frequencies.

Besides the possibility of studying aspects of the spatiotemporal BM response pattern other than CRT (e.g., response amplitude), a further step could be to investigate relations between individual estimates of CRT (disparities) and performance in other psychoacoustic tasks that have been discussed in the context of spatiotemporal processing (e.g., pitch perception and tone-in-noise detection). Here, the inclusion of hearing-impaired listeners may be crucial. Alterations in the spatiotemporal BM response, due to hearing impairment, might result in reduced performance in these tasks compared to normal-hearing listeners. The larger-than-normal across-listener variability within the hearing-impaired population may allow the study of such relations. Furthermore, it would be interesting to model the effect of CRT alterations in the framework of spatiotemporal models (cf. Carney, 1994). The lateralization method presented in this study might provide valuable information about such CRT alterations, particularly at low frequencies (below 500 Hz), where the accuracy of objective methods is limited.

ACKNOWLEDGMENTS

The authors wish to thank Dimitrios Christoforidis and Dr. James Harte for technical support and valuable scientific discussions about various aspects of this project. They also thank Brent C. Kirkwood, Torben Poulsen, Brenda L. Lonsbury-Martin, and two anonymous reviewers for their valuable comments on an earlier version of this manuscript. They are grateful to the listeners for their participation in testing. Part of this work was supported by the Danish Re-

search Foundation, the Danish Graduate school SNAK "Sense organs, neural networks, behavior, and communication," and the Oticon Foundation.

[1]The reference stimulus was not balanced in loudness since, for matched-frequency tones, equal SPLs instead of equal loudness at the two ears would give rise to a percept centered at the midline. As discussed by Durlach *et al.* (1981), the binaural system adapts to between-ear gain differences in such a way that equal-SPL tones are perceived at the midline. In this way, the correlation of auditory perception with visual and tactile perceptions is maximized.

[2]The further assumption is made that the response time at a given CF place of the BM is the same for tonal stimulation with frequencies at and below this CF. This corresponds to constant group delays, i.e., constant slopes of the BM phase response (cf. Ruggero and Rich, 1987; Robles and Ruggero, 2001). For the mismatched 800|900-Hz tone pair, for example, the TWs in response to the 800- and 900-Hz tones would reach the 900-Hz CF place at the same time. Hence, CRT differences would reflect the travel time between the 800- and 900-Hz CF places.

[3]Zerlin (1969) used the cochlear frequency-place map supplied by von Békésy (1963a) in order to derive TW velocities from the centering ITDs. In the present study, the Greenwood (1961) cochlear map was taken as a basis of all TW velocity estimates. Therefore, the TW velocities shown here were derived directly from the ITDs reported by Zerlin (1969) using the Greenwood (1961) map.

[4]These deviations cannot be attributed to the fact that Zerlin (1969) applied loudness balancing. Velocity estimates based on the ITDs obtained with loudness balancing (not shown) always fell in the same range or above the ones obtained without loudness balancing, but never below as Zerlin's.

[5]In the data by Buus *et al.* (1984) the actual tone level at 500 Hz was 59 dB SPL, not 50 dB SPL.

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.

Buus, S., Scharf, B., and Florentine, M. (**1984**). "Lateralization and frequency selectivity in normal and impaired hearing," J. Acoust. Soc. Am. **76**, 77–86.

Carney, L. H. (**1994**). "Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness," Hear. Res. **76**, 31–44.

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (**2002**). "Auditory phase opponency: A temporal model for masked detection at low frequencies," Acta. Acust. Acust. **88**, 334–347.

Clarey, J. C., Barone, P., and Imig, T. J. (**1992**). "Physiology of thalamus and cortex," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 232–334.

Deatherage, B. H., and Hirsh, I. J. (**1959**). "Auditory localization of clicks," J. Acoust. Soc. Am. **31**, 486–492.

Deng, L., and Geisler, C. D. (**1987**). "A composite auditory model for processing speech sounds," J. Acoust. Soc. Am. **82**, 2001–2012.

Don, M., and Eggermont, J. J. (**1978**). "Analysis of the click-evoked brainstem potentials in man using high-pass noise masking," J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., and Elberling, C. (**1994**). "Evaluating residual background noise in human auditory brain-stem responses," J. Acoust. Soc. Am. **96**, 2746–2757.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1993**). "Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response," J. Acoust. Soc. Am. **94**, 2135–2148.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Durlach, N. I., Thompson, C. L., and Colburn, H. S. (**1981**). "Binaural interaction in impaired listeners. A review of past research," Audiology **20**, 181–211.

Eggermont, J. J. (**1976**). "Analysis of compound action potential responses to tone bursts in the human and guinea pig cochlea," J. Acoust. Soc. Am. **60**, 1132–1139.

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.

Elberling, C., and Don, M. (**1984**). "Quality estimation of averaged auditory brainstem responses," Scand. Audiol. **13**, 187–197.

Elberling, C., and Don, M. (**2008**). "Auditory brainstem responses to a chirp stimulus designed from derived-band latencies in normal-hearing subjects," J. Acoust. Soc. Am. **124**, 3022–3037.

Elberling, C., and Wahlgreen, O. (**1985**). "Estimation of auditory brainstem response, ABR, by means of Bayesian inference," Scand. Audiol. **14**, 89–96.

Gorga, M. P., Kaminski, J. R., Beauchaine, K. A., and Jesteadt, W. (**1988**). "Auditory brainstem responses to tone bursts in normally hearing subjects," J. Speech Hear. Res. **31**, 87–97.

Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

IEC 60318-1 (**1998**). "Electroacoustics—Simulators of human head and ear—Part 1: Ear simulator for the calibration of supra-aural earphones," International Electrotechnical Commission, Geneva.

IEC 60318-2 (**1998**). "Electroacoustics—Simulators of human head and ear—Part 2: An interim acoustic coupler for the calibration of audiometric earphones in the extended high-frequency range," International Electrotechnical Commission, Geneva.

IEC 60711 (**1981**). "Occluded ear simulator for the measurement of earphones coupled to the ear by ear inserts," International Electrotechnical Commission, Geneva.

ISO 389-8 (**2004**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones," International Organization for Standardization, Geneva.

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

Joris, P. X., de Sande, B. V., Louage, D. H., and van der Heijden, M. (**2006**). "Binaural and cochlear disparities," Proc. Natl. Acad. Sci. U.S.A. **103**, 12917–12922.

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.

Magezi, D. A., and Krumbholz, K. (**2008**). "Can the binaural system extract fine-structure interaural time differences from noncorresponding frequency channels?," J. Acoust. Soc. Am. **124**, 3095–3107.

Moore, B. C. J. (**1996**). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," Ear Hear. **17**, 133–161.

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., 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.

Neely, S. T., Norton, S. J., Gorga, M. P., and Jesteadt, W. (**1988**). "Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli," J. Acoust. Soc. Am. **83**, 652–656.

Norton, S. J., and Neely, S. T. (**1987**). "Tone-burst-evoked otoacoustic emissions from normal-hearing subjects," J. Acoust. Soc. Am. **81**, 1860–1872.

Parker, D. J., and Thornton, A. R. (**1978a**). "Frequency specific components of the cochlear nerve and brainstem evoked responses of the human auditory system," Scand. Audiol. **7**, 53–60.

Parker, D. J., and Thornton, A. R. (**1978b**). "The validity of the derived cochlear nerve and brainstem evoked responses of the human auditory system," Scand. Audiol. **7**, 45–52.

Ponton, C. W., Eggermont, J. J., Coupland, S. G., and Winkelaar, R. (**1992**). "Frequency-specific maturation of the eighth nerve and brain-stem auditory pathway: Evidence from derived auditory brain-stem responses (ABRs)," J. Acoust. Soc. Am. **91**, 1576–1586.

Richter, U., and Fedtke, T. (**2005**). "Reference zero for the calibration of audiometric equipment using 'clicks' as test signals," Int. J. Audiol. **44**, 478–487.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Ruggero, M. A. (**1992**). "Physiology and coding of sound in the auditory nerve," in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer-Verlag, New York), pp. 34–93.

Ruggero, M. A. (**1994**). "Cochlear delays and traveling waves: Comments on 'Experimental look at cochlear mechanics'," Audiology **33**, 131–142.

Ruggero, M. A., and Rich, N. C. (**1987**). "Timing of spike initiation in

cochlear afferents: Dependence on site of innervation," J. Neurophysiol. **58**, 379–403.
Ruggero, M. A., and Temchin, A. N. (**2007**). "Similarity of traveling-wave delays in the hearing organs of humans and other tetrapods," J. Assoc. Res. Otolaryngol. **8**, 153–166.
Scharf, B. (**1972**). "Frequency selectivity and sound localization," in *Symposium on Hearing Theory*, edited by B. L. Cardozo (IPO, Eindhoven, Germany), pp. 115–122.
Scharf, B., Florentine, M., and Meiselman, C. (**1976**). "Critical band in auditory lateralization," Sens. Processes **1**, 109–126.
Schubert, E. D., and Elpern, B. S. (**1959**). "Psychophysical estimate of the velocity of the traveling wave," J. Acoust. Soc. Am. **31**, 990–994.
Shamma, S., and Klein, D. (**2000**). "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," J. Acoust. Soc. Am. **107**, 2631–2644.
Shamma, S. A., Shen, N. M., and Gopalaswamy, P. (**1989**). "Stereausis: Binaural processing without neural delays," J. Acoust. Soc. Am. **86**, 989–1006.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
von Békésy, G. (**1933**). "Über den Knall und die Theorie des Hörens (Clicks and the theory of hearing)," Phys. Z. **34**, 577–582.
von Békésy, G. (**1963a**). "Hearing theories and complex sounds," J. Acoust. Soc. Am. **35**, 588–601.
von Békésy, G. (**1963b**). "Three experiments concerned with pitch perception," J. Acoust. Soc. Am. **35**, 602–606.
Yost, W. A. (**1974**). "Discriminations of interaural phase differences," J. Acoust. Soc. Am. **55**, 1299–1303.
Zerlin, S. (**1969**). "Traveling-wave velocity in the human cochlea," J. Acoust. Soc. Am. **46**, 1011–1015.

Efficient coding in human auditory perception

Vivienne L. Ming[a)]
Redwood Center for Theoretical Neuroscience, University of California at Berkeley, 156 Stanley Hall, MC 3220, Berkeley, California 94720

Lori L. Holt
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

(Received 8 February 2008; revised 14 March 2009; accepted 3 June 2009)

Natural sounds possess characteristic statistical regularities. Recent research suggests that mammalian auditory processing maximizes information about these regularities in its internal representation while minimizing encoding cost [Smith, E. C. and Lewicki, M. S. (2006). Nature (London) **439**, 978–982]. Evidence for this "efficient coding hypothesis" comes largely from neurophysiology and theoretical modeling [Olshausen, B. A., and Field, D. (2004). Curr. Opin. Neurobiol. **14**, 481–487; DeWeese, M., *et al.* (2003). J. Neurosci. **23**, 7940–7949; Klein, D. J., *et al.* (2003). EURASIP J. Appl. Signal Process. **7**, 659–667]. The present research provides behavioral evidence for efficient coding in human auditory perception using six-channel noise-vocoded speech, which drastically limits spectral information and degrades recognition accuracy. Two experiments compared recognition accuracy of vocoder speech created using theoretically-motivated, efficient coding filterbanks derived from the statistical regularities of speech against recognition using standard cochleotopic (logarithmic) or linear filterbanks. Recognition of the speech created using efficient encoding filterbanks was significantly more accurate than either of the other classes. These findings suggest potential applications to cochlear implant design. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158939]

PACS number(s): 43.66.Ba, 43.71.An, 43.66.Ts, 43.72.Gy [RSN] Pages: 1312–1320

I. INTRODUCTION

Perceptual systems are limited capacity channels in that they may encode and transmit only a finite amount of information over any period of time. Mirroring the bandwidth issues that plagued early telecommunications and now electronic information exchange, perceptual systems face the seemingly intractable dilemma of coding efficiency; they must balance high-fidelity information transmission against the overall encoding cost to the system.

Although the problem of transmitting a high-fidelity, low-cost code may seem intractable, information theory states that optimally efficient codes, which carry the most information at the lowest cost, should match the statistics of the signals they represent (Shannon, 1948; MacKay, 2003). A large body of evidence from theoretical and empirical research in vision (Olshausen and Field, 1996; Sharpee *et al.*, 2006) suggests that efficiency may be central to perceptual encoding (Barlow, 1961; Atick, 1992; Simoncelli and Olshausen, 2001; Laughlin and Sejnowski, 2003).

Recent empirical and theoretical research (Rieke *et al.*, 1995; Attias and Schreiner, 1998; Lewicki, 2002; Klein *et al.*, 2003) has indicated that these principles extend to the auditory system. Smith and Lewicki (2006), for example, showed that auditory nerve response matches a theoretically-predicted efficient code for representing the diverse sounds of natural acoustic environments. In other words, the cochlear code reflects the statistics, both spectral power and higher-order (phase) statistics, of natural sounds. At a neural level, increased coding efficiency of natural signals has been repeatedly demonstrated (Rieke *et al.*, 1995; Attias and Schreiner, 1998; Vinje and Gallant, 2002; Sharpee *et al.*, 2006). Afferent fibers from the peripheral auditory system of the bullfrog better encode sounds with the spectrum of mating calls than broad-band noise (Rieke *et al.*, 1995). Similarly, neurons in the cat's inferior colliculus exhibit increased coding efficiency for narrow-band noise with "naturalistic" amplitude modulations versus "non-naturalistic" modulations (Attias and Schreiner, 1998; Escabi *et al.*, 2003). Although these results support the efficient coding hypothesis in neural auditory processing, they provide no direct insight into the extent to which observed neural coding differences have behavioral consequences in human perception.

In the present research, we examine this question directly by measuring human speech recognition under challenging perceptual circumstances. The underlying hypothesis guiding this work is that if coding efficiency has behavioral consequences, complex sounds created to match the statistics of natural sounds should have a perceptual advantage over sounds that diverge from environmental statistics. We use noise-excited vocoder speech (often used to mimic cochlear implant output in normal-hearing listeners; Shannon *et al.*, 1995), to create a challenging auditory perceptual task within which this advantage might be measured as gradations in speech intelligibility.

[a)]Author to whom correspondence should be addressed. Electronic mail: neuraltheory@gmail.com

II. VOCODING

In noise-vocoded speech, sounds are stripped of their fine spectral resolution and left with only their amplitude envelope via an algorithm similar to that used in cochlear implants (Zeng *et al.*, 2004b). A filterbank composed of limited number of filters (6 in the present work) separates speech sounds into a set of band-limited channels, with the choice of filterbank determining the frequency bands (e.g., linear versus logarithmic frequency tiling). The amplitude envelope of each channel, the slowly time-varying dynamics of the speech within that frequency band, is separated from its fine spectral detail via half-wave rectification followed by a 150 Hz low-pass filter. Each of these resulting envelopes is used to modulate the output of the Gaussian noise, giving the noise the low-frequency temporal dynamics of the original speech. Finally, each channel of modulated noise is again filtered so that its frequency range matches that of the original channel, and they are added back together, producing a single waveform. Through this process, noise-vocoded speech preserves the temporal dynamics of its limited number of frequency channels but has no spectral resolution within each channel, though some spectral information can be recovered by integrating information across channels (Nie and Zeng, 2004) allowing listeners hear some or all of the original speech steam. The change between original speech and its vocoder counterpart is illustrated in Fig. 1 where four spectrograms show how the acoustic frequency changes across time. Natural speech has complex spectral characteristics [Fig. 1(a)]. After vocoder transformation with six frequency channels, the sound loses nearly all fine spectral detail; however, the temporal envelopes of the six channels remain [Figs. 1(b)–1(d) showing three different choices of filterbanks]. Although the frequency information is severely degraded, the envelope retains many important cues for speech perception (Shannon *et al.*, 1995; Smith *et al.*, 2002). The resulting sounds can be quite difficult to understand but are clearly speech-like and, for a six-channel vocoder, reasonably intelligible with some practice.

Key to the our investigation of the consequences of efficient coding on human auditory perception, the content of each channel (and the qualities of sound produced) depends on the characteristics of the filterbank. Their experiments manipulate the form of the six filters comprising the filterbank affecting, for example, how they tile the frequency dimension, as illustrated in Fig. 2. The set of filters shown in Fig. 2(a) ("linear") simply tiles temporally-symmetric, equal-bandwidth band-pass filters linearly across the frequency dimension. The second set of filters [Fig. 2(b), "cochleotopic"] is more natural in that it mimics the near-logarithmic frequency-coding characteristics of the cochlea whereby lower frequencies are sampled with finer resolution (smaller bandwidths) than are higher frequencies (Bekesy, 1960; Greenwood, 1961). A spectrogram of vocoder speech using this filterbank is shown in Fig. 1(b). Noise-excited vocoder speech processed with a cochleotopic filterbank is generally better understood than speech processed using a linearly-tiled filterbank (Shannon *et al.*, 2003), but it is unclear whether the advantage reflects a better match to the frequency representation of the auditory system or the spectral statistics of natural sounds.

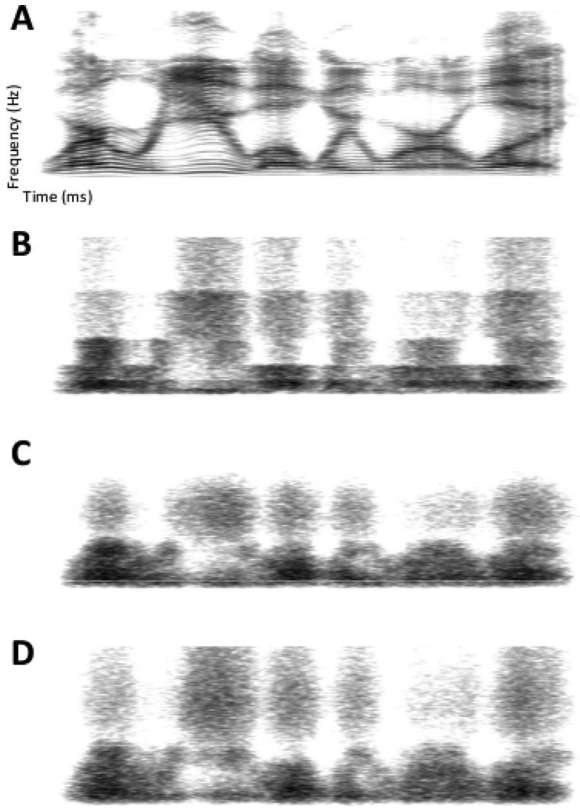


FIG. 1. Spectrograms of the utterance "Where were you while we were away?" Time unfolds along the *x*-axis, and frequency is presented along the *y*-axis and amplitude is illustrated with intensity (dark is higher amplitude). Unmodified speech (a) possesses fine spectral detail. Vocoder transformations of this utterance using three different, six-channel filterbanks [(b) cochleotopic; (c) efficient gammatone-smoothed; (d) efficient spline-smoothed, here with six channels] compress the spectral information within a channel so that only temporal modulation of six coarse frequency bands remains. The choice of filterbank determines how the spectral information is partitioned and influences the amplitude modulations extracted from each channel.

It is also possible, however, that the cochleotopic filterbank better reflects the statistical structure of speech acoustics, and that it is this quality which drives the improved performance. It is not possible to distinguish these two hypotheses comparing perception using cochleotopic versus linear filterbanks as the cochleotopic set matches both the biology and the sound statistics better than does the linear set. To address this confound, we will use a machine learning algorithm to analyze the statistics of speech acoustics and design a new filterbank to reflect the statistical structure.

III. EFFICIENT CODING HYPOTHESIS

According to the efficient coding hypothesis, perception should be optimally adapted to the statistics of natural signals such that they carry the most information at the least cost. Therefore, perceptual performance should be best when sensory codes match the statistics of environmental stimuli. To test this prediction, we use a computational model of efficient auditory coding (Smith and Lewicki, 2006) to opti-

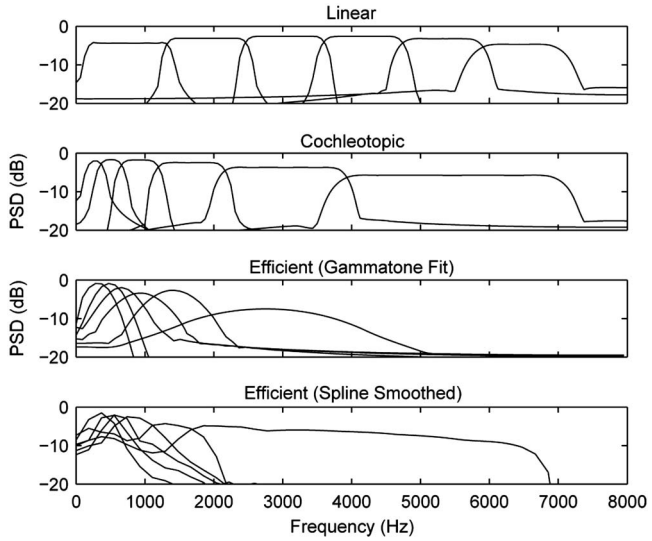


FIG. 2. The power spectra for all six filters from the four different filterbanks are shown. The top row shows the frequency tiling of the "linear" filterbank. The linear frequency tiling of the set can be clearly seen to the right. The cochleotopic frequency tiling is shown in row B. The next two rows show the learned "efficient" filters smoothed either by gammatone fitting or spline-smoothing.

mize a set of functions with respect to the information carried by the large TIMIT speech corpus training set (Garofolo *et al.*, 1990). This model allows an explicit prediction of the dimensions of perceptual sensitivity. In it, sound, $x(t)$, is generated by a linear superposition of a set of functions, $\varphi_1, \ldots, \varphi_M$, which can be positioned arbitrarily and independently in time. The mathematical form of the representation with additive noise is

$$x(t) = \sum_{m=1}^{M} \sum_{i=1}^{n_m} s_i^m \phi_m(t - \tau_i^m) + \varepsilon(t), \tag{1}$$

where τ_i^m and s_i^m are the temporal position and coefficient of the ith instance of kernel ϕ_m, respectively. The notation n_m indicates the number of instances of ϕ_m, which need not be the same across kernel functions. The kernel functions are not restricted in form or length, and both the kernel shapes and their lengths were adapted to optimize coding efficiency; in the results below, the kernels take on a variety of shapes and range in length from 10 to 100 ms. This provides a mathematical description of sound waveforms that has sufficient flexibility to encode arbitrary acoustic signals and encompass a broad range of potential auditory codes.

The key theoretical abstraction of the model is that the acoustic signal can be encoded most efficiently by decomposing it in terms of discrete acoustic elements, each of which has a precise amplitude and temporal position. This also yields a code that is time-relative and does not depend on artificial blocking of the signal (Smith and Lewicki, 2005a, 2005b). One interpretation of each analog τ_i^m, s_i^m pair is that it represents a local population of (binary) auditory nerve spikes firing probabilistically in proportion to the underlying analog value.

To code speech sounds efficiently, we need to determine both the optimal values of τ_i^m and s_i^m (*encoding*) and the optimal kernel functions ϕ_m (*learning*). From Eq. (1), coding efficiency can be defined approximately as the number of "spikes" (nonzero coefficient values) required to achieve a desired level of precision, which is defined by the variance of the additive noise $\varepsilon(t)$. This assumes that the goal of coding is to represent the entire acoustic signal and that coding efficiency is most closely related to the number of spikes in the code. Other definitions are possible within this framework, but this definition has the advantage of starting from a minimal set of assumptions.

Although the generative form of the model is linear, in other words the signal is a linear function of the representation, inferring the optimal representation for a signal is highly non-linear and computationally complex. Here we compute the values of τ_i^m and s_i^m for a given signal by using a matching pursuit algorithm (Mallat and Zhang, 1993), which iteratively approximates the input signal and has been shown to yield highly efficient representations for a broad range of sounds (Smith and Lewicki, 2005a, 2005b). In matching pursuit, the current residual signal (initialized as the original sound) is projected onto the dictionary of kernel functions. The projection with the largest inner product is subtracted out, and its coefficient and time recorded. For the results reported here, the encoding halts when s_i^m falls below a pre-set "spiking" threshold.

The goal of learning in the efficient coding model is to find a set of functions for which the coefficients are maximally efficient (i.e., carry the most information about the sound at the lowest cost) with respect to the given training data. We can rewrite Eq. (1) in probabilistic form in which we assume that the noise is Gaussian and the prior probability of a spike, $p(s)$, is sparse (i.e., comes from a probability distribution which produces very few nonzero values). The kernel functions are optimized by performing gradient ascent on the approximate log-data probability,

$$\begin{aligned}\frac{\partial}{\partial \phi_m}\log(p(x|\phi)) &= \frac{\partial}{\partial \phi_m}\log(p(x|\phi,\hat{s})) + \log(p(\hat{s})) \\ &= \frac{1}{2\sigma_\varepsilon}\frac{\partial}{\partial \phi_m}\left[x - \sum_{m=1}^{M}\sum_{i=1}^{n_m} \hat{s}_i^m [x - \hat{x}_{\tau_i^m}]\right]^2 \\ &= \frac{1}{\sigma_\varepsilon}\sum_i \hat{s}_j^m [x - \hat{x}]_{\tau_i^m}, \end{aligned} \tag{2}$$

where $[x-\hat{x}]_{\tau_i^m}$ indicates the residual error over the extent of kernel ϕ_m at position τ_i^m. The estimated kernel gradient is thus a weighted average of the residual error. For training here, we restrict the set to six functions, which were initialized as 100-sample Gaussian noise, and the spiking threshold (minimum value of s_i^m) was set at 0. Filters were derived from the resulting kernel functions using reverse correlation (Smith and Lewicki, 2006).

The filterbanks shown in Fig. 2(c) ["efficient (gammatone fit)"] and Fig. 2(d) ["efficient (spline smoothed)"] were learned using the efficient coding model (Smith and Lewicki, 2006) using two different smoothing methods to regularize the functions. These filters represent an optimal code for the statistical properties of the speech database when only six channels are available. The frequency tiling from the efficient coding model [Figs. 2(c) and 2(d)] is much more biased

to the low frequencies than the cochleotopic model [Fig. 2(b)]. This can also be seen in the difference between the vocoder spectrograms in Fig. 1, which shows the difference between speech transformed using the cochleotopic vocoder [Fig. 1(b)], the gammantone-smoothed "efficient" vocoder [Fig. 1(c)] and the spline-smoothed efficient vocoder [Fig. 1(d)]. Moreover, the form of the efficient functions is not fixed; it combines both gammatone-like filters in the lower frequencies with broadly-tuned, symmetric filters at the higher frequencies.

If there is efficient coding in auditory processing, one would expect perceptual performance to align with the efficient filters. The distinction between the filters in Figs. 2(c) and 2(d) and the cochleotopic filters of Fig. 2(b) may seem counter-intuitive given that for larger filterbanks (30 +filters), the optimal filterbank very closely match both individual structure and population statistics of filters estimated from single-unit recordings of auditory nerve fibers (Smith and Lewicki, 2006). This correspondence is lost when many fewer are channels available, as with noise-vocoded speech or cochlear implants. Spectral resolution is limited and the resulting filter characteristics change from the cochleotopic filterbank typically thought to best characterize the frequency processing of the cochlea. For six-channel noise-vocoded speech, the optimally efficient code and the cochlear code diverge, providing a means to dissociate them experimentally.

If efficient coding carries perceptual benefits, speech recognition accuracy should be greatest for noise-vocoded speech created with efficient filters because these filters best characterize the statistics of speech within the limited capacity of six channels, preserving the available information. We explicitly tested this prediction by having adult human listeners transcribe noise-vocoded speech produced with the filterbanks shown in Fig. 2.

IV. METHODS

Following the approach of previous vocoder experiments (Shannon *et al.*, 1995), we measured speech intelligibility in two distinct tasks: identifying words in continuous speech (sentences, Experiment 1) and identifying phonemes from non-word utterances (non-words, Experiment 2).

For Experiment 1, there were 168 distinct English sentences (42 sentences/condition), each spoken by a different native-English speaker (TIMIT corpus: Garofolo *et al.*, 1990). The assignment of sentences to filtering conditions was counter-balanced such that, across participants, each sentence was presented in each of the four conditions but no sentence or speaker was repeated for an individual participant. Sentences ranged in length from 8–16 words (approximately 1–6 s) for a total of 1564 words. The sentences used in the experiment were drawn from the TIMIT testing set and were distinct from those used in the training of the computational model that produced the efficient filterbanks.

Four stimulus conditions were created by synthesizing vocoder versions of each item using one of three filterbanks plus using the original, unmodified speech as a control. The filterbanks were composed of six finite impulse response filters, with the number of filters chosen to produce sufficiently challenging stimuli so as to avoid ceiling effects. The linear and cochleotopic filterbanks were composed of six Hanning-window band-pass filters. The filters were tiled across 0–7 kHz with either linear or cochleotopic (logarithmic) placement (see Fig. 2).

For the efficient filterbanks, the "raw" (unsmoothed) filters comprising them were identical in both experiments. Kernel functions were trained on the 4956 sentences from the TIMIT training set. Training involved encoding a batch of 100 full sentences on each iteration and then updating the kernel functions based on the gradient estimated from the batch. Training continued until the set reach convergence, about 10,000 iterations. Filters were then derived from the functions. These filters were then smoothed to regularize the filters, removing residual noise from the learning algorithm. In experiment 1, the efficient filters were fitted with gammatone functions, a parametrized approximation of the learned functions composed of sine wave modulated by a gamma function [Fig. 2(c)].

Sixteen listeners participated in Experiment 1. Participants were college-age native-English speakers from Carnegie Mellon University with no reported or obvious speaking or hearing disorders. Participants received undergraduate Psychology course credit for participation. Seated in individual sound-attenuated booths, participants listened to each stimulus and typed what they heard. In Experiment 1, participants were told that some of the sentences may be difficult to understand, but a response must be made on each trial. They were allowed to hear each sentence only once. In neither experiment was there a pre-exposure or training period for the participants with the vocoder speech. Each participant listened to 62 stimuli from each condition (186 total). In both experiments, order of stimulus presentation was randomly permuted for each participant.

The ALVIN experiment-control software (Hillenbrand and Gayvert, 2005) was used for stimulus presentation and data collection. Acoustic presentation was under the control of Tucker Davis Technologies (Alachua, FL) System II hardware; stimuli were converted from digital to analog, amplified, and presented dichotically over linear headphones (Beyer DT-150, Berlin, Germany) at approximately 70 dB SPL(A).

The stimuli for Experiment 2 consisted of non-word syllables spoken in isolation (Shannon *et al.*, 1999). Each stimulus was composed of two distinct utterances of the same syllable separated by 500 ms of silence. The stimuli consisted of both vowel-consonant-vowel (VCV) syllables such as "aba" and consonant-vowel (CV) syllables such as "bi" representing the full range of combination described in Shannon *et al.*, 1999. Stimuli were drawn from a corpus of ten speakers (five male, five female) to include 46 unique syllables from each speaker. Both the sentence and non-word stimuli were sampled at 16 kHz with 16-bit resolution.

In Experiment 2, we used spline-smoothing to regularize the raw filters [Fig. 2(d)], which offers a less biased estimate of the underlying function than gammatone fitting. As can be seen in Fig. 2(d), the gammatone fitting rounded the power spectra of the filters, even cutting off the high-end of the

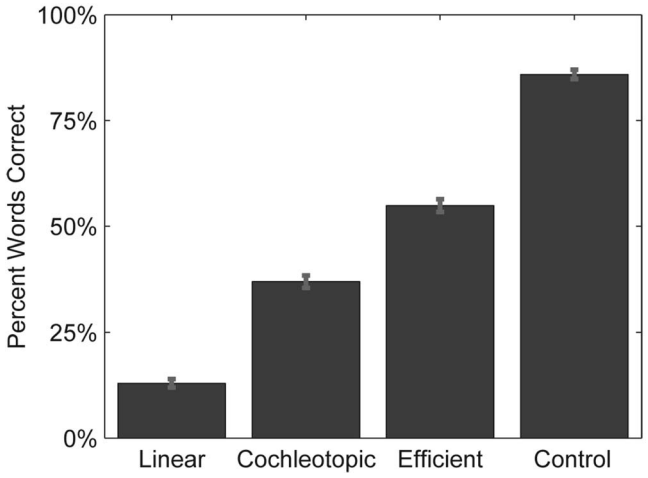


FIG. 3. Average speech intelligibility across participants as a function of condition. The control condition is unaltered speech. Error bars show 95% confidence interval of the mean.

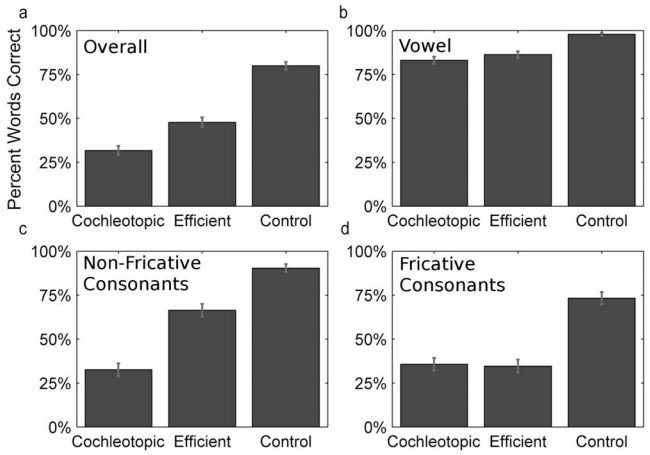


FIG. 4. Speech intelligibility for non-word stimuli. Each subfigure shows the accuracy of non-word speech identification across three conditions: cochleotopic filtering, efficient filtering, or control speech. The subplots show mean performance for (a) whole items, (b) vowels, (c) non-fricative consonants, and (d) fricatives. Error bars show the 95% confidence interval of the mean.

highest-frequency filter. The spline smoothed set better-preserved these features. In both cases, however, the exact same set of raw filters, derived from learning the statistics of the TIMIT training set, were used. Only the choice of smoothing techniques changed.

Fourteen CMU undergraduate students, none of whom had participated in Experiment 1, participated in Experiment 2. They were instructed that they would hear a speech sound repeated twice and they were to type what they heard. For simplicity, the linear condition was removed to focus on the comparison of interest, cochleotopic versus efficient. All other methodological details were identical to Experiment 1.

V. RESULTS

Experiment 1. Each participant's performance was evaluated by comparing every word transcribed against the set of words in the original sentence. The typed responses were hand coded as "correct" if a match could be found. Minor alterations, such as adding -s or -ed, were not scored as correct but homophones were (e.g., sea versus see). Compound words (e.g., houseboat, bittersweet, sleepwalk, etc.) were treated as multiple words. Data from three participants were coded independently by two coders; the two sets of scores were highly correlated ($r>0.99$).

Each word in the original sentence was treated as independent (see below for further discussion of this issue) and the overall probability of a correct response was computed for each filterbank condition. As shown in Fig. 3, although intelligibility was greatly degraded for vocoded speech relative to original speech, intelligibility of vocoded speech was highly influenced by filterbank choice. The mean percent correct across participants for each condition were $13\pm4\%$, $37\pm6\%$, $56\pm6\%$, and $86\pm4\%$ (mean $\pm$ 95% CI) for the linear, cochleotopic, efficient, and control conditions, respectively (planned Bonferroni-corrected pairwise comparisons for all results were highly significant, $p<0.001$). 15 of the 16 participants were significantly more accurate at transcribing speech synthesized with efficient versus cochleotopic filters ($p<0.01$). Across participants, performance differed by an average of 19% (efficient and cochleotopic filters, 56% versus 37%, respectively; $p<0.0001$).

Participants' performance was little influenced by various lexical variables. A small correlation was found between word frequency (averaged from Kucera and Francis, 1967; Brown, 1984) and participant accuracy ($r=0.23$). There was no significant relationship of word type (noun/verb/other) and intelligibility ($p=0.33$). There was a small, but significant, increase in accuracy for words near the end of a sentence versus those occurring near the beginning or middle (5.4%, $p<0.0001$). There was no significant effect of either speaker or participant gender ($p=0.13$ and 0.33, respectively). Comparing performance between the first and second half of the experiment shows a significant increase in percent correct for both the cochleotopic (+4.2%; p-value$=0.007$) and efficient conditions (+4.9% increase; p-value<0.0001). There was, however, no significant interaction between early-late training and filter condition (p-value$=0.3$).

Experiment 2. For the non-word task, performance was measured by hand coding the response to each vowel and consonant in an item as correct or "incorrect." The entire item was coded correct if all of its phonetic elements were correct. As shown in Fig. 4(a), overall accuracy was slightly lower than Experiment 1: $31\pm8\%$, $49\pm9\%$, and $80\pm7\%$, for the cochleotopic, efficient, and control conditions, respectively (mean $\pm$ 95% CI). Performance with vocoded speech remained best with efficient filters (18% greater than cochleotopic filters; $p<0.0001$).

Performance gains differed based on the acoustic properties of the speech sounds. Participants were very accurate at identifying vowels [Fig. 4(b)], nearing ceiling in the control condition (98% correct) and achieving 83% and 86% correct in the efficient and cochleotopic conditions, respectively. The small difference between vocoded-speech conditions was not reliable ($p=0.10$). Relative to vowels, accuracy was much lower for consonants across all conditions (34% for cochleotopic, 51% for efficient, and 82% for control).

Performance in the efficient condition improved significantly between the first and second half of the experiment (+15%; $p<0.0001$) and a more modest improvement in performance across experiment halves was observed for the cochleotopic condition (+5.5% increase; $p=0.03$). There was no reliable difference in this learning effect across the different filter types ($p=0.1$).

Results of the theoretical modeling of Smith and Lewicki (2006) suggest that noise-like, ambient natural sounds represent a dimension in natural sound statistics distinct from acoustic transients. Based on this distinction, we separated the consonant stimuli into two classes, fricative and non-fricative consonants. Reanalyzing the data based on these classes produced very different results. As shown in Fig. 4(c), performance on non-fricatives (e.g., stop consonants, nasals and glides, such as /b/, /n/, and /l/) differed markedly, 33% and 66% ($p<0.0001$) in the cochleotopic and efficient conditions, respectively. In contrast, performance with fricatives [Fig. 4(d)] was quite low in all conditions (35%, 36%, and 73% for cochleotopic, efficient, and control) and it did not differ significantly between the vocoded-speech conditions ($p=0.4$).

There was a small, but unreliable, trend for better overall performance with the VCV-stimuli compared to CV (53% versus 50%; $p=0.10$). As with the sentences, there was no significant effect of either speaker ($p=0.08$) or participant gender ($p=0.19$).

VI. DISCUSSION

The results are consistent with a marked perceptual benefit of efficient coding. Speech recognition accuracy was greatest for noise-vocoded speech created with efficient filters. Given that these filters were created such that they best characterized the regularities of speech within the limits of six channels, it appears that the acoustic dimensions conveyed by the efficient filters provided listeners with more information with which to identify words and phonemes. The effect was dramatic. In the linear condition of the continuous speech task, participants typically understood only one word per sentence, consistent with previous findings that linear frequency mapping degrades perceptual performance (Fu and Shannon, 1999). On average, participants understood nearly twice as many words synthesized with the cochleotopic filters. In accordance with the efficient coding hypothesis, though, the efficient representation further increased performance, with participants identifying more than half the words in each sentence, four times more than the linear condition.

Even with a completely different stimulus set and nonsense syllables in Experiment 2, the efficient filters produced greater accuracy than cochleotopic filters for the non-word stimuli. The non-word task also revealed that the benefit of the efficient filters stemmed largely from benefits in non-fricative consonant intelligibility. Nearly the entire increase in performance between the filter conditions came from those items. However, it should be noted that it is possible that there is a ceiling effect confounding any effect on vowel recognition.

The pronounced increase in speech intelligibility in both experiments strongly suggests that participants are sensitive to the dimensions of speech acoustics predicted by the efficient coding hypothesis, but it is not yet clear what drives these improvements. One simple possibility is that the frequency tiling learned by the efficient filters maximizes its channel capacity (i.e., each channel carries an equal amount of information). To test this, we computed the variance of the envelope output from each channel for each filterbank across all stimuli used in the continuous speech task. Ideally, assuming independent, equal capacity channels, the variance across all six channels should be equal, implying that each channel is carrying equal amounts of independent information. If the variance in any channel is low relative to the others, then the total capacity of the system is underutilized.

The efficient filterbanks (using either gammatone- or spline-smoothing) make fuller use of their channel capacity than do the standard filterbanks; the higher-frequency filters in the linear and cochleotopic filterbanks are relatively unused, forcing all of the information about the sound to be carried by only two to four channels. The pressure to fully utilize channel capacity explains why the frequency tiling of the efficient filterbanks (as shown in Fig. 2) appears biased to the low frequencies; these filters more equitably carry information about speech.

Whereas this analysis makes use of signal statistics to differentiate the information carried across filters in the filterbank, it is also possible to consider what drives listeners' sensitivity to the dimensions of speech acoustics predicted by the efficient coding hypothesis from a psychoacoustic perspective. The articulation index (AI) has long been used to evaluate the importance of different frequency bands for speech recognition using perceptual measures (Fletcher and Steinberg, 1929; French and Steinberg, 1947; Studebaker *et al.*, 1987; ANSI, 1997). It is possible that perceptual weighting across frequency of importance for speech is similar to the information-theoretic optimum. Figure 5(a) illustrates the band importance values for normal speech calculated using the AI (ANSI, 1997; 1/3 octave). For each filter in each filterbank, we computed the filter response at each frequency band. The importance weighting for each filter can be estimated as the dot product of the filters' frequency response (power spectrum) and the band importance values shown in Fig. 5(a). This provides a score for each filter that indicates its importance to intelligibility. As is clear from Fig. 5(b), the efficient filterbank more evenly distributes band importance across its constituent filters (see Table I), with a significantly higher mean and lower variance compared to the cochleotopic and linear filterbanks. Thus, like the analysis based on signal statistics, analysis based on speech psychophysics (via the AI) also indicates that the efficient coding filters make fuller use of the channel capacity.

The results of the non-word task suggest that intelligibility of non-fricative consonants, in particular, drives the increase in intelligibility. In the case of stop consonants, which make up 47% of the non-fricative consonants, the distinguishing characteristics are not purely functions of frequency resolution but reflect higher-order, temporal structure of sound (e.g., voice onset time, Lisker and Abramson, 1964;

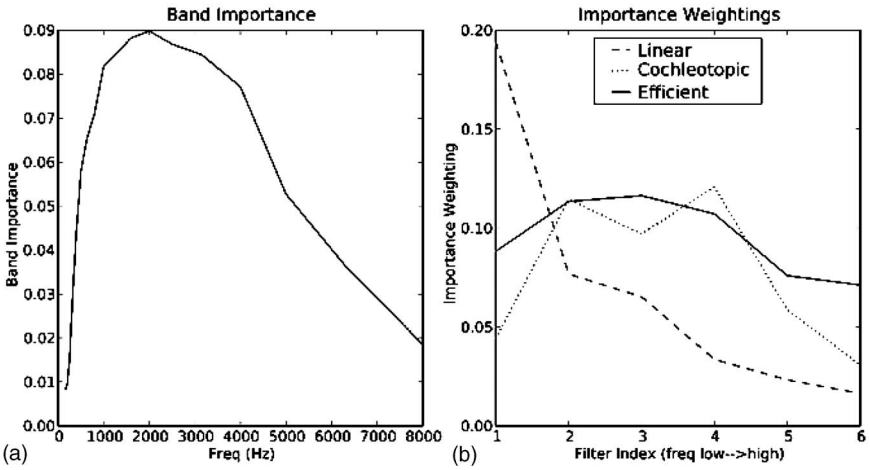


FIG. 5. (a) Band importance values for normal speech calculated using the AI (ANSI, 1997; 1/3 octave) indicates the contribution of different frequency bands to speech perception. The lowest and highest-frequency bands contribute much less to speech perception than the range from 1 to 3 kHz. (b) Importance weightings for each filter are computed as the dot product of the band importance values in (a) with the normalized filter's response at each frequency band and indicates how the filter pools information from each band.

Steinschneider *et al.*, 1999). The temporal asymmetry of the efficient filters may play a significant role here. This would agree with the finding that significant linguistic information is available in the temporal as well as spectral contents of speech (Shannon *et al.*, 1995). Increased sensitivity to temporal features like onsets suggests an influence of higher-order sound structure on the dimensions of perceptual sensitivity; second-order characteristics (i.e., the power spectrum) are not particularly sensitive to transient, edge-like signal structure (Field, 1987).

Our decision to treat each word in the continuous speech task as an independent measure greatly simplified analysis, but it is not realistic. It is known that syntactic and semantic contexts provided cues for sentence-level processing (Boothroyd and Nittrouer, 1988; Bronkhorst *et al.*, 1993; Gibson, 1998). Although the sentences in the TIMIT corpus have little predictability, participants clearly exhibited evidence of sentence-level processing. Correct responses were more likely to occur in pairs than would be expected at random and they were more likely to occur near the end of a sentence. In the non-word task, though, participants were required only to produce a single syllable. It is unlikely that cognitive load or context effects influenced the results, although there may be contrast effects (Lotto and Aravamudhan, 2004). Nonetheless, listeners' performance in the linear and cochleotopic conditions was significantly lower than that observed by Shannon *et al.* (1995); this is possibly a result of our choice of test materials (Zeng *et al.*, 2005). Syllables processed with an efficient filterbank were significantly better recognized than those processed with a cochleotopic filterbank.

In Experiment 1, we chose to smooth the learned efficient filters by fitting them with gammatones, allowing us to preserve the basic form of the learned filters using a model of auditory filters common to the auditory modeling literature (Patterson *et al.*, 1988; Slaney, 1993; Lyon, 1996). In Experiment 2, we aimed to address some limitations of the gammatone fitting by switching to spline-smoothing. For example, with the highest-frequency efficient filter, the best-fit gammatone truncated the highest frequencies whereas the best-fit spline did not. The smoothing in each experiment was performed on the same set of the raw filters produced by the computational model. Thus, the filters in both conditions reflect the statistics of the training set. The only difference between them was the smoothing technique. Spline smoothing, having many more free parameters, preserved more of the true spectral shape of the optimized filters. It should be noted that smoothing therefore introduced some differences between the experiments; specifically, whether (Experiment 1) or not (Experiment 2) the highest-frequency (5–7 kHz) information was incorporated into the sixth channel of the vocoder. The consistent patterning of results across the two experiments suggests that this difference did not have a significant impact on the results or their interpretation.

A possible criticism of this research is the use of classic linguistic categories (vowel, fricatives, etc.) that presuppose a particular structure to speech. Phonemic categories were used for stimulus categories as a rough approximation of the natural structure of speech acoustics. Their use here should not be taken as an assumption that phonemes represent a fundamental of acoustic or cognitive representation. Rather, their role here is only as a loose stand-in for dimensions of perceptual variability. Given the efficient coding hypothesis, ultimately it may be preferable to identify these dimensions using efficient coding algorithms similar to those used to train the filters here.

TABLE I. The mean and variance of the importance weightings across the filterbanks computed using the articulation index (AI). Higher means and smaller variances indicate greater and stronger correspondence between a given filterbank and the AI.

Filter type	Mean (CI)	Variance (CI)
Linear	0.0682 (0.0138–0.1225)	0.0072 (0.0019–0.0190)
Cochleotopic	0.0776 (0.0462–0.1089)	0.0024 (0.0007–0.0063)
Efficient	0.0953 (0.0792–0.1114)	0.0006 (0.0001–0.0016)

VII. CONCLUSION

As a unique compliment to the growing body of empirical and theoretical literature on efficient coding in neural systems (Barlow, 1961, Olshausen and Field, 2004), these results provide direct behavioral evidence for the role of coding efficiency as a general principle in human auditory perception. Yet to be addressed is the relevance of coding efficiency to higher-level representation. Methodologies that

further meld theoretical-experimental designs to test listeners' sensitivity to the statistics of complex everyday sounds will be important for future exploration of efficiency in auditory processing. For example, by adapting our generative model to the acoustics of different spoken languages, we can generate acoustic stimuli directly from the model that reflect the differing low-level statistics of sounds from different languages absent any high-level, linguistic content.

Experiments with normal-hearing participants and vocoder speech previously have been useful in modeling cochlear implant hearing (Shannon *et al.*, 2003). It is possible that consideration of the computational principles of efficient coding may provide insight in cochlear implant applications.

Cochlear implants are by far the most successful neuroprosthetic devices and the only one in standard clinical use. They employ direct, electrical stimulation of auditory nerve fibers along the tonotopic axis of the cochlea to restore some degree of hearing in individuals with peripheral hearing loss, even in cases of profound deafness (Wilson *et al.*, 1991; Zeng *et al.*, 2004a). Unfortunately, despite 20 years of research and wide clinical application, speech perception in cochlear implant users remains highly variable and often quite degraded (Shannon *et al.*, 2003).

In general, the present results emphasize the significance of perceptual theory in neuroprosthetic design. Mimicking the surface features of a perceptual system, as in the cochleotopic filtering scheme, may not provide as much leverage as understanding a perceptual system's computational principles. The efficient coding hypothesis claims a specific computational principle: optimally efficient codes which carry the most information at the lowest cost should match the statistics of the signals they represent. Here, we found that the set of filters derived from a computational model trained to optimally extract the statistics of a corpus of speech passed more information normal-hearing participants could use to identify speech in sentence and non-word contexts than did more standard filtering schemes (linear, cochleotopic).

Of course, there remain many open questions for this line of research, and the specific algorithm used here may not necessarily produce the same dramatic improvements in speech intelligibility outside that laboratory. For example, the algorithm used to learn the efficient filters has not taken issues of electrode placement into account, which are essential in optimizing cochlear implant performance. Perceptual performance is known to degrade sharply as the mismatch between the frequency of the input channel and tonotopy of the cochlea increases (Shannon *et al.*, 1998; Fu and Shannon, 2002; Baskent and Shannon, 2005). Although it is beyond the scope of the current work, exploring issues regarding adaptation by cochlear implant users to changes in place-frequency mapping (Rosen *et al.*, 1999; Fu *et al.*, 2002b) would be an important extension of this research. Alternatively, expanding the efficient coding algorithm to incorporate constraints relevant to cochlear implants, such as frequency-place mappings, might be even more valuable.

We have shown in this study that recognition performance for perceptually degraded, vocoder speech improves when the vocoder filterbank matches the statistical structure of speech acoustics. A machine learning algorithm based on the efficient coding hypothesis was used to adapt the filterbank to speech structure. In two experiments, using stimuli from two unrelated speech corpora, recognition accuracy was superior for speech generated by the adapted filterbanks than recognition using cochleotopic filterbanks. The adapted filterbanks show greater spectral resolution in the frequency range of speech formants, which plays a large role in the higher recognition accuracy.

ACKNOWLEDGMENTS

V.L.M. and L.L.H. designed the experiments, planned the analyses and wrote the paper. V.L.M. implemented the filters, produced the stimuli, and ran the experiments and analyses. He was supported by NSF IGERT training Grant No. DGE-9987588. This research was supported by Grant No. 2R01DC004674-04A2 from the National Institutes of Health and Grant No. 0345773 from the National Science Foundation. We would like to thank Christi Gomez for help collecting and coding data.

ANSI (**1997**). "Methods for calculation of the speech intelligibility index," American National Standards Institute, New York.

Atick, J. J. (**1992**). "Could information-theory provide an ecological theory of sensory processing?" Networks **3**, 213–251.

Attias, H., and Schreiner, C. E. (**1998**). "Coding of naturalistic stimuli by auditory midbrain neurons," in *Advances in Neural Information Processing Systems*, edited by M. I. Jordan, M. J. Kearns, and S. A. Solla (MIT, Cambridge, MA), Vol. **10**.

Barlow, H. B. (**1961**). "Possible principles underlying the transformation of sensory messages," in *Sensory Communication*, edited by W. A. Rosenbluth (MIT, Cambridge, MA), pp. 217–234.

Baskent, D. E., and Shannon, R. V. (**2005**). "Interactions between cochlear implant electrode insertion depth and frequency-place mapping," J. Acoust. Soc. Am. **117**, 1405–1416.

Bekesy, G. (**1960**). *Experiments in Hearing* (McGraw-Hill, New York), pp. 503–509.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Bronkhorst, A. W., Bosman, A. J., and Smoorenburg, G. F. (**1993**). "A model for context effects in speech recognition," J. Acoust. Soc. Am. **93**, 499–509.

Brown, G. D. A. (**1984**). "A frequency count of 190,000 words in the London Lund corpus of English conversation," Behav. Res. Methods Instrum. **16**, 502–532.

Escabi, M. A., Miller, L. M., Read, H. L., and Schreiner, C. (**2003**). "Naturalistic auditory contrast improves spectrotemporal coding in the cat inferior colliculus," J. Neurosci. **23**, 11489–11504.

Field, D. (**1987**). "Relations between the statistics of natural images and the response profiles of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394.

Fletcher, H., and Steinberg, J. C. (**1929**). "Articulation testing methods," Bell Syst. Tech. J. **8**, 806–854.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Fu, Q.-J., and Shannon, R. V. (**1999**). "Recognition of spectrally degraded and frequency-shifted vowels in acoustic and electric hearing," J. Acoust. Soc. Am. **105**, 1889–1900.

Fu, Q.-J., and Shannon, R. V. (**2002**). "Frequency mapping in cochlear implants," Ear Hear. **23**, 339–348.

Fu, Q.-J., Shannon, R. V., and Galvin, J. (**2002**). "Perceptual learning following changes in the frequency-to-electrode assignment with the Nucleus-22 cochlear implant," J. Acoust. Soc. Am. **112**, 1664–1674.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., Dahlgren, N. L., and Zue, V. (**1990**). TIMIT Acoustic-Phonetic Continuous Speech Corpus.

Gibson, E. (**1998**). "Linguistic complexity: Locality of syntactic dependen-

cies," Cognition **68**, 1–76.
Greenwood, D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.
Hillenbrand, M. J., and Gayvert, T. R. (**2005**). "Open source software for experiment design and control," J. Speech Lang. Hear. Res. **48**, 45–60.
Klein, D. J., Konig, P., and Kording, K. P. (**2003**). "Sparse spectrotemporal coding of sounds," EURASIP J. Appl. Signal Process. **2003**(7), 659–667.
Kucera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English* (Brown University Press, Providence, RI).
Laughlin, S. B., and Sejnowski, T. J. (**2003**). "Communication in neuronal networks," Science **301**, 1870–1874.
Lewicki, M. S. (**2002**). "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Lotto, A. J., and Aravamudhan, R. (**2004**). "Phonetic context effects in cochlear implant listeners," in Meeting of the Acoustical Society of America, San Diego, CA.
Lyon, R. F. (**1996**). "The all-pole gammatone filter and auditory models," Forum Acusticum, Antwerp, Belgium.
MacKay, D. J. C. (**2003**). *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge).
Mallat, S. G., and Zhang, Z. (**1993**). "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process. **41**, 3397–3415.
Nie, K. B., and Zeng, F. G. (**2004**). "Speech perception with temporal envelope cues: Study with an artificial neural network and principal component analysis," The 26th IEEE EMBS Conference, San Francisco.
Olshausen, B. A., and Field, D. (**1996**). "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," Nature (London) **381**, 607–609.
Olshausen, B. A., and Field, D. (**2004**). "Sparse coding of sensory inputs," Curr. Opin. Neurobiol. **14**, 481–487.
Patterson, R., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "Implementing a gammatone filter bank. SVOS final report: The auditory filter bank," Report No. 2341, Medical Research Council Applied Psychology Unit, University of Cambridge Medical School.
Rieke, F., Bodnar, D. A., and Bialek, W. (**1995**). "Naturalistic stimuli increase the rate and efficiency of information transmission by primary auditory neurons," Proc. R. Soc. London, Ser. B **262**, 259–265.
Rosen, S., Faulkner, A., and Wilkinson, L. (**1999**). "Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants," J. Acoust. Soc. Am. **106**, 3629–3636.
Shannon, C. E. (**1948**). "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 379–423, 623–656.
Shannon, R. V., Fu, Q.-J., Galvin, J., and Friessen, L. (**2003**). "Speech perception with cochlear implants," in *Cochlear Implants: Auditory Prostheses and Electric Hearing*, Springer Handbook of Auditory Research, edited by F.-G. Zeng, A. N. Popper, and R. R. Fay (Springer, New York), pp. 334–376.
Shannon, R. V., Jensvold, A., Padilla, M., Robert, M., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Shannon, R. V., Zeng, F.-G., and Wygonski, J. (**1998**). "Speech recognition with altered spectral distribution of envelope cues," J. Acoust. Soc. Am. **104**, 2467–2476.
Sharpee, T., Sugihara, H., Kurgansky, A., Rebrik, S., Stryker, M. P., and Miller, K. D. (**2006**). "Adaptive filtering enhances information transmission in visual cortex," Nature (London) **439**, 936–942.
Simoncelli, E., and Olshausen, B. (**2001**). "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1216.
Slaney, M. (**1993**). "An efficient implementation of the Patterson–Holdsworth auditory filter bank," Technical Report No. 35, Apple Computer, Cupertino, CA.
Smith, E. C., and Lewicki, M. S. (**2005a**). "Efficient coding of time-relative structure using spikes," Neural Comput. **17**, 19–45.
Smith, E. C., and Lewicki, M. S. (**2005b**). "Learning efficient auditory codes using spikes predicts cochlear filters," in *Advances in Neural Information Processing Systems*, edited by L. K. Saul, Y. Weiss, and L. Bottou (MIT, Cambridge, MA), Vol. **17**, pp. 1289–1296.
Smith, E. C., and Lewicki, M. S. (**2006**). "Efficient auditory coding," Nature (London) **439**, 978–982.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Steinschneider, M., Volkov, I. O., Noh, M. D., Garell, P. C., and Howard, M. A. (**1999**). "Temporal Encoding of the Voice Onset Time Phonetic Parameter by Field Potentials Recorded Directly From Human Auditory Cortex," J. Neurophysiol. **82**, 2346–2357.
Studebaker, G. A., Pavlovic, C. V., and Sherbecoe, R. L. (**1987**). "A frequency importance function for continuous discourse," J. Acoust. Soc. Am. **81**, 1130–1138.
Vinje, W. E., and Gallant, J. L. (**2002**). "Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1," J. Neurosci. **22**, 2904–2915.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Zeng, F.-G., Nie, K., Liu, S., Stickney, G., DelRio, E., Kong, Y.-Y., and Chen, H. (**2004a**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.
Zeng, F.-G., Popper, A. N., and Fay, R. R. (**2004b**). *Cochlear Implants: Auditory Prostheses and Electric Hearing, Springer Handbook of Auditory Research* (Springer, New York).

Pitch discrimination by ferrets for simple and complex sounds

Kerry M. M. Walker[a)]
Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, University of Oxford, Oxfordshire OX1 3PT, United Kingdom

Jan W. H. Schnupp
Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, University of Oxford, Oxfordshire OX1 3PT, United Kingdom and Department of Robotics, Brain, and Cognitive Sciences, Italian Institute of Technology, Via Morego 30, 16163 Genova, Italy

Sheelah M. B. Hart-Schnupp, Andrew J. King, and Jennifer K. Bizley
Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, University of Oxford, Oxfordshire OX1 3PT, United Kingdom

(Received 14 August 2008; revised 22 June 2009; accepted 23 June 2009)

Although many studies have examined the performance of animals in detecting a frequency change in a sequence of tones, few have measured animals' discrimination of the fundamental frequency (F0) of complex, naturalistic stimuli. Additionally, it is not yet clear if animals perceive the pitch of complex sounds along a continuous, low-to-high scale. Here, four ferrets (*Mustela putorius*) were trained on a two-alternative forced choice task to discriminate sounds that were higher or lower in F0 than a reference sound using pure tones and artificial vowels as stimuli. Average Weber fractions for ferrets on this task varied from ~20% to 80% across references (200–1200 Hz), and these fractions were similar for pure tones and vowels. These thresholds are approximately ten times higher than those typically reported for other mammals on frequency change detection tasks that use go/no-go designs. Naive human listeners outperformed ferrets on the present task, but they showed similar effects of stimulus type and reference F0. These results suggest that while non-human animals can be trained to label complex sounds as high or low in pitch, this task may be much more difficult for animals than simply detecting a frequency change.
 [DOI: 10.1121/1.3179676]

PACS number(s): 43.66.Gf, 43.66.Hg, 43.66.Fe, 43.80.Lb [BCM] Pages: 1321–1335

I. INTRODUCTION

To interpret a vocal call appropriately, animals must perceive a number of attributes of complex sounds, including loudness, timbre, and pitch. In the context of human speech, pitch perception is particularly pertinent to identifying the speaker (Smith *et al.*, 2005) and inferring their emotional state (Johnson, 1990) and it is thought to play similarly important roles in vocal communication among non-human primates (Koda and Masataka, 2002; Kojima *et al.*, 2003), songbirds (Nelson, 1989), and even frogs (Capranica, 1966). But although pitch is clearly a fundamental perceptual attribute of sound, few studies to date have examined how well animals can judge the physical correlate of pitch in naturalistic sounds—that is, the fundamental frequency (F0) of complex periodic stimuli.

The physiological mechanisms underlying pitch perception for complex sounds are also incompletely understood. In the case of pure tones, perceived pitch correlates directly with frequency, and different pure tones maximally stimulate different parts of the cochlea. These widely appreciated and fundamental facts can lead to the tempting, but probably incorrect, conclusion that discriminations of the pitch of periodic sounds are the result of a simple frequency discrimination problem that can be easily solved through place coding in the tonotopically organized ascending auditory pathway. However, most naturally occurring sounds are spectrally complex, (i.e., they carry acoustic energy at large numbers of frequencies). Consequently, the relationship between frequency content and perceived pitch is not straightforward since there is no one-to-one relation between the fundamental frequency of complex sounds and the resulting pattern of activation within a tonotopic map. Instead, pitch perception may rely on a combination of spectral template matching and "time domain" information about the periodicity of sounds carried by temporally phase-locked neural discharges early in the auditory pathway (Moore, 2003). These neural representations of sound periodicity need not necessarily result in an anatomically ordered topographic representation analogous to tonotopic maps. Some studies have provided evidence in favor of periodotopic arrangements, which might serve as pitch maps at the level of the inferior colliculus in cats (Schreiner and Langner, 1988) and the primary auditory cortex in gerbils (Schulze and Langner, 1997; Schulze *et al.*, 2002). Others have failed to find any clear topographic arrangement of periodicity preference in the auditory cortex of either ferrets (Nelken *et al.*, 2008) or marmosets (Bendor and Wang, 2005). Therefore, it remains unclear whether or to

a)Author to whom correspondence should be addressed. Electronic mail: kerry@oxfordhearing.com

what extent common physiological mechanisms are used to encode the fundamental frequency of pure tones and more naturalistic sounds.

For human listeners, the percept of pitch height, unlike that of timbre, can be described along a monotonic scale, from low to high, and one desirable feature of a topographic pitch map is that it might provide a simple neural mechanism by which listeners could judge the direction of a pitch change. Alternatively, specialized pitch "shift detection" operations have been proposed to underlie human listeners' ability to identify the direction of subtle F0 changes (Demany and Ramos, 2005). Before the relevance of such models to mammalian neurophysiology can be examined, it is first necessary to demonstrate that non-human animals do experience changes in the fundamental frequency of complex sounds as changes along an ordered, low-to-high pitch scale.

Numerous studies have shown that diverse species of animals are sensitive to changes in the F0 of tones or more complex stimuli. Most commonly, this is tested using a "go/no-go" task in which sounds are presented continuously and animals are conditioned to make a response if, and only if, the sound changes. This paradigm has been used to measure pure tone frequency discrimination thresholds in macaques (Sinnott *et al.*, 1985; Pfingst, 1993), chinchillas (Nelson and Kiester, 1978; Shofner, 2000), cats (Elliott *et al.*, 1960; Witte and Kipke, 2005), rats (Syka *et al.*, 1996; Talwar and Gerstein, 1998, 1999), mice (Ehret, 1975), guinea pigs (Heffner *et al.*, 1971), budgerigars (Dooling and Saunders, 1975), and electric fish (Marvit and Crawford, 2000). This experimental approach has also been used to demonstrate that electric fish, songbirds, and chinchillas can detect changes in the F0 of complex sounds (Marvit and Crawford, 2000; Shofner, 2000; Dooling *et al.*, 2002). What is not known, however, is whether the animals engaged in these tasks perceive the change in F0 as a change in pitch height, and whether they can discriminate upward from downward pitch changes. Recent studies of human listeners have emphasized that the ability to detect changes in the pitch of ongoing sounds does not necessarily imply that the listener can order these sounds along a pitch scale (Semal and Demany, 2006). While adults and children with cochlear implants are severely impaired on tasks that require discrimination of the direction of pitch changes (Fujita and Ito, 1999; Gfeller *et al.*, 2002; Pressnitzer *et al.*, 2005; Vongpaisal *et al.*, 2006), children with cochlear implants have been shown to exhibit much finer pitch acuity in tasks that do not require them to report whether the pitch in a sequence of complex sounds has increased or decreased (Vongpaisal *et al.*, 2006).

It is much more difficult to train animals on the types of psychophysical tasks required to address whether they order pitch height. Only a handful of studies of this kind have been undertaken so far, and none have used complex sounds. Go/no-go tasks, in which animals are trained to respond to a particular direction of frequency change (i.e., frequency increases or decreases) within a sequence of tones, have been used to demonstrate that primates (D'Amato, 1988; Brosch *et al.*, 2004) and birds (Page *et al.*, 1989; Cynx, 1995) do have the capacity to make relative pure tone frequency judgments. However, as noted above, the neural mechanisms of pure tone frequency discrimination may be quite different from those used to judge the pitch of a complex sound. Investigations of pitch discrimination in humans tend to favor two-alternative forced choice (2AFC) designs, in which the subject must not only detect pitch changes but also identify the change in each trial as a relative increase or decrease in pitch from a standard reference value (Wier *et al.*, 1977). This type of task is not commonly used in animal psychoacoustics largely because of the difficulty in training animals on 2AFC auditory discrimination tasks (Burdick, 1979). However, these differences in task design make it problematic to compare the difference limens of humans and animals directly since frequency discrimination performance has been shown to vary with task design (Nelson and Kiester, 1978; Burdick, 1980; Talwar and Gerstein, 1999).

Establishing a successful regime for training animals on 2AFC auditory discrimination tasks would also make novel investigations of the neurophysiological correlates of pitch perception possible. In the field of vision and somatosensory research, "neurometric" studies, which combine electrophysiology and 2AFC discrimination tasks, have identified the neural events that are likely to give rise to perceptual judgments (e.g., Liu and Newsome, 2005; de Lafuente and Romo, 2006). To use similar methodologies in hearing science, the authors require an animal model, like the ferret, that is suited to both psychophysical tasks and electrophysiological recordings.

In the present study, the authors trained ferrets (*Mustela putorius*) to discriminate the direction of pitch changes on a positively conditioned 2AFC discrimination task. In each trial, two artificial vowel sounds were presented in succession, and the ferret had to indicate, by choosing a water spout either to the left or to the right, whether the second sound was higher or lower in pitch than the first. The anatomy and physiology of the ferret auditory system are well documented, and there are a number of observations that suggest that the ferret may be a suitable species in which to study the role of pitch cues in vocalization processing. Ferrets are highly sensitive to low-frequency pure tones (Kelly *et al.*, 1986), there is evidence that they are sensitive to the harmonic fusion of tone complexes (Kalluri *et al.*, 2008), and many of their vocalizations contain low-frequency energy and are strongly periodic, within the pitch range. The responses of ferret auditory cortical neurons encode information about the F0 of artificial vowels (Bizley *et al.*, 2009) and support the discrimination of human speech sounds (Mesgarani *et al.*, 2008). However, the ability of ferrets to discriminate the pitch of complex sounds has not been previously measured.

The authors measured pitch discrimination in ferrets using both pure tones and artificial vowels in an identical paradigm, so that difference limens and Weber fractions could be directly compared. The artificial vowel stimuli were sufficiently complex to exhibit key features that are commonly found in vocalization sounds used in human and animal communication, yet simple enough to be described by a small number of numerical parameters. Finally, the authors also measured the discrimination performance of naive human listeners for comparison, using a similar paradigm.

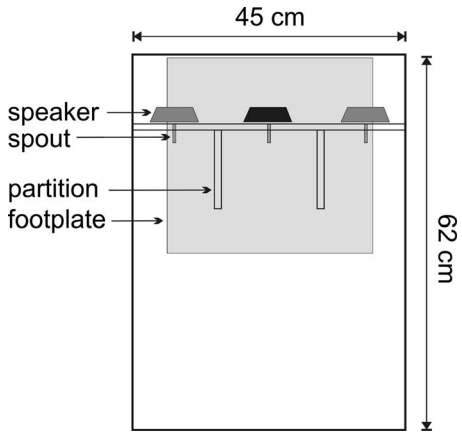


FIG. 1. Schematic of the training apparatus, shown from above. Ferrets made behavioral responses and received water rewards from the three stainless steel water spouts while standing on the aluminum footplate. In the first training stage, sounds were presented from the central loudspeaker (black) as well as the two peripheral loudspeakers (gray). In later training stages and during testing, sounds were presented from the central loudspeaker only.

II. METHODS

A. Subjects

Four adult pigmented ferrets (one male) were trained in this study. Ferrets were housed either singly (males) or in groups of two or three (females), with free access to high-protein food pellets and water bottles. On the day before training, water bottles were removed from the ferret's home cages, and they were replaced on the last day of a training run. Training runs lasted for 5 days or less, with at least 2 days between each run. On training days, ferrets received drinking water as positive reinforcement while performing a sound discrimination task. Water consumption during training was measured and was supplemented as wet food in home cages at the end of the day to ensure that each ferret received at least 70 ml of water per kilogram of bodyweight daily. Regular otoscopic and typanometry examinations were carried out to ensure that both ears of the animals were clean and healthy. All experimental procedures were approved by the local ethical review committee and were carried out under license from the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986.

B. Training apparatus

Ferrets were trained to discriminate sounds in custom-built testing chambers constructed from double glazing units that incorporated a sound-insulating vacuum. The ceiling of the chambers was covered in sound-absorbing foam. The testing chambers were approximately 45 cm wide, 62 cm long, and 54 cm high (Fig. 1). A Plexiglass wall, 12 cm from the back of the box, separated the animal from the electronics and tubing of the apparatus. Three metallic water spouts were mounted on the Plexiglass wall, one centrally positioned "start spout" and two "response spouts" positioned to the left and right. An aluminum 32×34 cm^2 footplate covered the floor below the water spouts. When the ferret licked a water spout while standing on the footplate, a small change in voltage between the steel spout and the aluminum plate resulted, allowing the authors to register the animal's licking responses using electronic circuitry, as described by Hayar *et al.* (2006). Sound stimuli as well as acoustic feedback signals were delivered via three loudspeakers (Visaton FRS 8), which were mounted above the spouts. These speakers produce a flat response (± 2 dB) from 200 Hz to 20 kHz, with an uncorrected 20 dB drop-off from 200 to 20 Hz. Plexiglass partitions, 13 cm long and 15 cm high, were positioned between the central spout and each of the peripheral spouts to increase the perceived cost involved in the ferrets' response, as initial testing had indicated that ferrets were less likely to pay careful attention to the acoustical cues if hopping between response spouts required essentially zero time or effort.

The behavioral task, data acquisition, and stimulus generation were all automated using custom software running on personal computers, which communicated with TDT RM1 real-time signal processors (Tucker-Davis Technologies, Alachua, FL).

C. Training

Ferrets were trained on a 2AFC discrimination task using drinking water as a positive reinforcer. To assist learning, animals were trained in several stages of 2AFC tasks, and each ferret was advanced to the next stage when they reached a criterion of at least 85% correct on three consecutive sessions. Each of the five training stages and the final testing stage are described in detail below. Ferrets ran two training sessions daily within each 5-day training "run" and were typically completed between 60–150 trials per session.

During a pre-training stage, ferrets learned to lick the spouts in the testing chamber for a water reward. The animal was required to maintain contact with a spout for about 1 s before receiving a water reward from the spout. During pre-training, ferrets were required to alternate between the central and peripheral spouts in order to receive a water reward, but no sound stimuli were presented at this time. The amount of water used to reward a single response varied across animals, but for all animals the water reward presented from the peripheral response spouts (0.3–0.5 ml per trial) was larger than the water reward presented at the central start spout (0.1–0.2 ml per trial).

1. Training stage 1

In the first training stage, the ferrets performed a pure tone frequency discrimination task but with a localization cue to assist. At the start of each trial, a "reference signal" consisting of a continuously repeated pure tone (5 kHz, 100 ms duration, 150 ms inter-tone interval, and 0.5 ms cosine ramped rise/fall) was presented from the central speaker until the ferret activated the central spout to start the trial [Fig. 2(a)]. This activated a small water reinforcer at the central spout and was followed by a second "target" pulse of pure tones (300 ms duration, 150 ms inter-tone interval, and 0.5 ms cosine ramped rise/fall) that differed in frequency from the initial reference tones. The target pulse continued to play until the ferret activated one of the peripheral spouts. If the target frequency was higher than the reference, the animal

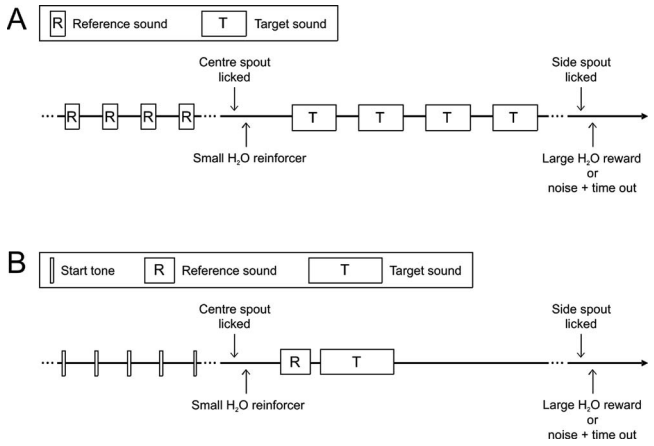


FIG. 2. Discrimination trial schematics. (A) In the first three stages of training, a reference stimulus was repeated until the ferret initiated a trial by activating the central spout. Then, a target stimulus of a different frequency was presented repeatedly until the animal responded. The ferrets' task was to indicate whether the fundamental frequency of the target sound was higher or lower than that of the reference by activating the right or left peripheral spouts, respectively. In training stages 1 and 2, the stimuli were pure tones, and in stage 3 they were click trains. (B) In training stages 4 and 5 and in testing, a ready signal consisting of a repeating 5000 Hz tone pip was presented to indicate to the animal that a trial could be started by licking the central start spout. The reference and target sounds were each then presented only once, in quick succession. As above, the ferrets' task was to indicate the direction of pitch change between the reference and target sounds by responding at the correct peripheral spout. The sounds to be discriminated were click trains in training stage 4 and an artificial vowel sound (formant-filtered click trains) in the final training and testing stages.

was required to move to the right response spout to receive a water reward. For frequencies lower than the reference, the animal was required to move to the left. Responses at the central spout were ignored, and responses at the incorrect side spout resulted in a 500-ms broadband noise at the central speaker that indicated the onset of a 10–12 s timeout. In all training and testing conditions, the central spout remained unresponsive for 2 s following a correct response to provide the animal with a quiet period between trials in which it could drink the water reward.

To make this task easy, the authors provided an additional spatial cue during the initial training phases: The target tones were presented only from the loudspeaker above the correct water spout for a given trial. Thus, if the target was higher in frequency than the reference, then the target sounds were presented from the right peripheral speaker and right spout responses were rewarded with water.

If an animal responded incorrectly on a given trial, the same stimuli were presented on the next trial. Such "correction trials" continued until the animal responded correctly. For all training stages, the F0 of the reference sound was fixed during a session, and the target sound varied between two F0s across trials—one that was at least an octave higher than the reference and another that was at least an octave lower than the reference.

2. Training stage 2

Once animals learned to perform the above task, the localization cue was removed so that the ferret was left only with frequency as an acoustic cue to the correct response. For this and all further tasks, all stimuli were presented from the middle speaker above the start spout only.

3. Training stage 3

Once criterion had been reached on the pure tone frequency discrimination task, the tones were replaced with click trains, where the click rate (and hence the F0) of the reference (300–500 Hz) differed from the target sounds by at least an octave. Human listeners perceive this stimulus as a rich buzzing sound with a pitch corresponding to the click rate. Thus, the reference and target sounds were both broadband but differed in F0.

4. Training stage 4

In stage 4 of training, the previous pitch discrimination task was modified so that the sounds to be discriminated were presented only once per trial, rather than being repeated until the animal made its response [Fig. 2(b)]. Before each trial, a series of tone pips (5 kHz tone, 20 ms duration, 0.5 ms rise/fall time, and 200 ms inter-tone interval) was presented as a "ready signal" to let the ferret know that the central spout could be triggered to start the trial. When the ferret triggered the central spout, a small water reward was administered from the spout on 10% of trials, chosen at random. On the remaining 90% of the trials, rewards were only given from the peripheral spouts for correct responses to the test sounds. The test sounds consisted of two consecutive click trains presented from the central speaker: A 200-ms-long reference click train (5 ms rise/fall time, with a click rate of approximately 400 Hz) followed by a 50-ms-long inter-stimulus interval of silence and then a 500-ms-long target click train (5 ms rise/fall time, with click rates at least one octave away from that of the reference). As before, in order to receive a reward, the ferret was required to activate the right peripheral spout if the target was higher in F0 than the reference sound and to activate the left peripheral spout if the target pitch was lower in F0 than the reference. Incorrect responses were again negatively reinforced with a noise and timeout of 10–12 s and were followed by correction trials. If an animal failed to make a response within 15 s of the onset of the target stimulus, the trial was reset without timeout or water reinforcement, forcing the animal to restart the trial.

5. Training stage 5

In this stage, click train stimuli were replaced with artificial vowel sounds that were created using custom software based on an algorithm adapted from Malcom Slaney's Auditory Toolbox (http://cobweb.ecn.purdue.edu/~malcolm/interval/1998-010/). The vowel sounds were composed of click trains that were bandpass filtered to add "formants" centered at 430, 2132, 3070, and 4100 Hz (Fig. 3) and then given an envelope with 5 ms rise and fall times. These formants correspond to the first four formants of the English vowel /*i*/ (as in "pill"). The overall spectral distribution of the vowel was largely determined by the position of the formants and was thus similar across all F0s tested. The click rates of the target and reference sounds in this task were similar to those used in stage 4. The switch from unfiltered

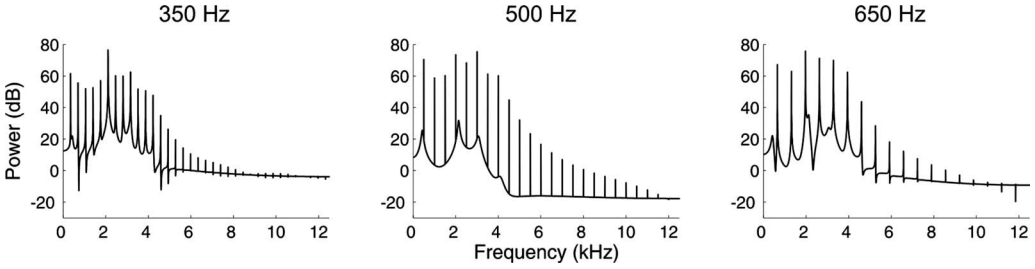


FIG. 3. Power spectra of the artificial vowel stimulus used in this experiment (|*i*|), shown with a fundamental frequency of 350 (left panel), 500 (center panel), and 650 (right panel) Hz.

click trains to artificial vowels did not cause any transient drop in the animals' performance, indicating that the animals were able to generalize the pitch task very rapidly across the two types of complex sounds. Across trials, the sound levels of the reference and target vowels were varied independently. Each was set to one of ten possible values, spanning a range of approximately 15 dB, and chosen at random with a uniform distribution. Sound levels were calibrated using a B&K sound level meter and free-field, $\frac{1}{2}$ in. microphone type 4191 (Brüel & Kjær, Nærum, Denmark). The average sound level of the artificial vowel was approximately 80 dB SPL (sound pressure level) (±3 dB across different F0s). The ±7.5 dB random variation in sound level was introduced to ensure that ferrets were not inadvertently provided with relative level difference cues within particular spectral bands across the reference and target sounds. Once animals had reached criterion on a given session of this task, the two target F0s used were occasionally jittered (±150 Hz) across additional sessions. By randomizing levels and target pitches, the authors encouraged the animals to follow pitch cues, rather than mapping other acoustical features of the training targets onto the left and right spouts.

D. Pitch discrimination testing

Once ferrets performed at ≥85% correct in at least three consecutive sessions of training stage 5, the authors switched the ferrets from training to testing. The testing task was very similar to stage 5 of training, except that 30 different target F0s were now presented within a given session, rather than just two target F0s. Animals that had reached criterion on training stage 5 generalized to the 30-target task well, often performing at their best from the very first testing session. Within each weekly testing run, composed of ten individual testing sessions, the F0 of the reference was held constant. Each weekly run was initiated with a session resembling training stage 5, wherein only two target sounds with fixed F0, one octave above and one octave below the reference, were used. Animals were required to perform at ≥85% on this task before progressing to the "variable target" condition, and most animals reached this criterion in one to two testing sessions. In the variable target condition, the F0 of the target sound was varied from trial to trial across 30 values. Fifteen of these targets had a higher F0 than the reference and 15 had a lower F0, with a sampling density of 12 steps per octave. To help ensure that the animals were still attending to the pitch of the sounds when the task was very difficult, if the ferret responded incorrectly when the target was within 5/12 octaves of the reference, then the correction trial was presented at the most extreme value in the range in the appropriate direction. Pitch discrimination testing was repeated across sessions using a constant reference F0 for typically 300–600 trials per reference. The animal was then restarted on stage 5 of training using a different reference F0 at the beginning of the next testing run (i.e., the following week). This procedure was repeated for a number of references at 200–1200 Hz.

E. Human psychophysics

Five adult humans (two male, ages 24–40 years) were tested on a similar pitch discrimination task, using both the artificial vowel and tones as stimuli. Human psychophysical procedures were carried out under the guidelines of the Central University Research Ethics Committee of the University of Oxford. The authors attempted to make the pitch discrimination task performed by human subjects as similar as possible to the task performed by the ferrets. The stimuli were presented from the same central speaker of the testing chamber in which ferrets were trained, and the sounds were presented with the same random level variation across trials as in the animals' task. The human subjects, being too large to fit in the ferret testing chamber, listened to the sounds through the opened lid, and initiated trials and responded by pressing keys on a keyboard positioned near the chamber. Upon initiation of each trial, a reference and a target sound, identical to those used with the ferrets, were presented. However, humans performed considerably better than ferrets at these discrimination tasks, so the range of periodicities sampled around each reference was adjusted based on pilot data (not shown) and was also occasionally readjusted for individual subjects based on their performance on the first 50–100 trials. Instead of water rewards, subjects received feedback on a computer monitor after each trial, and incorrect choices resulted in a broadband noise and timeout of 2–4 s. Each subject completed 250–350 trials for each of four reference F0s in both artificial vowel and tone versions of the discrimination task. The reference F0 was constant within any one testing run, tone and vowel trials were presented in blocks, and the order of reference periodicities presented was pseudo-randomized across subjects. Only one subject (H5) was musically trained, and none were given extensive training on the task prior to testing.

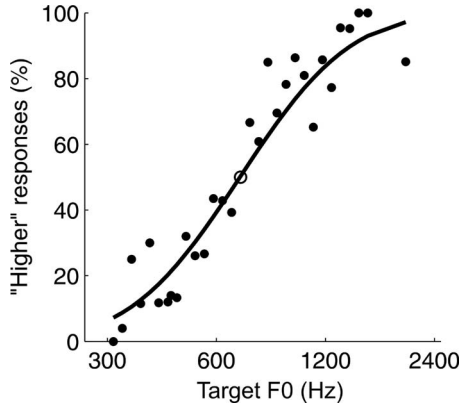


FIG. 4. Performance of one ferret (subject F1) on the discrimination task when the reference vowel had a fundamental frequency of 700 Hz. The percentage of right spout choices is plotted as a function of the fundamental frequency of the target sound (black dots), and the reference F0 is indicated by an open circle plotted at 50% choice probability. The psychometric curve, a fitted cumulative Gaussian distribution function, is also shown (black line).

F. Data analysis

Correction trials were excluded from the data analysis, as were trials on any testing session in which the subject scored less than 60% correct overall. Each animal completed at least 300 trials for each reference F0, and in many cases the total number of trials was closer to 1000. Data were pooled across testing sessions that used a common reference F0.

Figure 4 plots the performance of one ferret on trials in which the reference F0 (open circle) was 700 Hz. As expected, the ferret responded correctly more often when the F0 of the reference and target vowel were further apart. The functions relating the proportion of trials on which the animal responded at the right spout to the log of the target F0 were sigmoidal in shape and approximated a cumulative Gaussian distribution function (black dots, Fig. 4). Therefore, psychometric curves were estimated from ferrets' raw choice probabilities by fitting cumulative Gaussian distributions using probit generalized linear models (black line, Fig. 4). The difference limens for each reference F0 were calculated from the fitted psychometric curves as half of the distance between the F0 at which the right spout was chosen on 69.15% of trials and the F0 at which the right spout was chosen on 30.85% of trials. A threshold of 69.15% was chosen because this level of performance on the authors' 2-alternative discrimination task is equivalent to a d' of 1 (Wickens, 2002), making their results comparable to previous studies of frequency discrimination in non-human animals on go/no-go tasks. Weber fractions were calculated as the ratio of the difference limen for pitch divided by the F0 of the reference sound.

Pearson correlations and analyses of variance (ANOVAs) were used throughout to test whether differences in performance were significant and whether performance was related to parameters such as stimulus type (artificial vowel or pure tone) and reference pitch. A significance level (alpha) of 0.05 was used as a criterion for null hypothesis rejection, and where multiple comparisons were carried out across a number of subjects and/or reference values, significance levels were Bonferroni corrected.

III. RESULTS

A. Ferrets' discrimination of the pitch of artificial vowels

All four ferrets reached criterion on the five training stages after 1–2 weeks of training per stage. Figure 5 shows the psychometric functions obtained from all four ferrets for each reference F0 tested. The pitch acuity of ferrets is indicated by the steepness of the psychometric curves and can also be expressed as the minimum F0 difference required for the animal to reach the authors' criterion of 69.15% (i.e., the difference limen for pitch). Figure 6(a) illustrates how ferrets' difference limens for the pitch of the artificial vowel depend on the F0 of the reference. Pearson correlations showed that these difference limens increased significantly with the F0 of the reference vowel ($r=0.86$, $p<0.001$). Weber fractions were derived by normalizing each difference limen by the corresponding reference F0 [Fig. 6(b)]. These normalized measures of pitch acuity showed no linear change across the range of reference F0s tested ($r=-0.01$, $p=0.951$). Nevertheless, Weber fractions did differ significantly across reference F0s (one-way ANOVA; $F(12,25)=3.98$, $p=0.002$). *Post hoc* tests indicated that the Weber fraction measured using a reference of 200 Hz was significantly higher than the Weber fractions for references of 300–417 Hz and 700 Hz (Tukey's Honestly Significant Difference test; $p<0.05$).

In Fig. 5, some of the fitted psychometric curves do not pass through the reference F0 at the 50% choice probability point. These small shifts of the psychometric curve are partly attributable to statistical errors in the fit of the psychometric function but might also indicate a response bias; i.e., when uncertain about the correct response, the ferret did not make "higher" (right) or "lower" (left) responses with exactly equal probability but exhibited a small preference for either higher or lower responses. The authors quantified this bias in each psychometric curve as the distance, in hertz, between the 50% right choice probability point of the psychometric function and the reference F0. In this calculation, higher response biases are represented as negative values and lower biases as positive. In Fig. 7, the biases are plotted as a percentage of the reference F0, and these values are compared across ferrets and across references. Some animals were more likely to show response biases than others, and most of the significant biases on this task (see Fig. 7 caption) were toward the right spout (i.e., higher pitch).

To ensure that the ferrets were making pitch judgments, the intensities of the reference and target sounds were independently and randomly varied across trials. To confirm that the sound level of the target sounds did not affect discrimination performance, psychometric functions were fitted independently for target sounds at each intensity level for each reference F0. A two-way ANOVA was then carried out on the slopes of the fitted psychometric functions (expressed as percent choice probability per octave) using ferret identity and

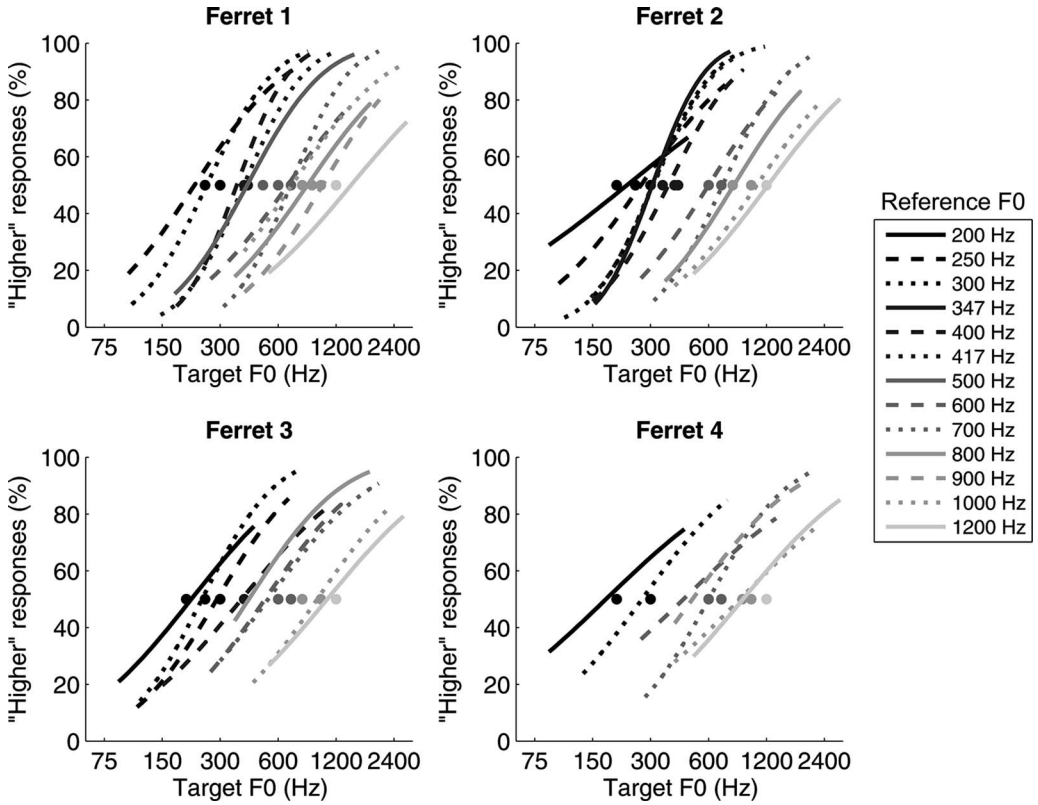


FIG. 5. Psychometric curves describe the performance of four ferrets trained to discriminate the pitch of an artificial vowel. Each panel shows the psychometric functions for a single animal, and the style and grayscale of each psychometric curve correspond to the fundamental frequency of the reference vowel (see legend). The reference F0 for each curve is also indicated by a grayscale circle at 50% choice probability.

target intensity as predictor variables. The authors found no significant effect of target sound intensity [$F(19,436)=1.09$, $p=0.359$], nor did the intensity significantly interact with the performance of individual ferrets [$F(57,436)=1.15$, $p=0.221$]. Furthermore, responses on individual trials were not predicted by the intensity differences between the reference and target vowel [two-way ANOVA; $F(28,569)=1.25$, $p=0.174$] nor by the interaction between this intensity difference and ferret identity [$F(84,569)=1.23$, $p=0.090$]. Therefore, across the 15 dB range tested, ferrets' performance on the pitch discrimination task did not depend on sound intensity.

Ferrets could adopt at least three strategies to solve the pitch discrimination task. They could compare the relative pitches of the target and reference presented on each trial, as human listeners report doing. They could also build up an internal "memory" of the reference during training at the beginning of the week and judge each target as high or low relative to this internalized reference. While this internal reference would be reinforced by the presentation of that same reference on each trial throughout the testing run, performance may persist without it. Finally, the ferrets may not compare the targets to a single reference at all but instead compare a given target to internalized "low" and "high" target templates, for instance, the high and low pitches presented during training at the beginning of a weekly run. The authors attempted to assess whether animals were relying on the reference presented at each trial by carrying out a week of "variable reference" testing sessions in which the refer-

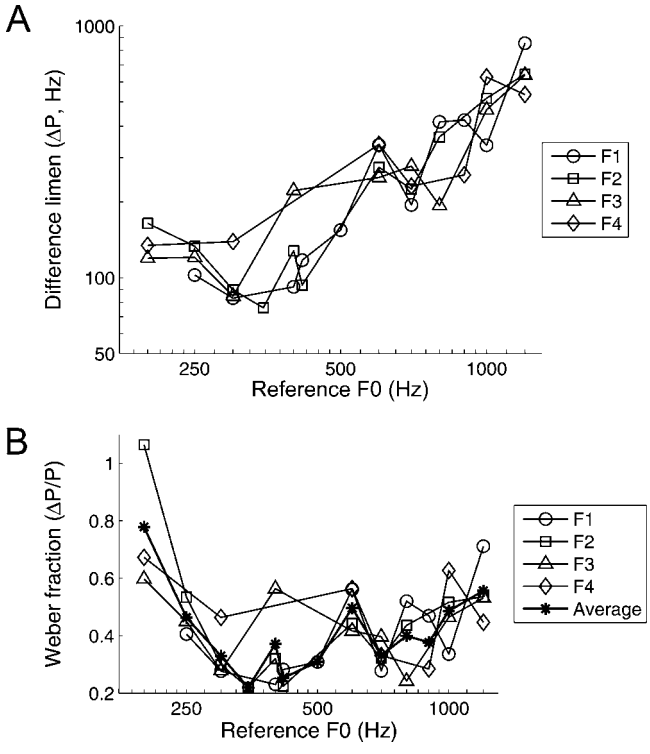


FIG. 6. (A) Difference limens for pitch of four ferrets across a range of reference F0s. Data for each ferret are displayed with a unique symbol and connected with solid lines. (B) Weber fractions for the four ferrets (F1–F4) across a range of reference F0s, calculated using the difference limens in (A). The average Weber fractions across all ferrets are shown by the asterisks.

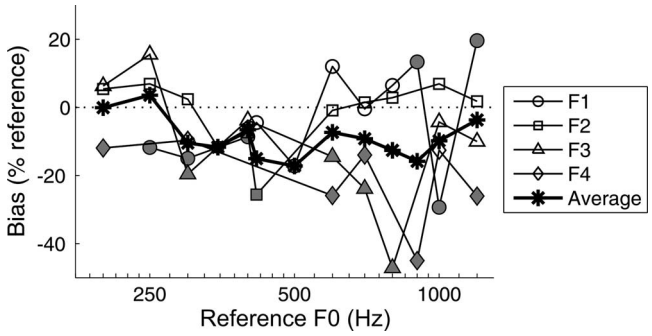


FIG. 7. Bias of four ferrets on the pitch discrimination task. Biases were calculated as the distances, in hertz, between the 50% right choice probability obtained from the fitted psychometric function and the corresponding reference F0, so that positive values indicate a bias toward the left ("lower") spout and negative values indicate a bias toward the right (higher) spout. These biases are normalized by the reference F0. Data for different animals are plotted with different symbols, and bias values that are significantly different from zero (99.87% confidence intervals; 0.05 alpha with Bonferroni correction) are shown in gray, with non-significant values in white. The average percentage of bias across all four animals is plotted with asterisks. A dotted line is shown at a bias of zero, for reference.

ence pitch could take one of two values, approximately two octaves apart. Three targets were presented: one about an octave below the low reference, one about an octave above the high reference, and one centered between the low and high references. If ferrets compared the pitch of the target to the reference pitch on each trial, they should perform this task well, responding high when the middle target was presented with the low reference and low when the middle target was presented with the higher target. If they matched targets to low and high templates, they might be expected to respond similarly to the middle target irrespective of the reference F0. If they rely on a stable memorized representation of the reference, they would be forced to adopt a new strategy in this paradigm because a stable reference is not presented, and so their performance on the variable reference task is less straightforward to predict.

After 4 days of testing with two references, ferrets discriminated the highest and lowest targets well, but responded to the middle target at chance (Fig. 8). Responses to the middle target did not differ across the two references (Tukey's HSD test). This indicates that these animals were not accustomed to comparing the pitch of the target to that of the reference presented on each trial, but they instead used a strategy that utilized internal representations.

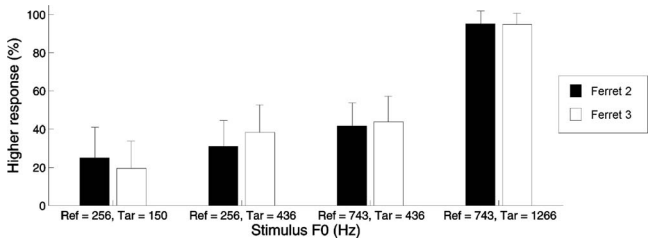


FIG. 8. Performance of two ferrets when the reference roved between two F0 values across trials. Right spout choice probabilities (mean+standard deviation, across eight consecutive testing sessions) are plotted across the four reference and target combinations presented in the variable reference testing sessions. This type of testing was carried out with ferret 2 (black bars) and ferret 3 (white bars).

B. Discrimination of the frequency of pure tones

Once trained to discriminate the pitch of artificial vowels, the same four ferrets were retested using pure tones instead of artificial vowels, but on an otherwise identical paradigm. On each trial, a reference tone was presented (200 ms duration and 20 ms rise/fall time), followed by a silent interstimulus interval (50 ms) and then a target tone (500 ms duration and 20 ms rise/fall time). Performance was again measured across a range of references between 200 and 1200 Hz. For each new reference tone, the animals were trained to criterion performance (85% correct) using two fixed targets before 30 new variable target frequencies were introduced in a single testing session, as described in Sec. II D. All ferrets generalized from the vowel to tone discrimination task well and reached criterion on the latter within the first 2 weeks of training.

Ferrets' psychometric curves for the tone version of the pitch discrimination task are shown in Fig. 9, while Fig. 10(a) shows the Weber fractions calculated from these psychometric curves, together with the Weber fractions measured with artificial vowels for comparison. The Weber fractions obtained with the vowel and tone versions of this task were clearly similar overall, and these values did not differ significantly between the two stimulus types across the reference range [two-way ANOVA; $F(1,43)=1.36$, $p=0.250$]. But note that one animal (F3) appeared to perform better on the vowel version of the task. Just as with the vowel data, the pure tone difference limens for pitch scaled with reference F0 ($r=0.88$, $p<0.001$), and when expressed as Weber fractions, discrimination performance did not show a systematic increase or decrease across the range of references tested ($r=-0.04$, $p=0.866$). Small but significant response biases were again observed in some cases, as shown in Fig. 10(b), and they again tended to favor higher (right) responses.

C. Discrimination of the pitch of artificial vowels by human listeners

In Fig. 11, the Weber fractions of human subjects are plotted together with the Weber fractions of ferrets on the same discrimination task. The pitch acuity of humans was clearly better than that of ferrets on these discrimination tasks (three-way ANOVA carried out on the difference limens across species, stimulus type, and reference periodicity; $F(1,82)=840.86$, $p<0.001$). As in the ferret data, the performance of the human listeners did not differ significantly when vowels or tones were used as stimuli [two-way ANOVA; $F(1,35)=1.81$, $p=0.188$], and difference limens for pitch scaled with the reference F0 in both the vowel ($r=0.74$, $p<0.001$) and tone ($r=0.78$, $p<0.001$) versions of the task, while Weber fractions did not change linearly across the range of references tested ($r=-0.25$, $p=0.297$ for vowels; $r=-0.04$, $p=0.855$ for tones). Analysis of the data collected for the subset of references that were tested in both human and ferret experiments showed that trends in difference limens across references did not differ between these two species (three-way ANOVA, interaction of species and reference F0; $F(3,65)=0.27$, $p=0.846$). A small but statistically significant bias was observed for 10 of the 40 tested

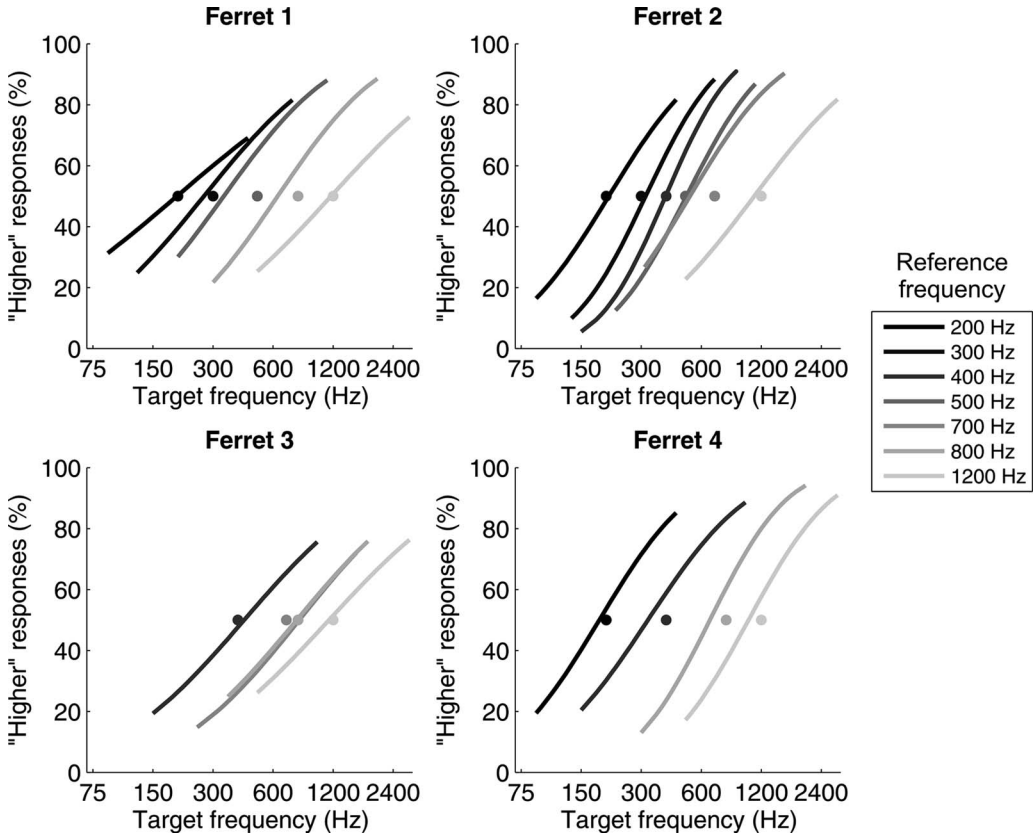


FIG. 9. Fitted psychometric curves showing the performance of four ferrets on a pure tone frequency discrimination task. The reference frequency for each curve is coded in grayscale, and the value of this frequency is also indicated by the position of the filled grayscale circles along the *x*-axis.

conditions ($p < 0.05/20$, for 20 subject-by-reference conditions tested with vowels and with tones, data not shown). In the majority (8 of 10) of these cases, the bias was again toward higher responses.

IV. DISCUSSION

In natural environments, the fundamental frequency often correlates with the size of a vibrating object. Consequently, the ability to order complex sounds along a pitch scale can be useful, as it provides information about the physical properties of sound sources (Smith *et al.*, 2005). Pitch ordering also plays an important role in vocal communication, as it often carries information not just about the gender and size, but also the emotional state of the vocalizing individual. The ability to perceive the pitch of periodic sounds along a continuous scale, from low to high, might facilitate estimating continuous valued properties of a sound source (heavy or light, large or small, relaxed or tense, empty or full). This would be a useful faculty for many species, but although there have been a number of previous investigations into frequency discrimination and, to a much lesser extent, the discrimination of periodic from non-periodic sounds in mammals, none so far have asked how well animals can judge the direction of a change in the F0 of complex sounds. The results of the present study show that ferrets can distinguish artificial vowels with F0s that are higher than a reference from those that are lower, which suggests that they, like humans, perceive pitch along an ordered scale from low to high. This study has also shown that ferrets' pitch discrimination performance for this class of stimuli closely matches their performance on an equivalent pure tone frequency discrimination task, even though the neural substrates for these two perceptual tasks could in principle be quite different.

A. Pitch direction judgments in animals

The few previous studies that have examined animals' ability to judge the direction of pitch changes have all used simple tones as stimuli. These studies have often been motivated by the question of whether animals can identify the direction of frequency changes by comparing the "relative" frequencies of several tones presented on a given trial or if they instead compare the "absolute" frequency of each tone to an internal frequency representation (i.e., they possess and use "perfect pitch"). For instance, Cynx and colleagues carried out psychophysical tasks in birds that required the animals to respond to increases or decreases in frequency across a sequence of tones (Page *et al.*, 1989; Cynx, 1995). While they showed that it is possible to train birds to make relative frequency judgments under carefully designed experimental conditions, these relative judgments do not appear to come easily. Their results suggest that both songbirds and non-songbirds prefer to label the absolute frequency of a sound to solve frequency discrimination tasks. Other groups have also experienced difficulty in training rats (D'Amato, 1988) and non-human primates (D'Amato, 1988; Izumi, 2001; Brosch *et al.*, 2004) to respond to the direction of relative frequency changes within a tone sequence.

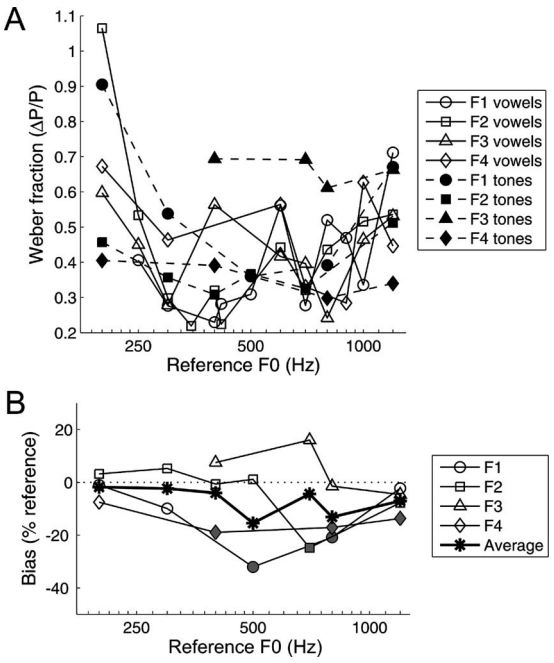


FIG. 10. Performance of four ferrets on a tone discrimination task. (A) Weber fractions of four ferrets across a range of reference F0s, measured with either the artificial vowel (open symbols, solid lines) or with pure tones (filled symbols, dashed lines). Data for each ferret are displayed with a unique symbol, as shown in the legend. (B) Bias of four ferrets on the tone discrimination task. Biases were calculated and are presented as described in Fig. 7.

The design of the task in the present study encouraged subjects to respond to the relative change in F0 between the reference and target sound, and human listeners carrying out this task report basing their decisions on relative pitch judgments. However, since the reference sound remained constant throughout a given session, the ferrets could use other strategies to perform this task. For instance, they may have formed an internal memorized representation of the reference during the first few trials in a session and then compared each target against this memorized representation instead of, or in addition to, comparing it against the reference presented at the start of each trial. Alternatively, they could have formed representations of the high and low targets presented during early training session(s) and judged whether the target presented on each testing trial was more similar to this high or low template. The results of the experiments in which the reference alternated randomly between two F0 values suggest that ferrets, like birds, rats, and primates, may adopt such an absolute pitch-matching strategy, rather than comparing pitch values across sounds presented in sequence. Whichever of these approaches were employed by the animals, in order to perform the task, the ferrets had to be able to evaluate the targets along an ordered perceptual scale that correlates with the F0 of the sound. In other words, they had to rate the pitch height of the target.

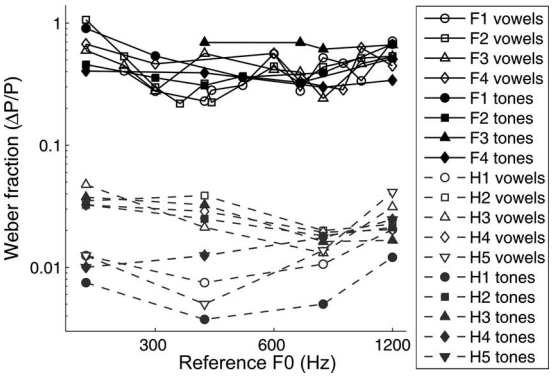


FIG. 11. Weber fractions of four ferrets and five human listeners on the pitch discrimination tasks. Data for each ferret are displayed with black symbols connected by black lines, and human data are displayed with gray symbols connected by gray lines. Weber fractions for each subject were measured using artificial vowels (open symbols) and tones (filled symbols).

A fourth strategy for solving the authors' discrimination task may be proposed, which would not be consistent with the interpretation that ferrets order the F0 of sounds. That is, the ferrets may have simply memorized the two targets and labeled them as "left response" and "right response" sounds, without making any judgment on the target sounds' relative pitch. During the early training stages, where only two different target sounds were presented per session, this would be a viable strategy. However, if the animals were only capable of a simple recognition of memorized sound samples, then that would leave them poorly prepared for the first testing session, in which 30 new targets were introduced. The authors found that the animals adapted very quickly to the testing regime and adopted a "go left for low and go right for high" strategy right from the first training session, which suggests that they quickly generalized their responses to the new target sounds as though they perceived the F0 of these harmonic sounds along an ordered pitch scale.

The authors have not been able to retrain the animals from this study on a task where references rove widely across more than 2 F0 values from trial to trial, which suggests that during their training the ferrets may have come to rely at least in part on a memorized pitch boundary. The perceptual demands of this task should not be very different, but from an animal training perspective it is much harder. Randomized references make the stimulus set a great deal more variable, and it is correspondingly harder for the animal to discover the rules by which it is to respond to the stimuli. Preliminary data from another laboratory suggest that ferrets can be trained to respond to the relative frequency of pure tones with roving references, provided the reference is varied from the beginning of training (Yin *et al.*, 2007).

B. Frequency acuity of ferrets

Ferrets have been trained on a number of sound localization (Kelly and Kavanagh, 1994; Parsons *et al.*, 1999; Kacelnik *et al.*, 2006; Bizley *et al.*, 2007; Nodal *et al.*, 2008) and detection tasks (Kelly *et al.*, 1986; Hine *et al.*, 1994; Kelly *et al.*, 1996; Fritz *et al.*, 2003; Fritz *et al.*, 2007) in the past, but rarely in discrimination tasks. In a series of studies by Shamma and colleagues, ferrets were trained on an aversively reinforced, go/no-go task to discriminate between two tones of fixed frequencies (Fritz *et al.*, 2005) or to discriminate pure tones from inharmonic tone complexes (Kalluri *et al.*, 2008). These experiments have demonstrated that ferrets can make discriminations based on simple spectral or

harmonic cues and that the response properties of primary auditory cortical neurons undergo systematic changes during this behavior. However, the present work is the first published measurement of frequency discrimination acuity in ferrets.

The pure tone Weber fractions of ferrets in the present high/low discrimination task were higher than those previously reported for frequency change detection tasks carried out in other species (reviewed by Heffner *et al.*, 1971; Fay, 1988; Shofner, 2005). For example, on frequency discrimination tasks that use a 500-Hz reference tone, Weber fractions have been measured to be 3.4% in chinchillas (Nelson and Kiester, 1978), 1.6% in guinea pigs (Heffner *et al.*, 1971), 1.4% in bushbabies (Heffner *et al.*, 1969b), 2.5% in tree shrews (Heffner *et al.*, 1969a), 3.6% in budgerigars (Dooling and Saunders, 1975), and 1.7% in cats (Elliott *et al.*, 1960). In comparison, the average Weber fraction measured for ferrets in the present direction judgment task at this reference frequency was about tenfold greater, at 36.4%. Even across studies that use the same species and very similar tasks, Weber fractions for pitch discrimination can vary widely. The average Weber fractions of chinchillas for a 250-Hz reference tone were measured on go/no-go tasks to be 4.6% by Nelson and Kiester (1978) and 21.2% by Shofner (2000). Large differences in pure tone frequency discrimination performance are also sometimes present across individual members of the same species within the same study. For instance, in one study, a group of nine male macaque monkeys tested with a 1000-Hz reference tone had individual Weber fractions that ranged from approximately 0.9% to 9.0% (Prosen *et al.*, 1990). The pitch direction judgments of human listeners have also been observed to vary widely across individuals (Semal and Demany, 2006).

Nevertheless, the relatively high-frequency difference limens measured here in ferrets warrant careful consideration. There are several factors which might explain the difference in frequency acuity measured on the authors' current 2AFC task and previous go/no-go studies. They include differences in the sensory comparison required, cognitive demands, the number of stimulus exposures per trial, variation in sound levels, and the reverberant environment.

The design of the present task required animals to classify a stimulus pitch as high or low relative to a given reference (a two-alternative-forced-choice, or 2AFC, task), while in most other previous animal studies, the participants were required only to report any detectable change in a continuously repeated sound (a go/no-go task) and did not need to be able to distinguish pitch decreases from increases. The design therefore requires a more demanding sensory judgment, and one might expect the animals' thresholds to be correspondingly higher.

Another, less intuitive, difficulty is associated with carrying out 2AFC frequency discrimination tasks in animals. Previous studies have suggested that when performing go/no-go tasks, animals tend to make response choices based on the "quality" of sounds (such as frequency or timbre), while in 2AFC tasks animals prefer to respond to the spatial location of the sound source [reviewed by Burdick (1979)]. The reasons for this task specificity, which appears to be much more pronounced in the auditory than in the visual modality, are still unclear, but they may include a difficulty in initially learning the rules required by a 2AFC task (which amount to simply "approaching the source" in localization tasks), and the working memory challenge involved with arbitrarily mapping sound quality onto two response options. Sound quality discriminations have been shown to be more difficult to train on 2AFC tasks than in go/no-go alternatives in a number of species, including guinea pigs (Upton, 1929), chinchillas (Burdick, 1980), cats (Elliott *et al.*, 1962), and monkeys (Elliott *et al.*, 1971). Of particular interest to the present discussion is a frequency discrimination study by Dobrzecka and Konorski (1968), described in Burdick (1979). They report that dogs more accurately discriminated the frequencies of pairs of tones in a go/no-go task than in a two-choice procedure (but for an alternative interpretation see Neill and Harrison, 1987).

The task design used here also differs from go/no-go studies of frequency discrimination in terms of the number of stimuli presented on each trial. In the present task, as in most human studies, the ferrets heard only one instance of the reference and target before making their response choice on each trial. In the typical go/no-go frequency discrimination task, a sequence of reference tones is presented and the signal for detection is a frequency alteration in two or more of the tones near the end of this sequence. Therefore, in these studies, the animals are provided with "multiple looks" at the sounds to be discriminated in any one trial, whereas in the 2AFC task a sensory decision is made after hearing each stimulus only once.

The randomization of stimulus level in the present study across 15 dB is larger than that used by most previous studies. The authors included this variation to prevent their animals from using simple spectral level cues and to encourage them to rely solely on the F0 of sounds to solve the task. This manipulation might also encourage subjects to rely more heavily on temporal pitch cues and/or spectral matching of lower harmonics in the stimuli, since the higher harmonics may have been resolved at low sound levels but unresolved at higher ones.

Finally, the potential effect of reverberations on pitch discrimination is worth considering. Here, stimuli were presented via a loudspeaker inside a chamber with vacuum glass walls, which provided good sound isolation, but are also reverberant. In most human studies, sounds are presented over headphones to avoid reverberation from the environment. In a recent study, Sayles and Winter (2008) demonstrated that reverberation can significantly impair one set of pitch cues, namely, those that may arise from regularities in the temporal envelope of the high-frequency part of the sound signal. However, the signals used in this experiment carry a substantial amount of pitch information that is far less affected by reverberation. This includes the temporal fine structure of the sound (i.e., frequency components at less than about 4 kHz to which auditory nerve fibers can phase lock), as well as resolvable low harmonics. In normal human listeners, spectral template matching is not thought to be the main strategy used to discriminate sounds that differ in pitch (Moore and Peters, 1992), but the presence of resolved harmonics can

nevertheless contribute to better pitch discrimination thresholds (Bernstein and Oxenham, 2006). At present, little is known about the degree to which other mammals normally rely on spectral and temporal cues for pitch discrimination.

In summary, the authors attribute the relatively elevated Weber fractions reported here for pitch judgments in ferrets to the higher cognitive and sensory demands posed by a low/high pitch judgment compared to a mere change detection. This result is consistent with a study of pitch discrimination in children with cochlear implants, who achieved better pitch discrimination thresholds in a change detection task than in a task that required them to make pitch direction judgments (Vongpaisal *et al.*, 2006). Vongpaisal *et al.* suggest that while subjects may have been able to solve the change detection task using spectral cues, these cues may have been insufficient to enable them to order the same sounds (synthesized piano notes) along a pitch scale. Thus, the performance in a change detection task may not reflect subjects' ability to tell high from low pitch. Along similar lines, Semal and Demany (2006) showed that listeners with frequency difference limens that are elevated but within the normal range find it easier to detect frequency changes than to identify the direction of those changes, while, counterintuitively, listeners with the best frequency acuity have better thresholds on a pitch direction-identification task than on a change detection task. Semal and Demany hypothesize that the human auditory cortex may contain frequency "shift detectors," which would enable the classification of small pitch changes (Demany and Ramos, 2005). The likely physiological basis for such shift detectors in the human brain, or indeed that of other mammals, remains unknown. From a comparative psychophysics point of view, it will be interesting for future studies to measure ferrets' difference limens for pitch on a go/no-go, F0 change detection task, in order to determine if and how thresholds on this task might differ from the authors' current 2AFC measurements.

The present results highlight interesting parallels between humans and ferrets. For both species, Weber fractions changed across reference F0s in similar ways, and thresholds were indistinguishable in the pure tone and the artificial vowel versions of the task. However, humans discriminated changes in F0 consistently and substantially better than ferrets. This was not entirely surprising, given that previous studies have already established that the pure tone frequency discrimination performance of humans is superior to that of many other mammals across the reference range tested here (Fay, 1988). It has been proposed that the superior performance of humans in these tasks might be due to differences in basilar membrane mechanics or higher densities of ganglion cells in the human cochlea (Elliott *et al.*, 1960). While humans' exceptional sensitivity to low-frequency pure tones may also facilitate pitch discrimination, ferrets are easily able to detect 200–1200 Hz tones presented at the levels used in the present study (Kelly *et al.*, 1986). Furthermore, the fundamental frequencies of ferrets' vocalizations are within this frequency range (unpublished observations from the laboratory of Didier Depireux), so it is reasonable to expect that they might perceive the pitch of artificial vowels.

C. Pitch discrimination for complex sounds verses pure tones

The authors observed that for both ferrets and humans, discrimination thresholds for the pitch of artificial vowels were not significantly different from those obtained with pure tone stimuli. However, previous studies of the pitch discrimination performance of humans and other species have shown that F0 acuity can depend on the type of periodic stimulus presented [reviewed by Shofner (2005)]. For example, in chinchillas (Shofner, 2000) and humans (Henning and Grosberg, 1968; Moore *et al.*, 1984), discrimination of the F0 of harmonic tone complexes can be more acute than that for pure tones at F0. In contrast, discrimination of the F0 of iterated rippled noise is poorer than pure tone frequency discrimination in humans (Yost, 1978) and chinchillas (Shofner *et al.*, 2007), while these stimulus types yield similar discrimination thresholds in goldfish (Fay *et al.*, 1983). Finally, discrimination of the modulation frequencies of sinusoidally amplitude-modulated noise bursts is poorer than the discrimination of pure tone frequencies in macaque monkeys (Moody, 1994), chinchillas (Long and Clark, 1984), parakeets (Dooling and Searcy, 1981), and humans (Formby, 1985).

Other authors have previously measured the pitch discrimination of human listeners using artificial vowel sounds (Flanagan and Saslow, 1958; Klatt, 1973), and subjects' thresholds on these tasks have been very modestly but consistently better than those for pure tone frequency discrimination. Weber fractions have been reported to be in the range of 0.23%–0.40% on pitch discrimination tasks that use a 120 Hz reference vowel (Flanagan and Saslow, 1958; Klatt, 1973). These values are slightly better than the Weber fractions of 0.44% measured for 120 Hz pure tones by Flanagan and Saslow (1958). The variability in performance across subjects in the authors' study may have been too large to observe more subtle effects of stimulus type on pitch discrimination, even though the expected effect of reference pitch on performance was clear in their data. The consensus between this study and previous ones seems to be that if there are differences between F0 difference limens for tones and vowels, these differences are small compared to the variation in difference limens across reference F0s.

D. Pitch discrimination thresholds of human listeners

The Weber fractions measured here for human subjects in the authors' study were larger than those previously reported (Flanagan and Saslow, 1958; Klatt, 1973; Wier *et al.*, 1977), and there are a number of possible explanations for this discrepancy. First, this could be due to the smaller number of training and testing trials used in the present study. Here human subjects carried out approximately 1200 pitch discrimination trials, while listeners in previous studies were much more highly practiced (Flanagan and Saslow, 1958; Klatt, 1973). Both human and animal frequency difference limens are known to continue to improve with training (Prosen *et al.*, 1990; Demany and Semal, 2002; Banai and Ahissar, 2004; Delhommeau *et al.*, 2005). The wide age range of the subjects may also have led to higher pitch

thresholds since the pitch discrimination thresholds of older adults for artificial vowels have been shown to be three times larger than those of young adults (Vongpaisal and Pichora-Fuller, 2007). Finally, while sounds were presented via a loudspeaker in the present experiments, previous studies of pitch discrimination in humans have presented stimuli over headphones. As mentioned above, reverberation may have impaired the use of high-frequency envelope pitch cues in the stimuli. While it is therefore likely that the results reported here do not reflect the limit of pitch discrimination in humans under ideal listening conditions, the authors would argue that their results are representative of the pitch discrimination capability of average human listeners functioning in everyday acoustic environments.

E. Pitch discrimination as a function of the reference F0

For both ferret and human listeners, difference limens in hertz were larger for higher-pitched references—an effect that has also been demonstrated in previous studies of pure tone frequency discrimination in humans and non-human animals (Fay, 1988). Figure 11 shows that while Weber fractions generally decrease across the range of references of 200–500 Hz, they tend to show an opposite trend for the range of references above 500 Hz. This is manifest in both the human and ferret data, though it is more pronounced in the latter. A similar trend has been observed in past studies of pure-tone frequency discrimination in humans (Rosenblith and Stevens, 1953; Moore, 1973; Wier *et al.*, 1977). While few non-human frequency discrimination studies have sampled reference frequencies below 500 Hz, those that have show compatible trends in the frequency difference limens of pigeons (Sinnott *et al.*, 1980), bushbabies (Heffner *et al.*, 1969b), chinchillas (Nelson and Kiester, 1978), and cats (Elliott *et al.*, 1960). That is, difference limens are constant for references below ~500 Hz and rise thereafter. The physiological mechanisms underlying pitch discrimination remain poorly understood, but the prevalence of such common trends suggests that similar mechanisms may be at work across a wide variety of species.

V. CONCLUSIONS

This study shows that ferrets can be trained to judge the pitch of complex sounds along a low/high scale on a 2AFC task. Ferrets' acuity on this pitch discrimination task is dependent on the F0 of the reference sound. The authors observed very similar discrimination performance for pure tones, where the periodicity information is mapped onto the tonotopic axis of the ascending auditory pathway, and for complex sounds, where the existence of an anatomical pitch map remains uncertain. These effects were consistent across ferret and human data, although the pitch acuity of humans was much better overall than that of ferrets.

ACKNOWLEDGMENTS

This research was supported by a Biotechnology and Biological Sciences Research Council Project Grant (Grant No. BB/D009758/1) to J.W.H.S., A.J.K., and J.K.B., a Rothermere Fellowship and Hector Pilling Scholarship to K.M.M.W., and a Wellcome Trust Principal Research Fellowship to A.J.K. We wish to thank the undergraduate dissertation students who have assisted in carrying out this work. Finally, we are grateful to the reviewers and editor of this manuscript, whose insightful suggestions helped to improve the quality of our report.

Banai, K., and Ahissar, M. (**2004**). "Poor frequency discrimination probes dyslexics with particularly impaired working memory," Audiol. Neuro-Otol. **9**, 328–340.

Bendor, D., and Wang, X. (**2005**). "The neuronal representation of pitch in primate auditory cortex," Nature (London) **436**, 1161–1165.

Bernstein, J. G. W., and Oxenham, A. J. (**2006**). "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," J. Acoust. Soc. Am. **120**, 3916–3928.

Bizley, J. K., Nodal, F. R., Parsons, C. H., and King, A. J. (**2007**). "Role of auditory cortex in sound localization in the midsagittal plane," J. Neurophysiol. **98**, 1763–1774.

Bizley, J. K., Walker, K. M. M., Silverman, B. W., King, A. J., and Schnupp, J. W. H. (**2009**). "Interdependent encoding of pitch, timbre and spatial location in auditory cortex," J. Neurosci. **29**, 2064–2075.

Brosch, M., Selezneva, E., Bucks, C., and Scheich, H. (**2004**). "Macaque monkeys discriminate pitch relationships," Cognition **91**, 259–272.

Burdick, C. K. (**1979**). "The effect of behavioural paradigm on auditory discrimination learning: A literature review," J. Aud. Res. **19**, 59–82.

Burdick, C. K. (**1980**). "Auditory discrimination learning by the chinchilla: Comparison of go/no go and two-choice procedures," J. Aud. Res. **20**, 1–29.

Capranica, R. R. (**1966**). "Vocal response of the bullfrog to natural and synthetic mating calls," J. Acoust. Soc. Am. **40**, 1131–1139.

Cynx, J. (**1995**). "Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon)," J. Comp. Psychol. **109**, 261–267.

D'Amato, M. R. (**1988**). "A search for tonal pattern perception in cebus monkey: Why monkeys can't hum a tune," Music Percept. **5**, 452–480.

de Lafuente, V., and Romo, R. (**2006**). "Neural correlate of subjective sensory experience gradually builds up across cortical areas," Proc. Natl. Acad. Sci. U.S.A. **103**, 14266–14271.

Delhommeau, K., Micheyl, C., and Jouvent, R. (**2005**). "Generalization of frequency discrimination learning across frequencies and ears: Implications for underlying neural mechanisms in humans," J. Assoc. Res. Otolaryngol. **6**, 171–179.

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: Perceiving shifts in non-perceived pitches," J. Acoust. Soc. Am. **117**, 833–841.

Demany, L., and Semal, C. (**2002**). "Learning to perceive pitch differences," J. Acoust. Soc. Am. **111**, 1377–1388.

Dobrzecka, C., and Konorski, J. (**1968**). "Qualitative versus directional cues in differential conditioning. IV. Left leg-right leg differentiation to nondirectional cues," Acta Biol. Exp. (Warsz.) **28**, 61–69.

Dooling, R. J., Leek, M. R., Gleich, O., and Dent, M. L. (**2002**). "Auditory temporal resolution in birds: Discrimination of harmonic complexes," J. Acoust. Soc. Am. **112**, 748–759.

Dooling, R. J., and Saunders, J. C. (**1975**). "Hearing in the parakeet (*Melopsittacus undulatus*): Absolute thresholds, critical ratios, frequency difference limens, and vocalizations," J. Comp. Physiol. Psychol. **88**, 1–20.

Dooling, R. J., and Searcy, M. H. (**1981**). "Amplitude modulation thresholds for the parakeet (*Melopsittacus undulatus*)," J. Comp. Physiol. **143**, 383–388.

Ehret, G. (**1975**). "Frequency and intensity difference limens and nonlinearities in the ear of the housemouse (*Mus. musculus*)," J. Comp. Physiol. **102**, 321–336.

Elliott, D. N., Frazier, L. A., and Haydon, R. C. (**1971**). "Relational and absolute cues in auditory discrimination by monkeys," Percept. Psychophys. **10**, 278–282.

Elliott, D. N., Frazier, L. A., and Riach, W. (**1962**). "A tracking procedure for determining the cat's frequency discrimination," J. Exp. Anal. Behav. **5**, 323–328.

Elliott, D. N., Stein, L., and Harrison, M. J. (**1960**). "Determination of absolute intensity thresholds and frequency difference thresholds in cats," J. Acoust. Soc. Am. **32**, 380–384.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, IL), pp. 451–458.
Fay, R. R., Yost, W. A., and Coombs, S. (**1983**). "Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system," Hear. Res. **12**, 31–55.
Flanagan, J. L., and Saslow, M. G. (**1958**). "Pitch discrimination for synthetic vowels," J. Acoust. Soc. Am. **30**, 435–442.
Formby, C. (**1985**). "Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners," J. Acoust. Soc. Am. **78**, 70–77.
Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (**2003**). "Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex," Nat. Neurosci. **6**, 1216–1223.
Fritz, J. B., Elhilali, M., and Shamma, S. A. (**2005**). "Active listening: Task-dependent plasticity of spectrotemporal receptive fields in primary auditory cortex," Hear. Res. **206**, 159–176.
Fritz, J. B., Elhilali, M., and Shamma, S. A. (**2007**). "Adaptive changes in cortical receptive fields induced by attention to complex sounds," J. Neurophysiol. **98**, 2337–2346.
Fujita, S., and Ito, J. (**1999**). "Ability of nucleus cochlear implantees to recognize music," Ann. Otol. Rhinol. Laryngol. **108**, 634–640.
Gfeller, K. E., Turner, C., Woodworth, G., Mehr, M., Fearn, R., Knutson, J., Witt, S., and Stordahl, J. (**2002**). "Recognition of familiar melodies by adult cochlear implant recipients and normal-hearing adults," Cochlear Implants Int. **3**, 29–53.
Hayar, A., Bryant, J. L., Boughter, J. D., and Heck, D. H. (**2006**). "A low-cost solution to measure mouse licking in an electrophysiological setup with a standard analog-to-digital converter," J. Neurosci. Methods **153**, 203–207.
Heffner, H. E., Ravizza, R. J., and Masterton, B. (**1969a**). "Hearing in primate mammals III: Tree shrew (*Tupaia glis*)," J. Aud. Res. **9**, 12–18.
Heffner, H. E., Ravizza, R. J., and Masterton, B. (**1969b**). "Hearing in primitive mammals IV: Bushbaby (*Galago senegalensis*)," J. Aud. Res. **9**, 19–23.
Heffner, R., Heffner, H., and Masterton, B. (**1971**). "Behavioral measurements of absolute and frequency-difference thresholds in guinea pig," J. Acoust. Soc. Am. **49**, 1888–1895.
Henning, G. B., and Grosberg, S. L. (**1968**). "Effect of harmonic components on frequency discrimination," J. Acoust. Soc. Am. **44**, 1386–1389.
Hine, J. E., Martin, R. L., and Moore, D. R. (**1994**). "Free-field binaural unmasking in ferrets," Behav. Neurosci. **108**, 196–205.
Izumi, A. (**2001**). "Relative pitch perception in Japanese monkeys (*Macaca fuscata*)," J. Comp. Psychol. **115**, 127–131.
Johnson, K. (**1990**). "The role of perceived speaker identity in F0 normalization of vowels," J. Acoust. Soc. Am. **88**, 642–654.
Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (**2006**). "Training-induced plasticity of auditory localization in adult mammals," PLoS Biol. **4**, e71.
Kalluri, S., Depireux, D. A., and Shamma, S. A. (**2008**). "Perception and cortical neural coding of harmonic fusion in ferrets," J. Acoust. Soc. Am. **123**, 2701–2716.
Kelly, J. B., and Kavanagh, G. L. (**1994**). "Sound localization after unilateral lesions of inferior colliculus in the ferret (*Mustela putorius*)," J. Neurophysiol. **71**, 1078–1087.
Kelly, J. B., Kavanagh, G. L., and Dalton, J. C. (**1986**). "Hearing in the ferret (*Mustela putorius*): Thresholds for pure tone detection," Hear. Res. **24**, 269–275.
Kelly, J. B., Rooney, B. J., and Phillips, D. P. (**1996**). "Effects of bilateral auditory cortical lesions on gap-detection thresholds in the ferret (*Mustela putorius*)," Behav. Neurosci. **110**, 542–550.
Klatt, D. H. (**1973**). "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception," J. Acoust. Soc. Am. **53**, 8–16.
Koda, H., and Masataka, N. (**2002**). "A pattern of common acoustic modification by human mothers to gain attention of a child and by macaques of others in their group," Psychol. Rep. **91**, 421–422.
Kojima, S., Izumi, A., and Ceugniet, M. (**2003**). "Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee," Primates **44**, 225–230.
Liu, J., and Newsome, W. T. (**2005**). "Correlation between speed perception and neural activity in the middle temporal visual area," J. Neurosci. **25**, 711–722.
Long, G. R., and Clark, W. W. (**1984**). "Detection of frequency and rate modulation by the chinchilla," J. Acoust. Soc. Am. **75**, 1184–1190.
Marvit, P., and Crawford, J. D. (**2000**). "Auditory discrimination in a sound-producing electric fish (*Pollimyrus*): Tone frequency and click-rate difference detection," J. Acoust. Soc. Am. **108**, 1819–1825.
Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (**2008**). "Phoneme representation and classification in primary auditory cortex," J. Acoust. Soc. Am. **123**, 899–909.
Moody, D. B. (**1994**). "Detection and discrimination of amplitude-modulated signals by macaque monkeys," J. Acoust. Soc. Am. **95**, 3499–3510.
Moore, B. C. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, London), Chap. 6, pp. 195–231.
Moore, B. C., Glasberg, B. R., and Shailer, M. J. (**1984**). "Frequency and intensity difference limens for harmonics within complex tones," J. Acoust. Soc. Am. **75**, 550–561.
Moore, B. C., and Peters, R. W. (**1992**). "Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity," J. Acoust. Soc. Am. **91**, 2881–2893.
Neill, J. C., and Harrison, J. M. (**1987**). "Auditory discrimination: The Konorski quality-location effect," J. Exp. Anal. Behav. **48**, 81–95.
Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J., and Schnupp, J. W. (**2008**). "Responses of auditory cortex to complex stimuli: Functional organization revealed using intrinsic optical signals," J. Neurophysiol. **99**, 1928–1941.
Nelson, D. A. (**1989**). "Song frequency as a cue for recognition of species and individuals in the field sparrow (*Spizella pusilla*)," J. Comp. Psychol. **103**, 171–176.
Nelson, D. A., and Kiester, T. E. (**1978**). "Frequency discrimination in the chinchilla," J. Acoust. Soc. Am. **64**, 114–126.
Nodal, F. R., Bajo, V. M., Parsons, C. H., Schnupp, J. W., and King, A. J. (**2008**). "Sound localization behavior in ferrets: Comparison of acoustic orientation and approach-to-target responses," Neuroscience **154**, 397–408.
Page, S. C., Hulse, S. H., and Cynx, J. (**1989**). "Relative pitch perception in the European starling (*Sturnus vulgaris*): Further evidence for an elusive phenomenon," J. Exp. Psychol. Anim. Behav. Process. **15**, 137–146.
Parsons, C. H., Lanyon, R. G., Schnupp, J. W. H., and King, A. J. (**1999**). "Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets," J. Neurophysiol. **82**, 2294–2309.
Pfingst, B. E. (**1993**). "Comparison of spectral and nonspectral frequency difference limens for human and nonhuman primates," J. Acoust. Soc. Am. **93**, 2124–2129.
Pressnitzer, D., Bestel, J., and Fraysee, B. (**2005**). "Music to electrical ears: Pitch and timbre perception by cochlear implant patients," Ann. N.Y. Acad. Sci. **1060**, 343–345.
Prosen, C. A., Moody, D. B., Sommers, M. S., and Stebbins, W. C. (**1990**). "Frequency discrimination in the monkey," J. Acoust. Soc. Am. **88**, 2152–2158.
Rosenblith, W. A., and Stevens, K. N. (**1953**). "On the DL for frequency," J. Acoust. Soc. Am. **25**, 980–985.
Sayles, M., and Winter, I. M. (**2008**). "Reverberation challenges the temporal representation of the pitch of complex sounds," Neuron **58**, 789–801.
Schreiner, C. E., and Langner, G. (**1988**). "Periodicity coding in the inferior colliculus of the cat. II. Topographical organization," J. Neurophysiol. **60**, 1823–1840.
Schulze, H., Hess, A., Ohl, F. W., and Scheich, H. (**2002**). "Superposition of horseshoe-like periodicity and linear tonotopic maps in auditory cortex of the Mongolian gerbil," Eur. J. Neurosci. **15**, 1077–1084.
Schulze, H., and Langner, G. (**1997**). "Periodicity coding in the primary auditory cortex of the Mongolian gerbil (*Meriones unguiculatus*): Two different coding strategies for pitch and rhythm?" J. Comp. Physiol. **181**, 651–663.
Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.
Shofner, W. P. (**2000**). "Comparison of frequency discrimination thresholds for complex and single tones in chinchillas," Hear. Res. **149**, 106–114.
Shofner, W. P. (**2005**). "Comparative aspects of pitch perception," in *Pitch: Neural Coding and Perception*, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer Science and Business Media, New York), pp. 56–98.
Shofner, W. P., Yost, W. A., and Whitmer, W. M. (**2007**). "Pitch perception in chinchillas (*Chinchilla laniger*): Stimulus generalization using rippled noise," J. Comp. Psychol. **121**, 428–439.

Sinnott, J. M., Petersen, M. R., and Hopp, S. L. (**1985**). "Frequency and intensity discrimination in humans and monkeys," J. Acoust. Soc. Am. **78**, 1977–1985.
Sinnott, J. M., Sachs, M. B., and Hienz, R. D. (**1980**). "Aspects of frequency discrimination in passerine birds and pigeons," J. Comp. Physiol. Psychol. **94**, 401–415.
Smith, D. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (**2005**). "The processing and perception of size information in speech sounds," J. Acoust. Soc. Am. **117**, 305–318.
Syka, J., Rybalko, N., Brozek, G., and Jilek, M. (**1996**). "Auditory frequency and intensity discrimination in pigmented rats," Hear. Res. **100**, 107–113.
Talwar, S. K., and Gerstein, G. L. (**1998**). "Auditory frequency discrimination in the white rat," Hear. Res. **126**, 135–150.
Talwar, S. K., and Gerstein, G. L. (**1999**). "A signal detection analysis of auditory-frequency discrimination in the rat," J. Acoust. Soc. Am. **105**, 1784–1800.
Upton, M. (**1929**). "The auditory sensitivity of guinea pigs," Am. J. Psychol. **41**, 412–421.
Vongpaisal, T., and Pichora-Fuller, M. K. (**2007**). "Effect of age on F0 difference limen and concurrent vowel identification," J. Speech Lang. Hear. Res. **50**, 1139–1156.
Vongpaisal, T., Trehub, S. E., and Schellenberg, E. G. (**2006**). "Song recognition by children and adolescents with cochlear implants," J. Speech Lang. Hear. Res. **49**, 1091–1103.
Wickens, T. D. (**2002**). *Elementary Signal Detection Theory* (Oxford University Press, New York, NY).
Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). "Frequency discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 178–184.
Witte, R. S., and Kipke, D. R. (**2005**). "Enhanced contrast sensitivity in auditory cortex as cats learn to discriminate sound frequencies," Brain Res. Cognit. Brain Res. **23**, 171–184.
Yin, P., Fritz, J., and Shamma, S. (**2007**). "Can ferrets perceive relative pitch?," Assoc. Res. Otolaryngol. Abstr., 141.
Yost, W. A. (**1978**). "Pitch and pitch discrimination of broadband signals with rippled power spectra," J. Acoust. Soc. Am. **63**, 1166–1175.

Iterated rippled noise discrimination at long durations

William A. Yost[a)]
Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 12 March 2009; revised 30 June 2009; accepted 6 July 2009)

Iterated rippled noise (IRN) was used to study discrimination of IRN stimuli with a lower number of iterations from IRN stimuli with a higher number of iterations as a function of stimulus duration (100–2000 ms). Such IRN stimuli differ in the strength of the repetition pitch. In some cases, the gain used to generate IRN stimuli was adjusted so that both IRN stimuli in the discrimination task had the same height of the first peak in the autocorrelation function or autocorrelogram. In previous work involving short-duration IRN stimuli (<500 ms), listeners were not able to discriminate between IRN stimuli that had different numbers of iterations but the same height of the first peak in the autocorrelation function. In the current study, IRN discrimination performance improved with increases in duration, even in cases when the height of the first peak in the autocorrelation was the same for the two IRN stimuli. Thus, future studies involving discrimination of IRN stimuli may need to use longer durations (1 s or greater) than those that have been used in the past.
 [DOI: 10.1121/1.3192345]

PACS number(s): 43.66.Hg [RLF] Pages: 1336–1341

I. INTRODUCTION

Iterated rippled noise (IRN) has been used to study the pitch and pitch strength of complex noise-like stimuli. IRN is generated by a cascade of add and delay networks as described by Yost *et al.* (1996). Models based on autocorrelation (see Yost *et al.*, 1996) have been successful in accounting for most of the data generated using IRN stimuli. Figure 1(a) shows the summary autocorrelogram[1] of an IRN stimulus generated with a 1.25-ms delay (d), gain (g) of 1, and four iterations (n). The reciprocal of the lag of the first (nonzero) peak in the autocorrelogram has been used to account for the pitch of IRN stimuli (Yost, 1996a), and the relative height of this peak has been used to account for the pitch strength of IRN stimuli, which is usually measured in discrimination experiments (Patterson *et al.*, 2000; Yost, 1996b; Yost *et al.*, 1996). The other peaks (at longer lags) in the autocorrelogram have usually not been used to account for either the pitch or the pitch strength of IRN stimuli (however, see Patterson *et al.*, 2000). The height of the first peak increases as gain (g) increases from 0 to 1.0 and as the number of iterations increases (i.e., the normalized height of the first peak of the autocorrelation function when $g=1.0$ is $n/n+1$). Of particular relevance to this paper, previous research (e.g., Patterson *et al.*, 2000; Yost *et al.*, 1996; and Yost, 1996b) has indicated that when the first peak in the autocorrelogram is the same for two different IRN stimuli, performance is at or near chance when listeners are asked to discriminate between two such IRN stimuli.

Figure 1(b) shows the auditory excitation pattern based on filtering a low pass (8000 Hz) IRN stimulus with a gammatone filter bank (see Patterson *et al.*, 1995). The stimulus for Fig. 1(b) was the same IRN stimuli used in Fig. 1(a). The excitation pattern shows spectral peaks at 800 Hz (the reciprocal of 1.25 ms) and its integer multiples out to about 6000 Hz (the upper limited of peripheral resolvability for this particular IRN stimulus). That is, the spectra of IRN stimuli have the characteristic spectral ripple with spectral peaks at one over the delay and its integer multiples and spectral valleys in between.

De Cheveigne (2007) warned that the statistics of IRN stimuli are complicated with several changing slowly over time. De Cheveigne (2007) described the slow change in spectral envelope (spectral ripples) that occurs for large number of iterations and suggested that these changes or ones related to them may be a possible cue for processing IRN stimuli. This implies that auditory processing of IRN stimuli may be different at long durations (greater than 250–500 ms, which is the typical duration of IRN stimuli used in the previous literature) as compared to shorter durations.

The main purpose of this paper was to investigate the ability of listeners to discriminate between IRN stimuli as a function of their duration. In addition, the ability of listeners to discriminate between IRN stimuli generated with different numbers of iterations was also investigated.

As the number of iterations (n) increases, so does the normalized height of the first peak in the autocorrelation function (AC1). For instance, for $n=1$, AC1=0.5, and for $n=2$, AC1=0.67. If gain (g) is decreased to 0.525 for the $n=2$ case, then AC1=0.5 for the $n=1$ case. In previous studies, using IRN stimuli that had the same value of AC1 made it very difficult for listeners to distinguish one from the other (for instance, Yost, 1996b). In this paper, the authors describe discrimination results in which the two IRN stimuli to be discriminated differed in n. In some conditions, the two stimuli were both generated with $g=1$. In other conditions, the value of g for the stimulus with the larger number of iterations was generated with a value of g so that AC1 was the same as that for the stimulus generated with the smaller number of iterations. These discrimination tasks were performed as a function of IRN duration.

[a)]Electronic mail: william.yost@asu.edu

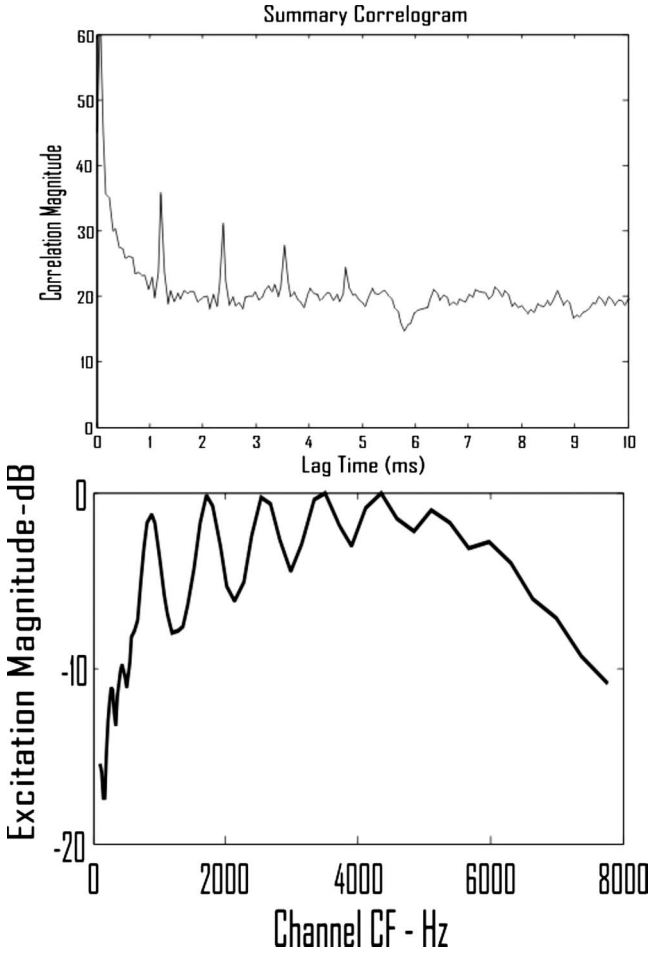


FIG. 1. The top panel represents a summary autocorrelogram [based on the auditory image model of Patterson *et al.* (1995)] for an IRN stimulus generated with a delay (d) of 1.25 ms, a gain (g) of 1.0, and four iterations (n). The large peak at a lag of 1.25 ms is used to account for the pitch of 800 Hz for the IRN stimulus, and the relative height of the first peak (and only the first peak) has previously been used to account for the pitch strength of IRN stimuli. But note the peaks at lags of 2.5 and 3.75 ms (integer multiples of 1.25 ms). The bottom panel is an auditory spectrum or excitation pattern of the IRN stimulus used for the top panel (but low passed filtered at 8 kHz). The spectral peaks at 800 Hz and its integer multiples indicate the resolved spectral ripples for this IRN stimulus.

II. METHOD

A. Participants

Six listeners who indicated they had normal hearing participated in the experiments. They ranged in age from 19 to 32, and four were female and two were male. All procedures were approved by Loyola University's Institutional Review Board (IRB).

B. Stimuli

The IRN stimuli were generated using the add-original procedure described by Yost *et al.* (1996). The IRN stimuli were generated independently during each interval on each trial, low pass filtered at 8000 Hz, and played out of an Echo Gina sound card at 44 100 samples/s. Table I indicates the values of g and n used in the experiments. The base number of iterations was increased by the additional number of iterations, and the gain (g) of the IRN stimulus with the higher number of iterations (base plus additional iterations) was the gain shown in Table I. The actual number of additional iterations used for each base number of iterations was determined in a pilot study. A d of 1.25 ms produces an 800-Hz pitch, and d of 16 ms produces a 62.5-Hz pitch (Bilsen and Ritsma, 1967/1968; Yost, 1996a). The values of g that were less than 1 were chosen so that AC1 for the IRN stimulus with the larger value of n would be the same AC1 value as for the base IRN stimulus.[2] The overall level of all IRN stimuli was randomly roved between 65 and 75 dB SPL (sound pressure level) across intervals of the same-different task. Duration ranged from 100 to 2000 ms, and the stimuli were shaped with 10-ms raised cosine rise-decay times. The stimuli were presented diotically over Sennheiser HD 280 Pro headphones, while the listeners were seated in a double-walled soundproof room.

C. Procedure

A same-different task was used to estimate percent correct. Half of the trials (determined randomly) consisted of two intervals with independent IRN noise samples but with the same values of d, g, and n used to generate the IRN stimuli. The number of iterations for these "same" trials was always the base number of iterations (see Table I), and g was always 1.0. The remaining half of the trials were "different" trials in that one of the two intervals (chosen at random) contained the IRN stimulus with the larger n (base plus additional number of iterations, see Table I), and the other interval contained the stimulus with the base number of iterations (smaller number of iterations). In different conditions, the value of g for the stimuli with the larger number of

TABLE I. Base number of iterations (n), additional number of iterations, and gain (g) used to generate the IRN stimuli.

Base n	Additional n	Gain (g)	Base n	Additional n	Gain (g)
1	1 (small)	1.0	4	8 (small)	1.0
1	1 (small)	0.525	4	8 (small)	0.82
1	3 (large)	1.0	4	16 (large)	1.0
1	3 (large)	0.51	4	16 (large)	0.8
2	3 (small)	1.0	8	16 (small)	1.0
2	3 (small)	0.56	8	16 (small)	0.84
2	8 (large)	1.0	8	32 (large)	1.0
2	8 (large)	0.65	8	32 (large)	0.89

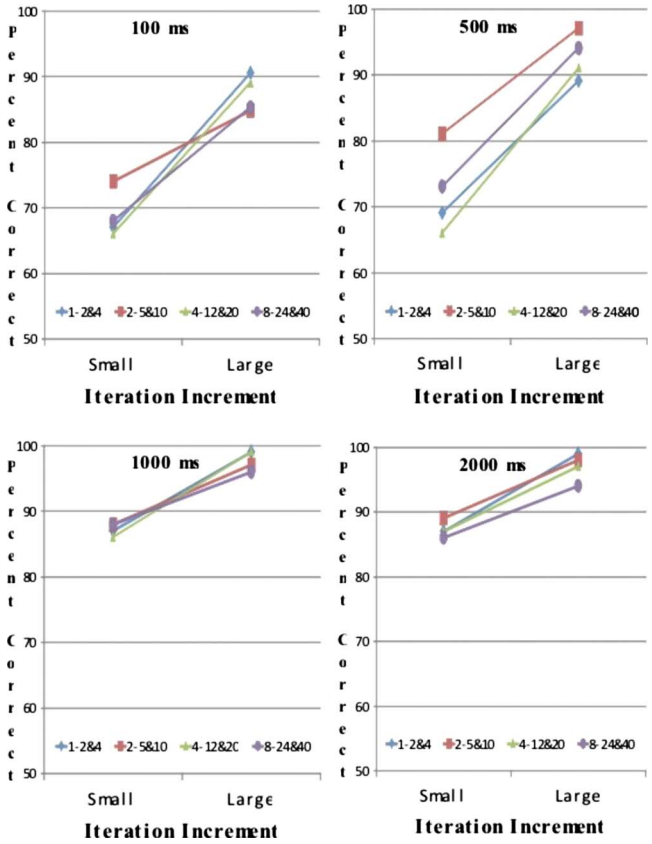


FIG. 2. (Color online) Four panels, one for each duration condition, indicating average (across listeners) percent correct discrimination as a function of a small or large increment in the number of iterations (base plus additional number of iterations) over the base number of iterations. The different curves in each panel represent the different conditions of the base number of iterations. The gain (g) was always 1.0 for all conditions shown in this figure. The data are for the delay of 1.25 ms condition, and the legend indicates the base number of iterations—the base plus the small number of additional iterations and base plus the large number of additional iterations.

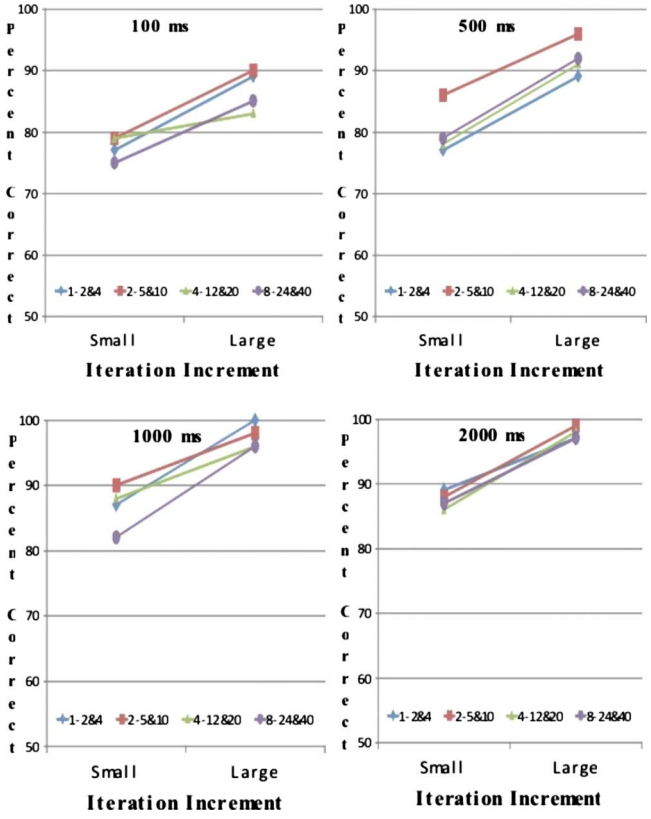


FIG. 3. (Color online) The same display shown in Fig. 2, but for data for a delay of 16 ms.

iterations was either 1.0 or set to the values shown in Table I. The two intervals were separated by 300 ms. Thus, the different trials contained IRN stimuli with different numbers of iterations. The listeners indicated whether a trial was the same or different. Feedback indicating the correct response was provided on each trial. Four 50-trial blocks were used to estimate percent correct for each condition.

III. RESULTS and DISCUSSION

Figures 2 and 3 show mean percent correct vs a small or a large difference in n (additional iterations, see Table I) for the ability of listeners to discriminate between IRN stimuli with different numbers of iterations (n) when the gain (g) was 1.0 (i.e., when the IRN stimulus with the higher number of iterations also had the higher value of AC1). For each base number of iterations, there was either a small increment in the number of iterations or a large increment (additional number of iterations). Each panel in each figure represents a different duration. Figure 2 shows the data for a delay of 1.25 ms (800-Hz pitch), and Fig. 3 shows the data for 16 ms (62.5-Hz pitch). Each item in the legend indicates the base number of iterations, followed by the base plus the small number of additional iterations, and then the base plus the large number of additional iterations (e.g., a base of 1 iterations with one and three additional iterations would be 1, 2, and 4)

Figure 4 shows data when the gain for the larger number of iterations was set so that each stimulus (small and large numbers of additional iterations) had the same value of the first peak in the autocorrelation function, AC1 [i.e., adjustments were made to g to make the stimulus with the larger number of iterations have the same value of AC1 as the IRN stimulus with the lower number of iterations[2] (see Table I)]. The data are plotted as average percent correct (averaged across listeners and the various numbers of iteration conditions for each duration) vs duration for the 1.25-ms delay (top panel) and 16-ms delay (bottom panel).

The overall result is that performance for all of the discriminations improved with increasing duration at least over the range from 100 to 1000 ms. While there was improvement in performance in all tasks with increasing duration, the improvement was not large. That is, performance increased by 10%–20%, which was statistically significant in some cases, but even at 2 s some listeners were not always 100% correct. In Figs. 2 and 3, as the number of additional iterations increased over the base number of iterations, so did discrimination performance. The change in number of iterations required to discriminate between two IRN stimuli that only differ in the number of iterations increases substantially as the base number of iterations increased. For instance, at 100 ms an $n=2$ condition can be discriminated from an $n=1$ condition with about 75% accuracy, but for an $n=4$ condition, it takes a difference of eight iterations ($n=4$ vs n

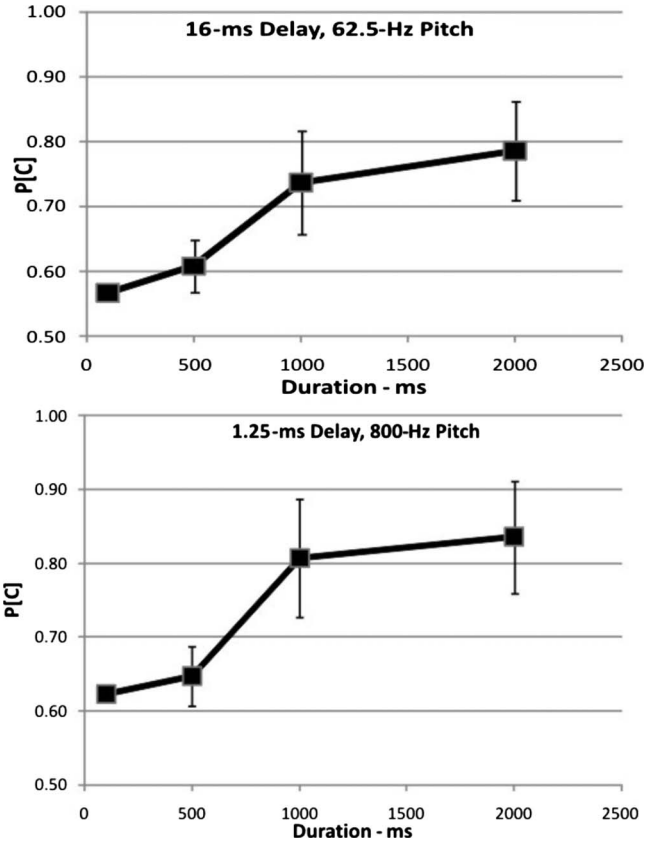


FIG. 4. The average percent correct discrimination between the smaller (base number of iterations) and larger numbers of iterations (base plus additional number of iterations) is shown as a function of duration when the gain (g, see Table I) for the larger number of iteration conditions was set such that the value of the first peak in the autocorrelation function (AC1) was the same for both IRN stimuli (small and large numbers of additional iterations) in the same-different task. The top figure represents data for a delay of 16 ms, and the bottom represents data for a delay of 1.25 ms. Error bars are one standard deviation of the mean threshold for each listener and IRN iteration condition.

=12) for this same level of performance. There was only a very small effect of delay (related to the pitch of the IRN stimuli) on the results for any of these conditions. Slightly better performance was obtained for the 1.25-ms condition over the 16-ms condition. There was no apparent change in performance based on the various base number of iteration conditions.

Yost (1996b) argued that the z-score transform of $(10^{AC1h}-10^{AC1l})/2$ describes listeners' performance in discriminating between IRN stimuli based on different heights of AC1.[3] In the equation, AC1h is the height of the first peak of the autocorrelation function for a condition with a high AC1 peak (e.g., the base plus additional number of iteration conditions), and AC1l is the height of the first peak of the autocorrelation function for a condition with a low peak height (e.g., base number of iteration conditions). In these experiments, the duration of the IRN stimuli was 500 ms. Table II shows a comparison of the results of the present study for an average of the 500-ms duration data when $g=1$ to the outcome of the function described by Yost (1996b). Except for perhaps the condition involving discrimination between $n=8$ and $n=40$, the data and the output of the equation agree reasonably well. There are at least two differences

TABLE II. Comparison of the percent correct discriminations for all of the 500-ms data of the present paper when $g=1$ as compared to the predictions of Yost (1996b). Predictions are the cumulative standard normal probabilities assuming standard scores (z-scores) determined by $(10^{AC1h}-10^{AC1l})/2$, where AC1h is the height of AC1 for the base plus additional number of iteration conditions and AC1l is the height for the base number of iteration conditions.

	Base vs base plus additional iterations			
	1 vs 2	2 vs 5	4 vs 12	8 vs 24
Data (%)	73%	83%	73%	76%
Predictions (%)	77%	83%	79%	75%
	1 vs 4	2 vs 10	4 vs 20	8 vs 40
Data (%)	89%	96%	91%	80%
Predictions (%)	94%	96%	91%	92%

in the experiments of Yost (1996b) and the present experiments that may impact comparisons between the two studies: (1) no feedback was used in Yost (1996b), and feedback was employed in the experiments of this paper; (2) the height of the first peak of AC1 was never as great in the Yost (1996b) experiments as it was for the $n=24$ and 40 conditions of the present paper. Given the difference in performance as a function of duration indicated in the current paper, the function used by Yost (1996b) would have to be modified to account for the results as a function of duration.

Two separate analysis of variance (ANOVA) tests were conducted; one when $g=1$ (Figs. 2 and 3) and one when g was not 1.0. (Fig. 4). When $g=1$, the ANOVA had two levels of delay (16 and 1.25 ms), four levels of base number of iteration (1, 2, 4, and 8), two levels of additional iterations (small and large), and four levels of duration (100, 500, 100, and 2000 ms). The main effect of additional number of iterations and the main effect of duration were both significant ($p<0.05$), while the main effect of base number of iterations and the main effect of delay were not significant ($p>0.05$), and no interactions were significant ($p>0.05$). When g was not equal to 1, the ANOVA had four levels of duration (100, 500, 1000, and 200 ms) and two levels of delay (16 and 1.25 ms). The main effect of duration was significant ($p<0.05$), while the main effect of delay and the interaction were not significant ($p>0.05$).

These data show that increasing duration improved performance in these IRN discrimination experiments. These results also indicate that at longer durations, IRN stimuli with the same height of the first peak in the autocorrelation (Fig. 4) can be discriminated from one another at performance levels above chance. This result is in contrast with previous studies (see Yost, 1996b) obtained at shorter durations (500 ms or less) that suggested that when the first peak of the autocorrelation was the same, discrimination was at or near chance. While the average discrimination performance was near chance in the present study, the average data at 100 and 500 ms were statistically different from chance using a T-test ($p<0.05$). No statistical tests to determine if the near chance discrimination performance was statistically different from chance were calculated by Yost (1996b). While the data

reported by Yost (1996b) appear to be closer to chance than those in the current paper, feedback was not employed in the discrimination tasks used by Yost (1996b) as it was in the present paper.

There was a fair amount of individual differences in overall performance, but all subjects showed the same pattern of results, as indicated in Figs. 2–4. Two subjects could perform at nearly 100% correct at 1000 and 2000 ms for the conditions in which $g=1.0$ and when there was a large increment in additional number of iterations. These listeners were close to 100% correct at 1000 and 2000 ms for the cases in which the value of the AC1 were the same for both stimuli (Fig. 4). The other four listeners always performed at less than 100% percent correct in all conditions.

It is the case that with increasing duration there is less variability from trial to trial in the spectral peaks and valleys and in the autocorrelation function peaks of the physical stimuli. Thus, the improvement in performance with increasing duration may simply reflect this change in IRN stimulus variability. Two analyses were performed to estimate this variability. Normalized autocorrelation functions were generated for all of the various numbers of iteration conditions used in this paper for the IRN stimulus generated with 1.25-ms delay, and the height of the peak at a lag of 1.25 ms was tabulated for 1000 50-trial blocks (50 000 trials). The comparison to be made was between the mean height over the 1000 blocks and the standard deviations over the 1000 blocks for the 100-ms conditions vs the 1000-ms condition. Across all of the various iterations, there was never a difference of more than 1% between the mean heights of the first peak in the autocorrelation function of the 100-ms and 1000-ms IRN stimulus conditions. While the variability (standard deviations) in the mean heights of the first peak in the autocorrelation function was always less for the 1000-ms conditions than for the 10-ms conditions, the differences were always less than 3%. A 3% change in the height of the fist peak in the autocorrelation function is not detectable (Yost, 1996b). Thus, it is unlikely that a difference in the variability in the autocorrelation functions can account for the increase in performance with increasing duration.

To analyze the results spectrally, a one-equivalent-rectangular-filter wide gammatone filter was centered at 800 Hz (the first peak in the spectrum of the 1.25-ms IRN stimulus) and a similar gammatone filter were centered at 1200 Hz (the valley above the first spectral peak). The peak-rms to valley-rms ratio expressed in decibels was calculated for 1000 50-trial blocks for the 100-ms duration condition and then for the 1000-ms duration condition. There was less than a decibel difference on average between the peak-to-valley ratios for the 100-ms condition and for the 1000-ms condition. The variability in the peak-to-valley ratios was always less for the 1000-ms condition than for the 100-ms condition, but the difference in the standard deviations between the 100- and 1000-ms duration conditions was always less than 1.39 dB. It is unlikely that this difference in spectral variability could account for the improved performance measured as a function of duration.

One basis for distinguishing between stimuli with different numbers of iterations, especially when the stimuli have the same height of the first peak in the autocorrelation function, is the height of the higher-order peaks [peaks at lags at higher integer multiples of the reciprocal of the delay (see Fig. 1) and the peaks at lags of 2.5 and 3.75 ms]. For instance, if g is adjusted so that the height of the first peak for the $n=10$ IRN stimulus is the same as that for the $n=2$ stimulus, then the height of the second peak for the $n=10$ case is close to 0.68, while the height of this second peak for the $n=2$ case is 0.33. Thus, the height of the second peak (and peaks at longer lags in the autocorrelation function) clearly differs when n differs, even when the heights of the first peak are adjusted to be equal. Thus, these peaks at longer lags are a possible cue for discrimination. The intervals in the IRN stimulus that lead to the correlations of 0.68 and 0.33 are at lags of twice d, i.e., 32 or 2.5 ms (for $d=16$ and 1.25 ms, respectively). Perhaps it takes a longer-duration stimulus to be able to sample enough of these longer intervals to aid in discrimination. No matter what cue is used for discriminating one IRN stimulus from another, performance does improve somewhat as the duration of IRN stimulation is increased out to a second or more.

One use of IRN stimuli where long-duration stimulation might need to be considered is the use of IRN stimuli in neural imaging studies (e.g., functional magnetic resonance imaging), where a considerable stimulus duration may be required to average responses. If long-duration IRN stimuli are used rather than the repeated shorter-duration pulsed stimuli, the neural image data might reflect that same ability for processing shown in this paper. That is, better processing might be suggested than would be obtained for shorter-duration stimuli.

Thus, De Cheveigne's (2007) caution seems warranted. Slow statistical changes in IRN stimuli (perhaps those associated in some way with peaks at longer lags in the autocorrelation function) may be usable in discriminating one IRN stimulus from another. As such, long-duration (of a second or more) IRN stimuli may be required to accurately determine the ability to process IRN stimuli in some conditions.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute on Deafness and Other Communication Disorders (NIDCD). The research was conducted while the author was at the Parmly Hearing Institute of Loyola University Chicago, and he is grateful for the interactions with Toby Dye, Dick Fay, and Stan Sheft. The author is also grateful to Dr. Chris Brown and Farris Walling at ASU for their comments on the paper.

[1]An autocorrelogram is computed by generating a neural activation pattern (NAP) using the auditory image model of Patterson *et al.* (1995). Then, each frequency channel of the NAP is subjected to an autocorrelation computation. These channel by channel autocorrelations are summed across channels and presented as the summary autocorrelogram. Summary autocorrelograms are similar, but not identical, to the autocorrelations of the stimulus.

[2]The equation for the normalized autocorrelation function of an ensemble average of an infinite wideband noise, described in footnote 3 of Yost *et al.* (1996), was used numerically to compute the value of g for a higher

number of iterations to be the same as that for a lower number of iterations. This equation is $AC1=[g^*(1-g^{2n})]/[1-g^{2(n+1)}]$. For $g=1.0$ conditions, g equaled 0.999 when this equation was used.

[3]Yost (1996b) described the function 10^{AC1} as a "pitch strength" function. Shofner and Selas (2002) used a power function to describe a pitch strength function and showed that it fit their data somewhat better than the function used by Yost (1996b). Since Shofner and Selas (2002) did not use discrimination procedures such as those used in Yost (1996b) and in this paper, the function (10^{AC1}) is used in this paper. The values of n used for all conditions of the present paper were chosen from a pilot study using 1000-ms long stimuli rather the 500-ms long stimuli used by Yost (1996b).

Bilsen, F. A., and Ritsma, R. J. (**1967**/68). "Repetition pitch mediated by temporal fine structure at dominant spectral regions," Acustica **19**, 114–115.

De Cheveigne, A. (**2007**). "Comment by de Cheveigne," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, New York), pp. 90 and 91.

Patterson, R. D., Allerhand, M., and Giguere, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1895.

Patterson, R. P., Yost, W. A., Handel, S., and Datta, J. A. (**2000**). "The perceptual tone/noise ration of merged iterated rippled noises," J. Acoust. Soc. Am. **107**, 1578–1588.

Shofner, W. P., and Selas, G. (**2002**). "Pitch strength and Stevens' power law," Percept. Psychophys. **64**, 437–450.

Yost, W. A. (**1996a**). "Pitch of iterated rippled noise," J. Acoust. Soc. Am. **100**, 511–518.

Yost, W. A. (**1996b**). "Pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3329–3335.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.

Tuning properties of the auditory frequency-shift detectors

Laurent Demany[a)]
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), Université de Bordeaux, BP 63, 146 Rue Leo Saignat, F-33076 Bordeaux, France

Daniel Pressnitzer
Département d'Etudes Cognitives, Laboratoire Psychologie de la Perception (UMR CNRS 8158), Université Paris-Descartes and Ecole Normale Supérieure, 29 Rue d'Ulm, F-75230 Paris Cedex 05, France

Catherine Semal
Laboratoire Mouvement, Adaptation, Cognition (UMR CNRS 5227), Université de Bordeaux, BP 63, 146 Rue Leo Saignat, F-33076 Bordeaux, France

(Received 11 December 2008; revised 16 April 2009; accepted 23 June 2009)

Demany and Ramos [(2005). J. Acoust. Soc. Am. **117**, 833–841] found that it is possible to hear an upward or downward pitch change between two successive pure tones differing in frequency even when the first tone is informationally masked by other tones, preventing a conscious perception of its pitch. This provides evidence for the existence of automatic frequency-shift detectors (FSDs) in the auditory system. The present study was intended to estimate the magnitude of the frequency shifts optimally detected by the FSDs. Listeners were presented with sound sequences consisting of (1) a 300-ms or 100-ms random "chord" of synchronous pure tones, separated by constant intervals of either 650 cents or 1000 cents; (2) an interstimulus interval (ISI) varying from 100 to 900 ms; (3) a single pure tone at a variable frequency distance (Δ) from a randomly selected component of the chord. The task was to indicate if the final pure tone was higher or lower than the nearest component of the chord. Irrespective of the chord's properties and of the ISI, performance was best when Δ was equal to about 120 cents (1/10 octave). Therefore, this interval seems to be the frequency shift optimally detected by the FSDs. [DOI: 10.1121/1.3179675]

PACS number(s): 43.66.Mk, 43.66.Hg [BCM] Pages: 1342–1348

I. INTRODUCTION

Auditory scene analysis has two facets: a segregation facet and an integration facet (Bregman, 1990). This is observable, for instance, when the scene is a rapid sequence of pure tones varying in frequency. Instead of being identified as the products of a single acoustic source, the elements of such a scene may be perceived as products of two or more concurrently active sources. This reveals the segregation facet of auditory scene analysis. On the other hand, the tones will also be perceptually linked to each other along the time dimension, into a single stream when no segregation occurs or into several streams otherwise. As a result of this sequential integration, the listener will perceive melodic patterns rather than independent sound events.

The mechanisms of auditory scene analysis remain largely unknown but are currently the topic of active research (Moore and Gockel, 2002; Snyder and Alain, 2007; Micheyl *et al.*, 2007a; Micheyl *et al.*, 2007b; Carlyon, 2004; Pressnitzer *et al.*, 2008; Elhilali *et al.*, 2009). For a scene such as the one considered above, it is possible that segregation and integration are governed by different neural processes. However, another possibility is that they have a common basis. van Noorden (1975) proposed a specific hypothesis in line with the latter idea. He suggested that the auditory system contains automatic "pitch motion detectors" working much like the spatial motion detectors of the visual system. Visual motion percepts can be elicited by discontinuous as well as continuous spatial shifts (Ekroll *et al.*, 2008). The core function of the neural machinery underlying visual motion perception is to bind successive stimuli and to give us an ability to identify them as one and the same physical object (Ullman, 1978; Shepard, 1984). In audition, similarly, pitch motion detectors might serve to bind successive tones into higher-order auditory entities subjectively emanating from a single acoustic source, even though each tone maintains its perceptual identity. Assuming that the strength of the bonds created by the detectors depends on the frequency and temporal relations of the tones, the detectors might participate in scene analysis as both integration tools and segregation tools.

Some evidence of the existence of pitch motion detectors was found in an experiment by Okada and Kashino (2004), where listeners had to judge as ascending or descending the direction of a frequency glide preceded by a repeating pair of discrete tones forming an ascending or descending melodic interval. The results showed that judgments of the glide direction were influenced by the direction of the previous melodic interval. This effect was consistent with the idea that the initial discrete tones adapted neural pitch motion detectors responding to both continuous and discrete frequency changes. Other experiments have shown that subjective judgments on the temporal order of tones can be influ-

a)Author to whom correspondence should be addressed. Electronic mail: laurent.demany@u-bordeaux2.fr

enced by previous stimuli consisting of glides (Okada and Kashino, 2003), and that streaming judgments on tone sequences can be influenced by previous sequences (Snyder *et al.*, 2008, 2009). The latter two findings can also be construed as reflecting the adaptation of pitch motion detectors. However, a subjective judgment was used in all of these studies so it is possible that the adaptating stimulus affected the listener's decision criterion (Wakefield and Viemeister, 1984; Okada and Kashino, 2003).

Demany and Ramos (2005) reported objective psychophysical observations that seem to provide stronger evidence for the existence of automatic pitch motion detectors. They constructed sound sequences in which a random "chord" of five synchronous pure tones, separated by frequency intervals of at least 0.5 octave, was followed after a 500-ms delay by a single pure tone (*T*). Because the components of the chord were synchronous, they were very difficult to hear out individually. This was objectively verified in an experimental condition where, on each trial, *T* could be either identical to one component of the chord (selected at random) or positioned halfway in frequency between two components. The task was to indicate if *T* was present in the chord or absent from it. Performance in this "present/absent" task was quite poor. In another condition, however, *T* was positioned one semitone above or below (equiprobably) one of the chord's components (selected at random), and the task was to indicate if *T* was higher or lower in pitch than the closest chord component. Surprisingly, this "up/down" task was performed much better than the present/absent task. The up/down task was relatively easy because, when *T* was relatively close in frequency to one component of the chord, the sequence formed by this component and *T* generally elicited a clear percept of directional pitch change, even though the chord's component was generally not consciously perceived. The fact that it is possible to hear a pitch change between two tones without consciously hearing one of them strongly suggests that the auditory system does contain the automatic pitch motion detectors invoked by van Noorden (1975). Demany and Ramos (2005, 2007) preferred to denominate these neural entities as "frequency-shift detectors" (FSDs).

In order to account for their findings, Demany and Ramos (2005) proposed a simple qualitative model, similar to a recent model of visual motion perception [see Ditterich *et al.* (2003); see also Allik *et al.* (1989) for related ideas in the auditory domain]. The model first assumes the existence, in the auditory system, of two subsets of FSDs respectively tuned to upward and downward frequency shifts. A second assumption is that, within each subset, the FSDs respond most strongly to *small* frequency shifts (of the same magnitude for the two subsets). The third and final assumption is that the consciously available information about the FSDs' activity is only the *difference* between the response strengths of the two subsets. This model correctly predicted that the up/down task should be easy because up and down trials were expected to activate the two subsets of FSDs in opposite ways. The model also predicted correctly that the present/absent task should be difficult because on both present and absent trials the two subsets of FSDs were expected to be activated with approximately the same strength. Finally, the model made sense of results obtained in a third condition, named "present/close," where the *T* tone could be either identical to one of the chord's components or one semitone away from one component (in either direction). The present/close task, in which listeners had to discriminate between these two types of trials, was found to be harder than the up/down task but easier than the present/absent task.

The model that we just described is only qualitative. It needs to be made more quantitative. One of its assumptions is that the FSDs respond more strongly to small frequency shifts than to large ones. But by definition, a FSD is not expected to respond strongly to a sequence of tones with identical frequencies. Thus, response strength must be maximal for frequency shifts with a certain magnitude (possibly depending on temporal parameters of the sound sequence). What is this optimal magnitude? We endeavored to estimate it in the two experiments reported here.

II. EXPERIMENT 1

A. Rationale

In this experiment, the up/down task described above was performed again using chords made up of six synchronous pure tones equally spaced on a log-frequency scale. We manipulated two independent variables. One of them was the magnitude of the frequency interval separating the components of the chord presented on a given trial; this interval (*I*) was equal to either 650 cents (1 cent=1/100 semitone =1/1200 octave) or 1000 cents. The second independent variable was the magnitude of the frequency interval Δ separating *T* (the pure tone following the chord) from the closest component of the chord; Δ was varied from 50 to 250 cents for $I=650$ cents and from 50 to 300 cents for $I=1000$ cents. Note that for each value of *I*, the maximum value of Δ was well below $I/2$, so that the task was always objectively unambiguous: Even for the maximum Δ, a listener who would be able to hear out individually the components of the chord was not expected to make errors due to an incorrect identification of the chord component closest to *T*. For such a listener, performance should either increase monotonically with Δ or increase up to some Δ value and then stay on a plateau. By contrast, for a listener unable to hear out individually the chord components but possessing automatic FSDs working as specified by the model defined above, performance could be maximal for some Δ value and worse for both smaller and larger Δ values. Logically, the value of Δ maximizing performance (Δopt) should correspond to the Δ value for which up trials and down trials elicit maximally different responses of the two hypothetical subsets of FSDs.

B. Procedure

Each tone had a total duration of 300 ms, including 5-ms raised-cosine amplitude ramps, and a nominal sound pressure level of 65 dB. The chord of six tones presented on each trial was randomly positioned between 125 and 4000 Hz (using a logarithmic frequency scale). It was followed by the *T* tone after a 500-ms interstimulus interval (ISI). As in the

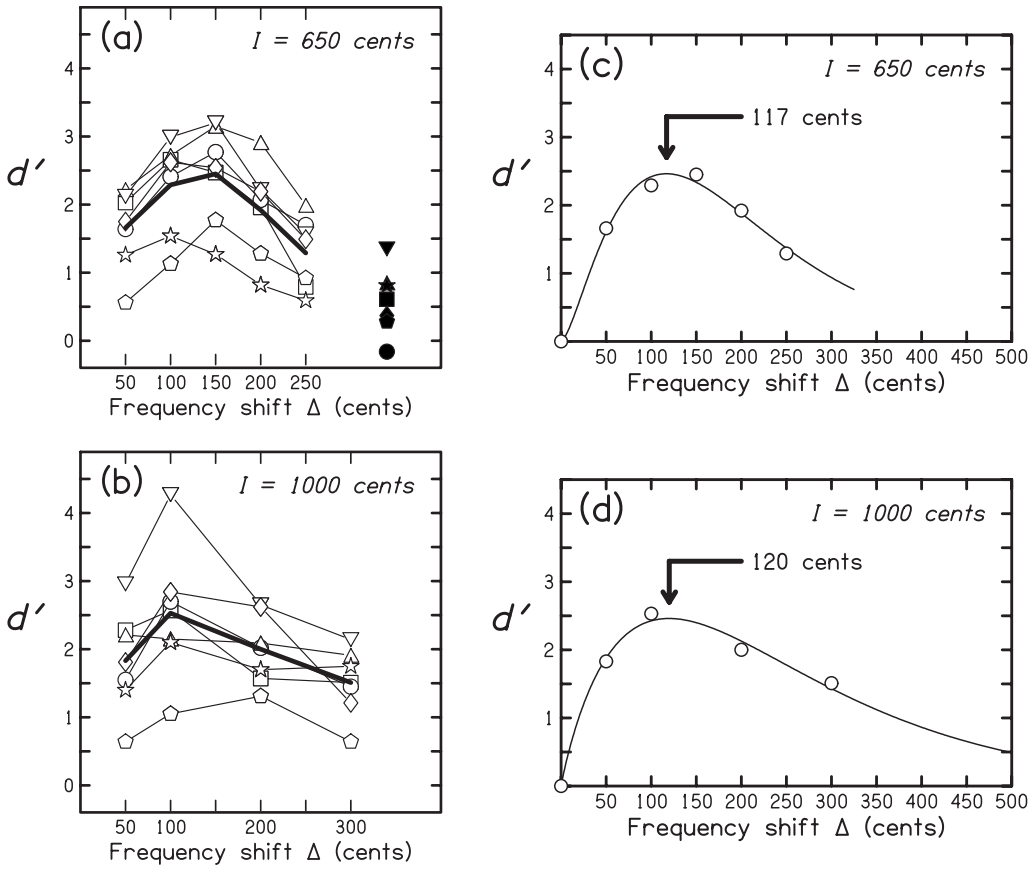


FIG. 1. (a) Results of experiment 1 for $I=650$ cents. Each listener is represented by a specific symbol shape. Open symbols show how d' varied as a function of Δ, the magnitude of the frequency shift to be judged as ascending or descending in the up/down task. Thick segments connect the mean values of d' in that task. Filled symbols represent the data obtained in the present/absent task. (b) Same as (a), but for $I=1000$ cents. The present/absent task was not performed with this value of I. (c) Circles represent the same data as those plotted with thick segments in (a), plus the expected data point for $\Delta=0$ ($d'=0$). The continuous curve is the best-fitting adjustment of Eq. (1) to the data. The Δ value corresponding to the maximum of this function is indicated by a broken arrow. (d) Same as (c), but for the data plotted in (b).

study of Demany and Ramos (2005), the chord was presented 600 ms after a random melody produced at the beginning of the trial. This random melody, serving both as a warning signal and as a pitch eraser [see Demany and Ramos (2005), footnote 1], consisted here of five immediately consecutive tones, with frequencies randomly selected between 125 and 4000 Hz.

There were ten experimental conditions. In nine of them, the listener had to perform the up/down task. On each trial, the T tone was randomly positioned Δ cents above or below (at random) one of the four "inner" components of the chord (a random choice was made among these four components). In five conditions, I was equal to 650 cents and Δ was equal to 50, 100, 150, 200, and 250 cents. In four other conditions, I was equal to 1000 cents and Δ was equal to 50, 100, 200, and 300 cents. In all nine conditions, the listener simply had to vote for up if the frequency shift Δ was positive, and for down if it was negative. In the tenth and last condition, the task to be performed was a present/absent task: The T tone could be, at random, either identical to one of the four inner components of the chord or halfway in (log) frequency between two adjacent components; one had to vote for present in the former case, and for absent in the latter case.

In each experimental session, ten blocks of 50 trials were run: one block in each of the ten conditions. These ten blocks of trials were randomly ordered. Eight sessions were run for each listener, so that overall 400 trials were performed for each condition and listener.

The stimuli were generated at a sampling rate of 44.1 kHz using a 24-bit sound card (Echo Gina). They were presented binaurally, via electrostatic headphones (Stax SR-007) in a double-walled soundproof booth (Gisol, Bordeaux). Listeners gave their responses by means of mouse clicks on two virtual buttons. Responses were not followed by immediate visual feedback, but listeners were allowed to look at their results following each of block of trials.

Seven listeners with normal hearing (five men, two women) were tested individually. This group included six students in their twenties and the first author (54 years). Most of these listeners were amateur musicians. Two of them had previously been tested in closely related experiments. The other five listeners had no previous experience with psychoacoustics. They were initially familiarized with the tasks during one or two practice sessions, using at first chords consisting of only three pure tones, very widely spaced in frequency.

C. Results and discussion

Performance was measured in terms of d' (Green and Swets, 1974). The results are displayed in Fig. 1(a) for I

TABLE I. Estimates of Δopt derived from the results of experiment 1. The table also indicates the associated values of r^2.

	$I=650$ cents		$I=1000$ cents	
	Δopt	r^2	Δopt	r^2
S1	102	0.968	72	0.958
S2	127	0.985	128	0.974
S3	127	0.970	54	0.999
S4	114	0.987	108	0.955
S5	119	0.997	127	0.993
S6	147	0.956	149	0.968
S7	91	0.997	151	0.955
Means of d'	117	0.990	120	0.990

=650 cents and in Fig. 1(b) for $I=1000$ cents. Each listener is represented in the two panels by a specific symbol shape.

The filled symbols in panel (a) represent the results obtained in the present/absent task. It can be seen that performance in this task was generally quite poor; d' exceeded 1 for only one of the seven listeners, and the average value of d' was 0.59. This poor performance indicates that the components of the chords were very difficult to perceive individually, at least for $I=650$ cents. For $I=1000$ cents, listeners informally reported that the chords' components were not noticeably easier to hear out.

Performance was generally much better in the up/down task, as indicated by the open symbols in panels (a) and (b). For this task, the grand mean of d' was 1.92 when I was 650 cents and 1.97 when I was 1000 cents. Thus, I had essentially no effect on overall performance. For each value of I, however, performance was markedly dependent on Δ, as confirmed by a repeated-measures analysis of variance [for $I=650$ cents: $F(4,24)=19.8$, $P<10^{-6}$; for $I=1000$ cents: $F(3,18)=7.8$, $P=0.0015$]. It can be seen that in most cases, as Δ increased from 50 cents to its maximal value, d' initially increased and then decreased. There were only two exceptions to this rule: For one listener, when I was 1000 cents, d' monotonically (but very slowly) decreased as Δ increased; for the other listener, when I was 1000 cents, d' was slightly larger for $\Delta=300$ cents than for $\Delta=200$ cents.

In order to estimate precisely, for the two types of chord, the value of Δ maximizing d' in the up/down task (i.e., Δopt), we fitted continuous curves to the individual and mean data, taking into account the fact that d' should be equal to 0 for $\Delta=0$. It was found that very good fits could be achieved with the scaled gamma distribution function,

$$d' = a \cdot \Delta^b \exp(-c\Delta). \quad (1)$$

The three parameters of this function, a, b, and c, were adjusted iteratively using the NCSS statistical software. Table I indicates, for each listener as well as for the mean data [plotted with thick lines in Figs. 1(a) and 1(b)], the resulting estimates of Δopt, as well as the associated r^2 statistics reflecting the proportion of variance accounted for by the best-fitting functions. In Figs. 1(c) and 1(d), the mean data are replotted as circles, and the functions fitted to them are displayed.

It can be seen that, on the basis of the mean data, Δopt was estimated at 117 cents for $I=650$ cents and 120 cents for $I=1000$ cents. These two figures are almost identical. They are also close to the figures obtained by averaging the individual estimates of Δopt across listeners; the corresponding means are 118 cents for $I=650$ cents and 113 cents for $I=1000$ cents. Note that there was no correlation between the individual estimates of Δopt for the two values of I ($r=0.07$). The variability of these individual estimates for a given value of I may arise mainly from the limited accuracy of our performance measurements and the rather coarse sampling of Δ (especially for $I=1000$ cents) rather than from genuine differences between listeners.

For $I=1000$ cents, Δopt was found to be smaller than $I/8$. Thus, it would be very unreasonable to ascribe the decrease in performance for Δ beyond Δopt to errors in the identification of the chord component closest to T. This would, in addition, presuppose unrealistically that the components of the chords could be perceived individually. The decrease in performance beyond Δopt must originate mainly from properties of the auditory system rather than merely from the chords' characteristics. Such a view is supported by the fact that Δopt did not appear to change when I changed from 650 to 1000 cents. Our interpretation of the results is that the auditory system contains automatic FSDs activated optimally by frequency shifts of about 120 cents, at least when the temporal parameters of the sound sequence are those used in this experiment.

III. EXPERIMENT 2

A. Rationale

Experiment 1 suggests that the FSDs are optimally activated by frequency shifts of about 120 cents, but is this optimal magnitude a constant or does it depend on the temporal characteristics of the sound sequence containing the frequency shift? We carried out experiment 2 to answer that question. This second experiment was a variant of experiment 1 in which we examined again the effect of Δ on performance in the up/down task, but using now shorter tones and a variable ISI between the chord and T.

B. Procedure

The stimuli were the same as those used in the up/down condition of experiment 1, except for the following differences: (1) each tone now had a total duration of 100 ms; (2) I was fixed at 1000 cents; (3) the ISI separating the chord from T could be equal to 100, 250, or 900 ms, thus producing stimulus-onset asynchronies (SOAs) of 200, 350, and 1000 ms; (4) the components of the random melody preceding the chord were always separated by 250-ms ISIs; (5) Δ could be equal to 50, 100, 150, 200, 250, or 300 cents.

In a first stage of the experiment, the ISI separating the chord from T took two possible values: 250 and 900 ms. Each session, in this stage, consisted of 12 blocks of 40 trials, one block for each combination of ISI and Δ; these 12 blocks were randomly ordered, and ten sessions were run for each listener. Then, in the second stage, the ISI was fixed at 100, and each session consisted of six randomly ordered

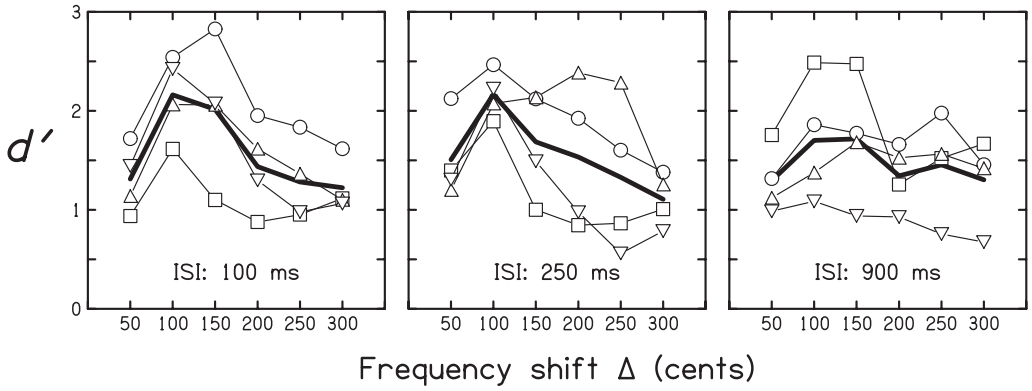


FIG. 2. Individual (open symbols) and average (thick curves) values of d' in experiment 2. Each panel represents the data obtained for a given value of the ISI separating the chord from the following *T* tone.

blocks of 40 trials, one block for each Δ; ten sessions were again run for each listener (but sessions were often grouped in pairs, run on the same day). Overall, therefore, each listener performed 400 trials in the 18 subconditions (three ISIs × six Δ values). Four male listeners, including the first author, were tested. All of them had previously taken part in experiment 1.

C. Results and discussion

The results are displayed in Fig. 2. When the ISI was 100 or 250 ms, d' was clearly a nonmonotonic function of Δ, as in experiment 1. When the ISI was 900 ms, a similar trend could again be discerned, but it was less clear. The mean values of d' for the three ISIs were very close to each other, differing by less than 0.1. A repeated-measures analysis of variance [Δ × ISI] revealed a significant main effect of Δ [$F(5,15)=7.4$, $P=0.001$], but no main effect of the ISI [$F(2,6)<1$], and no significant interaction of the two factors [$F(10,30)=1.5$, $P=0.18$].

As in experiment 1, continuous functions defined by Eq. (1) were fitted to the individual and mean data for each ISI, in order to estimate Δopt. It was again assumed, in so doing, that d' was equal to 0 for $\Delta=0$. Table II shows the obtained estimates of Δopt, as well as the associated values of r^2. The grand mean of r^2 (0.929) was lower than that found in experiment 1 (0.978) but still high enough to consider that, overall, the fits were satisfactory. For each ISI, moreover, the mean of the four individual Δopt estimates was close to the Δopt estimated from the means of d' across listeners. The functions fitted to these means are shown in Fig. 3, together with the means themselves. For clarity, the values of d' have been increased by 1 for the 250-ms ISI and by 2 for the 900-ms ISI; this is why the ordinate axis is not numbered. It can be seen that Δopt remained approximately constant when the ISI changed. The mean of the three Δopt values indicated in the figure is 121 cents. This estimate differs by only 1 cent from the one at which we arrived in experiment 1 when *I* had the same value as in the present experiment, i.e., 1000 cents.

Figure 3 suggests that, as the ISI increased, d' decayed less and less rapidly when Δ exceeded Δopt. However, the reliability of this trend is uncertain since the analysis of variance reported above did not demonstrate a significant interaction between Δ and ISI.

IV. GENERAL DISCUSSION

We infer from the present study that the automatic FSDs of the human auditory system (Demany and Ramos, 2005, 2007) are optimally sensitive to frequency shifts of about 120 cents, i.e., one-tenth of an octave, at least when the shifts take place in slow or moderately rapid sound sequences. They arrived at this estimate with tones lasting 300 ms (experiment 1) as well as 100 ms (experiment 2) and with four different SOAs: 200, 350, 800, and 1000 ms. Further work would be needed to check that the optimal frequency shift, measured in cents (that is to say, in logarithmic units), is largely independent of frequency. A previous study by Rose and Moore (2000) suggests that departures from this assumption are possible. We investigated auditory stream segregation in rapid tone sequences (ABA-ABA-,…) made up of two pure tones A and B differing in frequency. They found that the minimum frequency interval permitting to hear the tones A and B in two separate streams was less constant

TABLE II. Estimates of Δopt derived from the results of experiment 2. The table also indicates the associated values of r^2.

	ISI=100 ms		ISI=250 ms		ISI=900 ms	
	Δopt	r^2	Δopt	r^2	Δopt	r^2
S1	109	0.930	99	0.903	91	0.993
S2	131	0.985	157	0.935	179	0.989
S4	116	0.788	55	0.793	109	0.829
S5	128	0.960	95	0.995	159	0.952
Means of d'	122	0.947	109	0.965	131	0.969

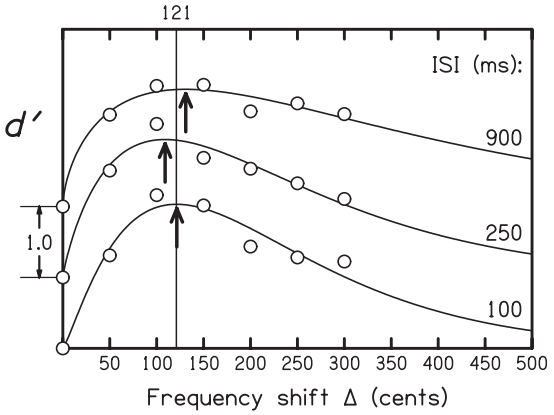


FIG. 3. Best fits of Eq. (1) to the mean data plotted in Fig. 2. These mean data are replotted here as circles, with vertical shifts: For the 250-ms ISI, the values of d' have been increased by 1; for the 900-ms ISI, the values of d' have been increased by 2. Arrows indicate the Δ values corresponding to the maxima of the fitted curves; the mean of these three Δ values is 121 cents.

across frequency when expressed in cents than when expressed in ERB units, that is, in terms of the bandwidth of the auditory filters (Glasberg and Moore, 1990; Moore, 2003).

Another goal for future research is to evaluate more precisely the shape of the FSDs' "tuning curves." We assumed for simplicity that there are only two opponent channels of FSDs, up and down, with symmetrical properties. Such a hypothesis is sufficient to account for the present and previous data (Demany and Ramos, 2005, 2007) if listeners base their judgments on the difference between the activities of the two channels. Suppose, in addition, that the (signed) frequency shift maximizing the activity of each channel does not activate significantly the other channel. If so, the absolute value of this frequency shift will be equal to the Δ value maximizing performance in the up/down task (i.e., the Δ value that we defined here as Δopt). More complex layouts for FSDs are possible, including, for instance, largely overlapping channels and/or a set of more than two opponent channels, but such additions to the original model of Demany and Ramos (2005) do not seem necessary to explain the available behavioral data.

It would be relevant to replicate experiment 2 using SOAs even shorter than 200 ms. For such SOAs, the FSDs might be maximally activated by frequency shifts smaller than 120 cents, and an even more likely possibility is that they become less sensitive to frequency shifts exceeding the optimal shift. This conjecture stems from the fact that the perception of temporal coherence in rapid melodic sequences of pure tones is limited by a tradeoff between their speed and the size of the melodic intervals: Increasing the speed of a sequence reduces the range of melodic intervals for which the sequence can be perceived as a single coherent melody (van Noorden, 1975; Micheyl *et al.*, 2007b). As pointed out by van Noorden (1975) (see also Bregman and Achim, 1973), this auditory phenomenon is analogous to the visual phenomenon known as Korte's third law of apparent motion (Korte, 1915; Lakatos and Shepard, 1997; Ekroll *et al.*, 2008). The link between FSDs and perceptual streaming, however, remains to be clarified. For melodic sequences with relatively short SOAs, frequency selectivity and adaptation are sufficient to predict stream segregation (Bee and Klump, 2005; Micheyl *et al.*, 2005; Pressnitzer *et al.*, 2008). Activity in FSD channels could provide an additional encoding of the sequences, signaling the binding between tones.

Since in any case the FSDs seem to be particularly sensitive to frequency shifts of about 120 cents over a wide temporal range, it is worth considering if such a frequency distance is also perceptually special in other ways. In this regard, we shall first note that 120 cents appears to be, over a wide frequency range, a good estimate of the minimum frequency interval permitting a perceptual segregation of two *simultaneous* pure tones, in the absence of other tones (Plomp, 1964). However, this might well be a fortuitous coincidence: Making sense of it is not straightforward.

Another coincidence could be more meaningful. In an experiment on pitch memory, Deutsch (1972) required listeners to make pitch comparisons (same/different judgments) on two pure tones separated by a sequence of interfering tones. All but one of the interfering tones were remote in frequency from the initial test tone. The independent variable was the frequency distance separating the remaining interfering tone from the initial test tone: This distance (D) varied from 0 to 200 cents in 33-cent steps. It was found (on both "same" and "different" trials) that listeners' error rate steadily increased when D was increased from 0 to 133 cents but then steadily decreased when D was further increased to 200 cents. Deutsch and Feroe (1975) confirmed this observation and proposed an interesting explanation of it. They assumed that pitch memory traces are stored in a tonotopically organized neural network where lateral inhibitory interactions take place and where inhibition is maximal for a frequency distance of about 133 cents. In support for a role of lateral inhibition, Deutsch and Feroe (1975) showed that the deleterious effect of an interfering tone I_1 on the memory trace of a test tone 133 cents away can be reduced by the presentation, following I_1, of another interfering tone I_2, 133 cents away from I_1 and 266 cents away from the test tone. A natural interpretation of this finding is that I_2 inhibits the trace of I_1 and, in doing so, disinhibits the trace of the test tone.

The critical frequency distance identified by Deutsch and Feroe (1975) is not significantly different from the one identified here, given among other things the individual variability of our data (see Tables I and II). Now if at this frequency distance a tone affects in a particularly strong way the internal representation of a previous tone [as shown by Deutsch and Feroe (1975)], then one can hypothesize that the initial tone also affects in a special way the encoding of the subsequent tone. However, this "forward interaction" hypothesis is not sufficient to account for our results in the up/down task; it must be supposed, in addition, that the effect of the initial tone (a chord component, C) on the encoding of the subsequent tone (T) crucially depends on the direction of the frequency shift. One could imagine, for example, that the auditory system's "normal" response to T is inhibited by C when the frequency shift is negative but enhanced by C when the frequency shift is positive. However, such a scenario would imply that the intensity of T cannot be encoded independently of the relationship between the frequencies of

C and *T*. Generally speaking, as pointed out by Demany and Ramos (2007) and Demany and Semal (2008), the mere existence of a forward interaction between two successive sounds in the auditory system does not immediately account for the perception of a relation between them; the listener must dissociate, in the neural activity concomitant to the presentation of the second sound, what is due to the relation between the two sounds from what could be due to intrinsic properties of the second sound. Another problem with the forward interaction hypothesis, in the present context, is that frequency shifts between tones can be detected automatically even for ISIs exceeding 1 s (Demany and Ramos, 2005), whereas forward masking (i.e., the effect of one sound on the absolute threshold of a subsequent sound) is only observable for ISIs smaller than 100–200 ms (Moore, 2003). We thus believe that the forward interaction hypothesis is not an adequate explanation of listeners' success in the up/down task. In other words, it seems unlikely that this task can be performed by considering only the internal representation of the tone following the chord. We argue instead that successful performance rests upon the existence of FSDs that do not participate in the encoding of the tones themselves. The precise mechanism of their action remains to be elucidated.

ACKNOWLEDGMENTS

We thank Makio Kashino, Brian Moore, and an anonymous reviewer for judicious comments on a previous version of this paper.

Allik, J., Dzhafarov, E. N., Houtsma, A. J. M., Ross, J., and Versfeld, N. J. (**1989**). "Pitch motion with random chord sequences," Percept. Psychophys. **46**, 513–527.

Bee, M. A., and Klump, G. M. (**2005**). "Auditory stream segregation in the songbird forebrain: Effects of time intervals on responses to interleaved tone sequences," Brain Behav. Evol. **66**, 197–214.

Bregman, A. S., (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Bregman, A. S., and Achim, A. (**1973**). "Visual stream segregation," Percept. Psychophys. **13**, 451–454.

Carlyon, R. P. (**2004**). "How the brain separates sounds," Trends Cong. Sci **8**, 465–471.

Demany, L., and Ramos, C. (**2005**). "On the binding of successive sounds: Perceiving shifts in nonperceived piches," J. Acoust. Soc. Am. **117**, 833–841.

Demany, L., and Ramos, C., (**2007**). "A paradoxical aspect of auditory change detection," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Heidelberg), pp. 313–321.

Demany, L., and Semal, C., (**2008**). "The role of memory in auditory perception," in *Auditory Perception of Sound Sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 77–113.

Deutsch, D. (**1972**). "Mapping of interactions in the pitch memory store," Science **175**, 1020–1022.

Deutsch, D., and Feroe, J. (**1975**). "Disinhibition in pitch memory," Percept. Psychophys. **17**, 320–324.

Ditterich, J., Mazurek, M. E., and Shadlen, M. N. (**2003**). "Microstimulation of visual cortex affects the speed of perceptual decisions," Nat. Neurosci. **6**, 891–898.

Ekroll, V., Faul, F., and Golz, J. (**2008**). "Classification of apparent motion percepts based on temporal factors," J. Vision **8**, 1–22.

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A., and Shamma, S. A. (**2009**). "Temporal coherence in the organization and representation of auditory scenes," Neuron **61**, 317–329.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M., and Swets, J. A., (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York).

Korte, A. (**1915**). "Kinematoskopische Untersuchungen [Cinematoscopic investigations]," Z. Psychol. Z. Angew. Psychol. **72**, 193–296.

Lakatos, S., and Shepard, R. N. (**1997**). "Constraints common to apparent motion in visual, tactile, and auditory space," J. Exp. Psychol. Hum. Percept. Perform. **23**, 1050–1060.

Micheyl, C., Carlyon, R. P., Gutschalk, A. , Melcher, J. R., Oxenham, A. J., Rauschecker, J. P., Tian, B., and Courtenay Wilson, E., (**2007a**). "The role of auditor cortex in the formation of auditory streams," Hear. Res. **229**, 116–131.

Micheyl, C., Shamma, S. A., and Oxenham, A. J. (**2007b**). "Hearing out repeating elements in randomly varying multitone sequences: A case of streaming?," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Heidelberg), pp. 313–321.

Micheyl, C., Tian, B., Carlyon, R. P., and Rauschecker, J. P. (**2005**). "Perceptual organization of tone sequences in the auditory cortex of awake macaques," Neuron **48**, 139–148.

Moore, B. C. J., (**2003**). *An Introduction to the Psychology of Hearing* (Elsevier, Amsterdam).

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. Acust. **88**, 320–332.

Okada, M., and Kashino, M. (**2003**). "The role of spectral change detectors in temporal order judgment of tones," NeuroReport **14**, 261–264.

Okada, M., and Kashino, M. (**2004**). "The activation of spectral change detectors by a sequence of discrete tones," Acoust. Sci. & Tech. **25**, 293–295.

Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Pressnitzer, D., Sayles, M., Micheyl, C., and Winter, I. M. (**2008**). "Perceptual organization of sound begins in the auditory periphery," Curr. Biol. **18**, 1124–1128.

Rose, M. M., and Moore, B. C. J. (**2000**). "Effects of frequency and level on auditory stream segregation," J. Acoust. Soc. Am. **108**, 1209–1214.

Shepard, R. N. (**1984**). "Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming," Psychol. Rev. **91**, 417–447.

Snyder, J. S., and Alain, C. (**2007**). "Toward a neurophysiological theory of auditory stream segregation," Psychol. Bull. **133**, 780–799.

Snyder, J. S., Carter, O. L., Hannon, E. L., and Alain, C. (**2009**). "Adaptation reveals multiple levels of representation in auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. In press.

Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E., and Alain, C. (**2008**). "Effects of context on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **34**, 1007–1016.

Ullman, S. (**1978**). "Two dimensionality of the correspondence process in apparent motion," Perception **7**, 683–693.

van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. dissertation, Institute for Perception Research, Eindhoven, The Netherlands.

Wakefield, G. H., and Viemeister, N. F. (**1984**). "Selective adaptation to linear frequency-modulated sweeps: Evidence for direction-specific FM channels?" J. Acoust. Soc. Am. **75**, 1588–1592.

An influence of amplitude modulation on interaural level difference processing suggested by learning patterns of human adults

Yuxuan Zhang[a] and Beverly A. Wright
Department of Communication Sciences and Disorders and Interdepartmental Neuroscience Program, Northwestern University, Evanston, Illinois 60208

(Received 24 April 2008; revised 16 June 2009; accepted 16 June 2009; corrected 16 December 2009)

Humans rely on interaural level differences (ILDs) to determine the location of sound sources, particularly for high-frequency sounds. Previously, ILD-discrimination performance with a 4-kHz pure tone was reported to improve with multi-hour training. Here the effect of the same training regimen on ILD discrimination with a 4-kHz tone sinusoidally amplitude modulated (SAM) at 0.3 kHz was examined. Ten of the 16 trained listeners improved more than untrained controls, demonstrating training-induced learning. However, compared to the learning previously obtained with the 4-kHz pure tone, learning with the SAM tone was less predictable based on starting performance, took longer to complete, and was characterized by specificity to stimulus type (SAM vs pure tones) rather than stimulus frequency. These differences demonstrate an influence of amplitude modulation on learning of ILD discrimination. This influence suggests that the auditory system makes use of amplitude envelope information in determining ILD-discrimination performance, a form of interaction between time and level processing in the binaural system.
 [DOI: 10.1121/1.3177267]

PACS number(s): 43.66.Pn [RLF] Pages: 1349–1358

I. INTRODUCTION

Many auditory skills improve with practice even in adults (for review, see Watson, 1980; Irvine and Wright, 2005; Wright and Zhang, 2006, 2009), raising the possibility that training programs can be used to treat individuals with hearing disorders and to create new hearing expertise. To help establish the principles of auditory learning, we have been examining how basic auditory skills in normal-hearing adults improve with training and how the improvements generalize to untrained conditions. We also use these learning patterns to make inferences about the neural processes underlying the trained skills (also see Karni and Sagi, 1991; Ahissar and Hochstein, 2004). Of interest here is how listeners learn to better detect changes in two primary sound-localization cues, interaural level differences (ILDs) and interaural time differences (ITDs). Sensitivity to these cues allows us to register the source of a sound in space as well as to separate concurrent sounds from different sources.

Human sensitivity to ILDs and ITDs represents an exquisite example of neural computation by the central nervous system through which information that is not available at the sensory periphery is extracted (Hafter, 1984). ILDs arise because a sound is attenuated by the head and torso during its transmission so that the sound level at the ear closer to the source is larger than that at the ear farther from the source. The ILD magnitude depends on the stimulus frequency as well as the stimulus source location. Because at low frequencies the sound wave largely "bends around" the head and torso and therefore undergoes little attenuation, ILDs are considered to function best at high frequencies. ITDs are created because it takes the sound less time to travel to the closer ear than to the farther ear. The magnitude of the ITD is almost solely determined by the difference in the sound-transmission length to the two ears. For pure tones, ITDs can only serve as an effective cue to sound source location at low frequencies ($<\sim 1.5$ kHz). However, they are effective at high frequencies for sounds with low-frequency amplitude envelopes.

Though ILD and ITD sensitivities have been extensively studied with both physiological and psychophysical methods, examinations of the change in such sensitivity with experience have only recently begun (Wright and Fitzgerald, 2001; Rowan and Lutman, 2006, 2007; Zhang and Wright, 2007). We previously observed learning on ILD, but not ITD, discrimination in normal-hearing adults induced by multiple-day training, suggesting a type of modifiability that influences level but not timing sensitivity in the binaural system. In these tasks, listeners discriminated sounds that differed only in their ILD or ITD value. Following convention, we term the task ILD or ITD discrimination based on the cue manipulated (e.g., see Stern *et al.*, 1983; Koehnke *et al.*, 1986; Yost and Dye, 1988). In our initial investigation, we trained one group of listeners for multiple days on ILD discrimination with a high-frequency (4 kHz) pure tone and another group on ITD discrimination with a low-frequency (0.5 kHz) pure tone (Wright and Fitzgerald, 2001). Training-induced learning (defined as significantly more improvement in trained listeners than untrained controls) was observed for ILD but not for ITD discrimination, and the ILD learning did

[a]Author to whom correspondence should be addressed. Electronic mail: y-zhang6@northwestern.edu

not generalize to the ITD condition. We subsequently tested whether the distinct learning patterns observed in our initial investigation resulted from the difference between the stimulus frequencies (4 vs 0.5 kHz) by training another group of listeners, using the same regimen as before, on ITD discrimination with a 4-kHz tone sinusoidally amplitude modulated (SAM) at 0.3 kHz (Zhang and Wright, 2007). The SAM tone was used because humans are not sensitive to ITDs in pure tones in the high-frequency region. There was no training-induced improvement on this 4-kHz SAM ITD condition, similar to the training results for ITD discrimination with the 0.5-kHz pure tone but different from those for ILD discrimination with the 4-kHz pure tone. Thus, the learning pattern varies with the cue manipulated, instead of with the stimulus frequency region used (Zhang and Wright, 2007). Taken together, these results demonstrate that our multi-day training regimen can differentially modify the neural processes governing performance on interaural discrimination when different cues are manipulated.

Taking advantage of this demonstration, here we used the same training regimen to investigate whether the timing characteristics of the stimulus influence ILD processing in terms of its modifiability. Specifically, we asked whether amplitude modulating the trained stimulus results in a different learning pattern on ILD discrimination from that previously obtained with the 4-kHz pure tone (Wright and Fitzgerald, 2001). While the binaural system is best known for the microsecond level of time calculation shown by ITD sensitivity, impressive temporal fidelity in the range of a few milliseconds is also present in the neural circuits traditionally considered as the ILD pathway (for review, see Tollin, 2003). Tollin (2003) suggested that such high temporal fidelity in ILD processing allows temporal variations in the stimuli arriving at each ear to be accurately represented and compared in a time-locked manner over short time intervals. However, to date, behavioral demonstrations that the timing characteristics of the stimulus influence ILD processing are lacking. Here, we compared ILD processing, in terms of its modifiability, between two stimuli with different timing characteristics: a 4-kHz pure tone [data from Wright and Fitzgerald (2001)] and a 4-kHz SAM tone (the present experiment). Toward this end, we trained a new group of listeners on ILD discrimination with the SAM tone using the same training regimen as in the previous ILD-discrimination training experiment with the pure tone. The amplitude modulation (AM) rate of the SAM tone was 0.3 kHz (a period of 3.3 ms), a rate that likely taxes the temporal fidelity in ILD processing. Thus, ILD performance with this stimulus could potentially benefit from improved temporal processing. A difference in ILD-discrimination learning patterns between the SAM tone and the pure tone would indicate an influence of AM on the neural processes governing ILD-discrimination performance.

II. MATERIALS AND METHODS

A. Listeners

Thirty-two normal-hearing human adults (19 women) between the ages of 18 and 36 years (average of 22.5 years) participated in the experiment. All were paid for their participation. None of the listeners had previous experience in any psychoacoustic experiment.

B. Experimental organization

The experiment consisted of a pretest session, nine ~1-h training sessions, and a post-test session conducted on consecutive days except weekends. Half of the listeners ($n=16$), referred to as the trained listeners, participated in all of the sessions. The other half ($n=16$), referred to as controls, only participated in the pre- and post-test sessions. In each training session, listeners obtained 12 threshold estimates (see Sec. II D) on a single condition, referred to as the trained condition. In the pre- and post-tests, they obtained five threshold estimates on each of six conditions, one trained and five related untrained conditions. The condition order was randomized across listeners but fixed for each listener between the pre- and post-tests.

C. Task and conditions

For all stimulus conditions, the listeners were asked to discriminate between two sounds, presented through headphones, that differed only in their ILD or ITD value. Discrimination ability was measured in a two-interval-forced-choice procedure. In each trial, stimuli were presented in two visually marked 300-ms observation intervals that were separated by a 660-ms silent period. In one interval randomly chosen on each trial, a standard stimulus was presented. In the other interval, a signal stimulus was presented that differed from the standard stimulus only by a variable ΔILD or ΔITD that always favored the right ear (Wright and Fitzgerald, 2001; Zhang and Wright, 2007). The listeners reported which interval they perceived as containing the signal stimulus by pressing a key on a computer keyboard. Visual feedback was provided after each response throughout the entire experiment. Before starting each condition in the pretest, listeners were presented with samples of the standard and signal stimuli that, as they reported verbally or by pointing, produced distinct lateral positions. These samples were also provided before each 60-trial block throughout the experiment.

Here, we report data from seven stimulus conditions, one trained and six untrained. The trained condition, ILD discrimination with a 4-kHz tone SAM at 0.3 kHz and a 0-dB standard ILD, was included in the pre- and post-tests of all listeners. Each of the six untrained conditions included in this report was tested in only half of the listeners (eight trained listeners and eight controls per condition). These untrained conditions differed from the trained one either only in the standard ILD (6 vs 0 dB), the carrier frequency (6 vs 4 kHz), the modulation rate (0.15 vs 0.3 kHz), the stimulus type [pure tone (4 or 0.3 kHz) vs SAM tone], or the cue manipulated (ITD vs ILD). Four other untrained conditions were tested for a purpose unrelated to the current experiment and thus are not described here.

D. Procedure

From each 60-trial block, a discrimination threshold for ILD or ITD was estimated using a 3-down, 1-up adaptive procedure. The ΔILD or ΔITD value was decreased after every three consecutive correct responses and increased after each incorrect response (Levitt, 1971). The signal values at which the direction of change switched from decreasing to increasing or from increasing to decreasing were denoted as reversals. When there were seven or more reversals within a block, we discarded the first three reversals (if the total number of reversals was odd) or four reversals (if the total number of reversals was even) and averaged the remaining reversals to estimate the ΔILD or ΔITD value required for 79% correct responses (threshold). When there were fewer than seven reversals, performance on that block was marked as "insufficient reversals." In the ILD conditions, the starting value of the ΔILD was 6 dB, and the step size was 0.5 dB until the third reversal and 0.25 dB thereafter. In the ITD condition, the starting value of ΔITD was 1 μs, forcing the listeners to guess on the first trial (see also Wright and Fitzgerald, 2001; Zhang and Wright, 2007). The step size was multiplications or divisions by $10^{0.2}$ until the third reversal and by $10^{0.05}$ thereafter (Saberi, 1995). The minimum value was 0 dB for ΔILD and 1 μs for ΔITD. The maximum value was 650 μs for ΔITD, approximately the maximum naturally occurring time delay between the two ears in humans (e.g., Feddersen *et al.*, 1957; Kuhn and Guernsey, 1983). There was no maximum for ΔILD. We chose these parameters for the adaptive algorithms to be consistent with our previous experiments on interaural discrimination learning (Wright and Fitzgerald, 2001; Zhang and Wright, 2007). The differences in the parameters between the ILD and ITD conditions do not appear to influence the effectiveness of threshold estimation (Zhang and Wright, 2007).

Consistent with our previous experiment (Zhang and Wright, 2007) for the ITD-discrimination condition with the trained SAM tone, a substantial proportion (36%) of blocks failed to yield valid threshold estimates. Because omitting these blocks can lead to underestimation of ITD discrimination thresholds (Zhang and Wright, 2007), we reevaluated performance in the invalid blocks following the approach used in that report. Briefly, for tracks that yielded fewer than seven reversals (and hence were marked as insufficient reversals), we estimated thresholds based on the last four reversals when there were six total reversals and excluded the tracks yielding fewer than six reversals in total. When a track yielded a threshold estimate greater than 650 μs by calling for nominal ΔITD values exceeding 650 μs, we replaced the estimate with 650 μs.

E. Stimulus generation

The SAM tones were synthesized by sinusoidally modulating the amplitude of a sinusoidal carrier to 100% depth. In all conditions, the stimuli to both ears started and ended simultaneously, and each stimulus had a total duration of 300 ms, including 10-ms rise/fall cosine ramps. For pure tones, the starting phase of the right-ear stimulus was randomized across intervals. For SAM tones, the starting phases of both

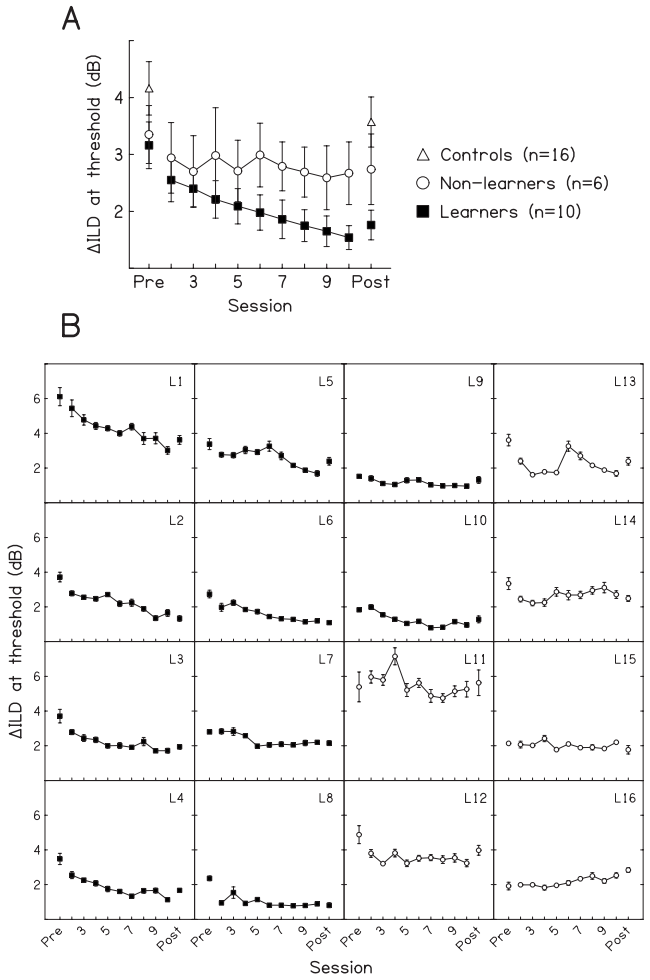


FIG. 1. Performance on the trained condition. (A) The mean ILD-discrimination thresholds for the 10 learners (filled squares), 6 non-learners (open circles), and 16 controls (open triangles) from the pretest, training sessions, and post-test. (B) Individual data for the ten learners (filled squares) and six non-learners (open circles). Error bars represent ±1 standard error (A) across or (B) within listeners.

the carrier and modulation waveforms to the right ear were randomized across intervals and were independent of each other. The nominal stimulus level was 50 dB SPL both for the pure tones and for SAM tones before modulation. This level was low enough to avoid the influence of combination products in the SAM-tone stimuli (Plomp, 1965). In the ILD conditions, the sound level was 50 dB SPL minus 0.5 times the desired ILD for the left-ear stimulus, and 50 dB SPL plus 0.5 times the desired ILD for the right-ear stimulus. There was no time or phase difference between the two ears. In the ITD condition, the desired ongoing ITD was set by delaying the starting phases of both the carrier and the modulator of the left-ear stimulus relative to those of the right-ear stimulus. There was no level difference between the two ears.

We used a digital-signal processing board (Tucker-Davis Technologies in Gainesville, Florida, AP2) to generate all stimuli. The stimuli to each ear were then delivered through separate 16-bit digital-to-analog converters (TDT DD1), anti-aliasing filters (8.5-kHz low-pass, TDT FT5), and programmable attenuators (TDT PA4). Finally, the stimuli were sent through a headphone buffer (TDT HB6) to headphones

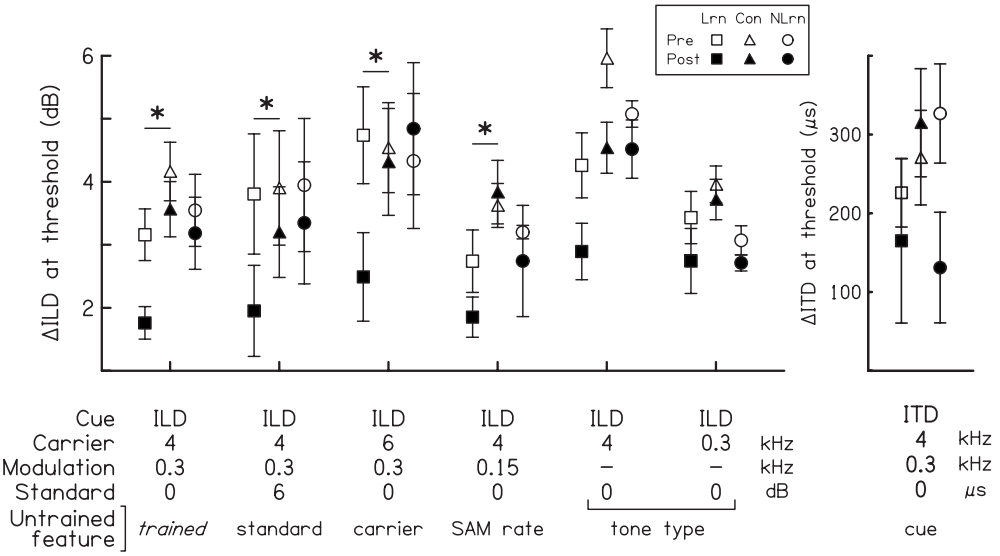


FIG. 2. Pre- and post-test performance on the trained and untrained conditions. The mean discrimination thresholds of the learners (squares), the non-learners (circles), and the controls (triangles) for both the pretest (open symbols) and post-test (filled symbols) in the six ILD conditions (left panel) and one ITD condition (right panel). The error bars represent ±1 standard error across listeners. Conditions are denoted on the abscissa by the interaural cue manipulated, the carrier frequency, the modulation rate, and the standard cue value. The trained condition is at the far left. For each untrained condition, the untrained feature is marked. Asterisks indicate that there was a significant difference in the amount of improvement from the pretest to the post-test between the learners and controls [$p<0.05$ for the group by time interaction in a two group (learners vs controls) by two time (pretest vs posttest) ANOVA].

with circumaural cushions (Sennheiser, HD265). Listeners were seated in a double-walled sound-attenuating booth.

III. RESULTS

A. Learning during the training sessions

Through multi-hour training, performance on the trained condition (ILD discrimination with a 4-kHz carrier SAM at 0.3 kHz and a 0-dB standard ILD) improved significantly in the majority of the trained listeners (Fig. 1). As a group, the trained listeners showed a significant decrease in their daily mean thresholds across the nine training sessions (not shown), as indicated by a repeated-measures one-way analysis of variance (ANOVA) ($p<0.001$) and a negative slope of a regression line fitted over training sessions (slope=−0.123 dB/session, $p=0.004$). However, the effect of training varied markedly across individual listeners. We identified an individual listener as having learned across training sessions only if that listener's performance met all of the following three criteria: (1) a significant one-way ANOVA on thresholds across training days without repeated measures, (2) a significant linear regression of thresholds fitted over training days, and (3) a negative slope of the regression. Alpha was set at 0.05 in all of the analyses. Using these criteria, the 16 trained listeners fell into two groups according to their performance during training. Ten out of the 16 listeners improved through training and are referred to as learners [Fig. 1(b), L1–L10, ANOVA: all $p\leq0.012$, regression: all $p\leq0.007$ and all slopes <0 dB/session]. The other six listeners did not meet the criteria and are referred to as non-learners [Fig. 1(b), L11–L16]. Among the six non-learners, one failed all three criteria (all p values ≥0.20, slope=0.785 dB/session), one failed both the criteria for ANOVA and for significant regression (all p values >0.24), three did not show a significant regression (all p values ≥0.38), and one did not have a significant ANOVA ($p=0.23$). The mean thresholds of the learners decreased steadily across all training sessions [Fig. 1(a), filled squares, regression: slope=−0.117 dB/session, $p=0.007$], while those of the non-learners remained at approximately the same level throughout training (open circles, regression: slope=−0.009 dB/session, $p=0.894$). Interestingly, the learners and non-learners had similar pretest thresholds (t test, $p=0.777$), so whether or not a listener would improve through training could not be predicted from starting performance. Because of their distinct patterns of behavior during the training sessions, we conducted the following analyses separately for the learners and non-learners (also see Wright *et al.*, 1997; Karmarkar and Buonomano, 2003).

B. Learning between the pre- and post-tests

On the trained condition, the learners improved significantly more than the controls between the pre- and post-tests (Fig. 2, the condition on the far left). The thresholds of the controls (triangles) decreased significantly between the pre- (open symbols) and post-tests (filled symbols), as revealed by a paired t test ($p=0.013$). However, the learners (squares) showed an even greater improvement, as indicated by a significant group by session interaction in a two group (learners vs controls) by two session (pretest vs post-test) ANOVA ($p=0.022$). Confirming this result, the pretest thresholds of the two groups did not differ significantly (independent t test, $p=0.151$), but the post-test thresholds were significantly lower in the learners than in the controls (independent t test, $p=0.006$). It is worth noting that, though not significantly different, the average pretest threshold of the learners (3.16 dB) was 1 dB lower than that of the controls (4.16 dB). Group differences in starting performance have been observed frequently in perceptual learning investigations with-

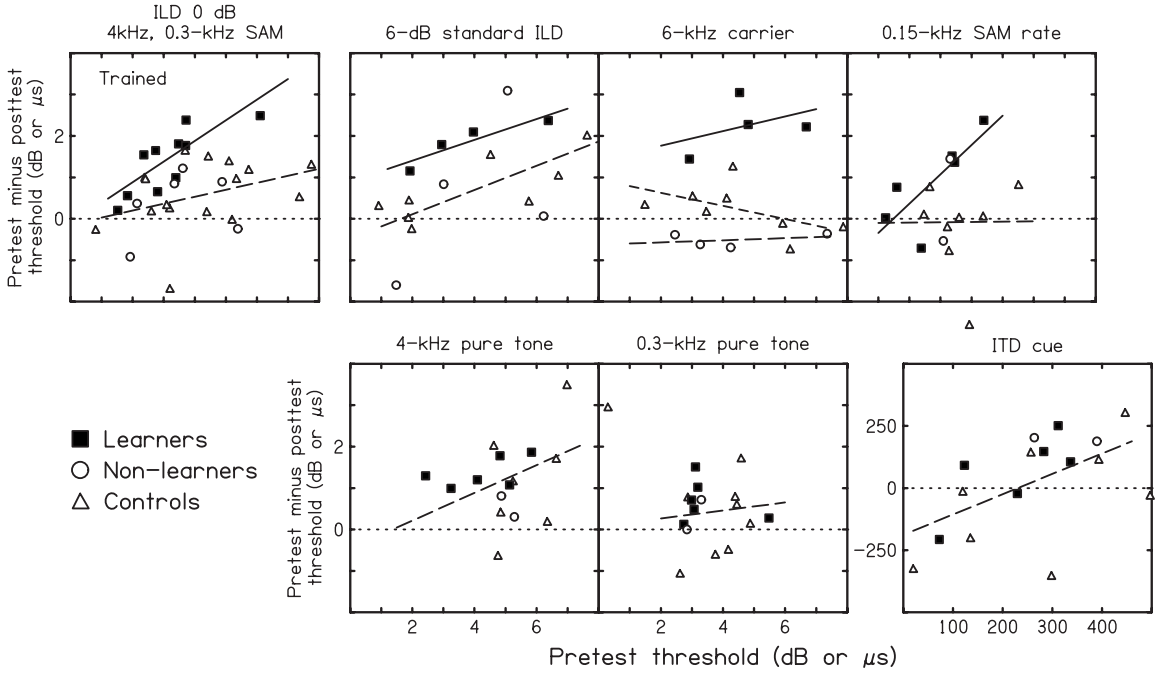


FIG. 3. The relationship between the pretest threshold and the amount of improvement. Pretest thresholds (abscissa) and pretest minus post-test thresholds (ordinate) for individual learners (filled squares), non-learners (open circles), and controls (open triangles) are shown for each of the seven tested conditions (panels). On each condition, the number of the regression lines fitted is based upon the results of between-group comparisons (see text): one line (all listeners, dashed lines), two lines (learners, solid lines; controls and non-learners, dashed lines), and three lines (learners, solid lines; non-learners, dashed lines; controls, short-dashed lines). The dotted lines represent zero improvement.

out consistent patterns or clearly identifiable causes (e.g., trained listeners started worse than controls: Wright and Fitzgerald, 2001; Fitzgerald and Wright, 2005; the opposite pattern: Mossbridge *et al.*, 2006). In the present case, the lower starting thresholds in the learners than in the controls actually strengthens the claimed effect of training because lower starting thresholds are associated with smaller training-induced improvements on this task (see below).

The learners also improved more than the controls on the untrained ILD-discrimination conditions with SAM stimuli but not on those with pure-tone stimuli or on the untrained ITD-discrimination condition. The controls (Fig. 2, triangles) showed significant learning between the pre-(open symbols) and post-tests (filled symbols) only on two of the six untrained conditions: the 6-dB standard ILD and the 4-kHz pure-tone conditions (paired t tests, $p \leq 0.038$; for other conditions, $p \geq 0.3$). Compared to the controls, the learners (squares) improved significantly more on all of the untrained ILD SAM conditions tested (Fig. 3, second to fourth condition from the left), including those with the untrained standard ILD (6 dB; ANOVA group by session interaction: $p=0.024$), the untrained carrier (6 kHz; $p<0.001$), and the untrained modulation rate (0.15 kHz; $p=0.086$, but $p=0.039$ according to an analysis of covariance that took into account the difference in pretest threshold between the two groups). For the remaining three untrained conditions (right three conditions), there was no difference between the learners and controls (all p values ≥ 0.304).

Unlike the learners, the non-learners (Fig. 2, circles) did not improve more than controls (triangles) between the pre-(open symbols) and post-tests (filled symbols) on any condition, trained or untrained ($p=0.038$ for the ILD 6-kHz carrier condition, but with more improvement in the controls than non-learners; $p \geq 0.202$ for the other conditions). Note, however, that on four out of the six untrained conditions, the analyses were based on only two non-learners and thus should be viewed as tentative. For example, the two non-learners both happened to improve (though neither significantly, independent-sample t tests, $p \geq 0.25$) between the pre- and post-tests on the ITD condition (Fig. 3, bottom right panel, circles) and appeared to show a trend of learning more than controls. However, considering the marked variability in the improvement of controls (Fig. 3, bottom right panel, triangles), the sample size of the non-learners was too small to support any specific conclusion.

Finally, the relationship between the amount of learning and pretest threshold varied across conditions, as well as across different listener groups. To examine the influence of pretest threshold on the amount of learning, we fitted regression lines to the amount of threshold improvement between the pre- and post-tests over the pretest threshold across individual listeners (Fig. 3). Based on the between-group comparisons reported above, on each condition, a separate line was fitted to each statistically different listener group. For the trained condition (Fig. 3, top left panel), because the statistical analyses revealed no difference between the non-learners (open circles) and controls (open triangles) but a difference between the learners (filled squares) and controls, we fitted a separate regression line for learners and one common line for the other two groups. The slope of the regression line for the learners was significantly different from zero (slope: 0.48; $p=0.006$; solid line), indicating that the magnitude of the improvement increased with increasing pretest threshold. In contrast, the slope of the regression line for

controls and non-learners did not differ significantly from zero ($p=0.104$; dashed line), indicating that the amount of improvement was approximately independent of the pretest value. For the untrained conditions, the magnitude of improvement tended to increase with increasing pretest threshold in general ($p \leq 0.085$), except in the cases of the 6-kHz carrier condition, the 0.3-kHz pure-tone condition, and the controls and non-learners in the 0.15-kHz SAM rate condition ($p \geq 0.168$).

IV. DISCUSSION

The primary purpose of the present experiment was to examine the extent to which AM influences ILD-discrimination learning. Here, we compare the current ILD-discrimination training results obtained with the SAM-tone (4-kHz carrier, 0.3-kHz AM) to those previously obtained with a pure tone (4 kHz) using the same training regimen (Wright and Fitzgerald, 2001) and discuss the implications of the differences between these results in terms of the neural processes governing ILD-discrimination ability. Before doing so, we first briefly discuss the lack of learning in a subset of the trained listeners (non-learners).

A. Non-learners

In the current experiment, a considerable portion of the trained listeners (6 out of 16, or 37%) failed to improve across training sessions. The lack of learning in a subset of trained listeners is not unusual. Large individual differences in improvement magnitude have been frequently observed in perceptual learning (e.g., Nagarajan *et al.*, 1998; Irvine *et al.*, 2000; Delhommeau *et al.*, 2002). In the auditory-learning investigations in which whether or not an individual listener improved over training was determined, the probability of learning across individual listeners varies considerably across tasks [e.g., 0%–50% for ITD discrimination (Wright and Fitzgerald, 2001; Rowan and Lutman, 2007; Zhang and Wright, 2007), 42%–79% for temporal-interval discrimination (Wright *et al.*, 1997; Karmarkar and Buonomano, 2003), 67%–86% for temporal-order discrimination (Mossbridge *et al.*, 2006; Mossbridge *et al.*, 2008), and 100% for SAM rate discrimination (Fitzgerald and Wright, 2005)] as well as across stimulus conditions for the same task (e.g., Karmarkar and Buonomano, 2003). To date, there is no systematic investigation into the cause of such individual differences in the ability to improve perceptually. In most cases, the listeners who failed to learn tended to have good starting performance, suggesting a ceiling effect for learning. However, in the current experiment, a ceiling effect could not explain the lack of learning in the non-learners because whether or not a listener improved through training could not be predicted from the starting performance. This homogeneity of starting performance of the learners and the non-learners also suggests that the non-learners did not fail to learn due to general factors that are likely to result in systematic changes in performance, such as an inability to maintain attention, confusion about the task, or the use of different strategies to solve the task (also see Sec. IV B 2 for further discussion). An absence of learning that could not be predicted from starting performance can also be seen on an auditory temporal-interval discrimination task in the individual-learning figure [Karmarkar and Buonomano, 2003, Fig. 1(b)]. Future experiments are needed to determine whether the lack of learning in individuals who have room to improve results from a general inability to improve perceptually or from task-dependent defects in cognitive or perceptual plasticity.

B. Differences between ILD-discrimination learning with pure and SAM tones

1. Learning patterns

While multi-hour training yielded significant improvement on ILD discrimination both with pure and SAM tones, the results differed between the two cases in the predictability, rate, and generalization pattern of that learning. First, whether or not a listener learned could be predicted from the starting performance when ILD discrimination was trained with the pure tone, but not with the SAM tone. With the pure tone, two out of eight trained listeners did not improve. These two listeners had starting thresholds among the best. With the SAM tone, however, as mentioned above, the six non-learners could not be distinguished from the learners based on pretest thresholds (Fig. 3, top left panel, open circles). Second, in those who learned, the rate of learning appeared to be slower with the SAM tone than with the pure tone, based on a two group (SAM trained vs pure-tone trained) by nine training session ANOVA (group by session interaction: $p=0.049$). With the pure tone [Fig. 4(a), hourglasses, data from Wright and Fitzgerald, 2001], performance improved rapidly in the first few sessions and leveled off in the following sessions, while with the SAM tone (squares), performance continued to improve at a constant rate throughout all of the nine sessions. Third, ILD-discrimination learning generalized across stimulus frequencies following training with the SAM tone, but not with the pure tone. With the pure tone, learning did not generalize from the trained 4-kHz tone to untrained tones at 6 or 0.5 kHz, while with the SAM tone, learning generalized from the trained 4-kHz carrier to the untrained 6-kHz carrier.

Another interesting difference between ILD-discrimination learning with the pure and SAM tones was that learning in one case (the SAM tone) did not generalize to the other (the pure tone). In the present experiment, the SAM-trained learners did not improve more than controls on ILD discrimination with untrained pure tones, either at the trained carrier frequency (4 kHz) or at the trained modulation rate (0.3 kHz), suggesting a failure of generalization that is attributable to stimulus type (SAM vs pure tone) rather than stimulus frequency. Strengthening this conclusion, for the ILD 4-kHz pure-tone condition, these learners, just like controls, improved significantly less than the listeners trained with the 4-kHz pure tone itself [Fig. 4(b), bottom panel; pure-tone trained data from Wright and Fitzgerald, 2001], according to a two group (SAM trained vs pure-tone trained) by two session (pretest vs posttest) ANOVA (group by session interaction, $p=0.034$). Thus, the current learners did not fail to generalize their learning to the pure tone at the trained carrier frequency because they had no room to improve.

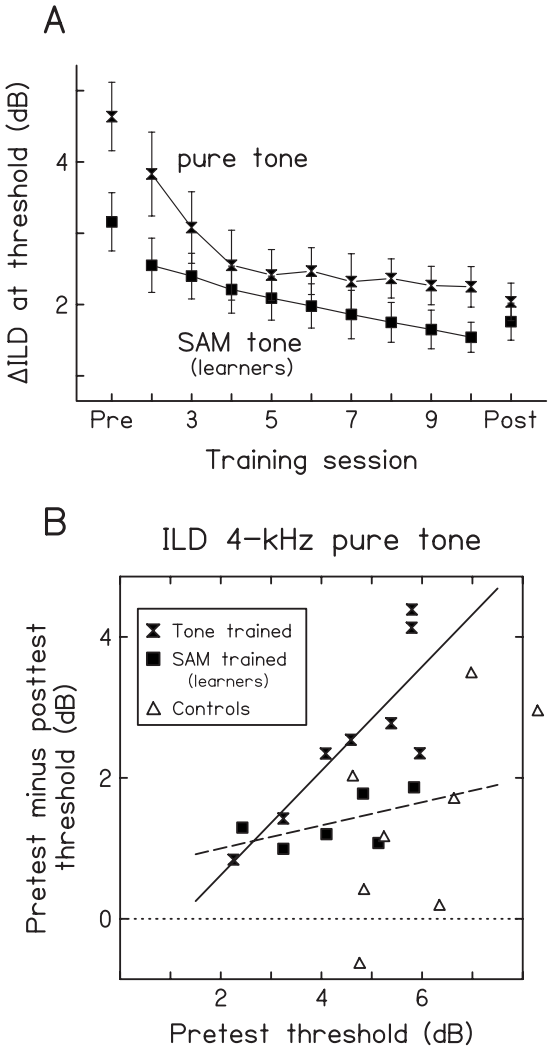


FIG. 4. Comparison between the current data and those previously obtained for ILD-discrimination training with a 4-kHz tone (Wright and Fitzgerald, 2001). (A) The mean ILD-discrimination thresholds across the test and training sessions for the present learners (n=6 of 8) trained with a 4-kHz tone SAM at 0.3 kHz (SAM tone, filled squares) and the previous listeners (n=8) trained with a 4-kHz tone (pure tone, filled hourglasses). (B) The amount of improvement (ordinate) on the ILD 4-kHz pure-tone condition plotted against the pretest threshold (abscissa) for each of the present SAM-trained learners (filled squares, n=6 of 8), the present controls (open triangles, n=8), and the previous listeners trained with the 4-kHz tone (filled triangles, n=8). Regression lines were fitted separately for the present learners (dashed line) and the previous listeners (solid line). Zero improvement is denoted by the dotted line.

These different learning and generalization patterns for ILD discrimination with the SAM and the pure tones suggest that the current multi-hour training regimen differentially influenced the neural system in the two cases. First, the neural processes affected by training with the SAM tone and the pure tone appear to have different characteristics. In the pure-tone case, the affected neural process seems to influence perception of different ILD values in a frequency-specific manner and to be readily modifiable unless its output is already near optimal. In the SAM-tone case, the modified neural process seems to influence ILD-discrimination performance in a stimulus-type-specific manner, responding to amplitude modulated tones regardless of location, carrier frequency, and modulation rate but not to pure tones even at the same frequency as the trained SAM tone. These different characteristics could result from either different neural circuitries being modified or from different modifications in the same neural circuitry. Further, the differences in the predictability and time course of learning observed with the two stimulus types also suggest different characteristics of the modifications themselves. With the pure tone, the modification took place readily with training, improving ILD sensitivity except in listeners with the best starting performance, and was completed relatively rapidly within a few training sessions. In contrast, with the SAM tone, the modification occurred with a probability of approximately 60%, independent of the initial performance, and proceeded at a nearly constant rate for more than ten days. These different characteristics of modifications hint at different types of neural changes resulting from training with the SAM tone and the pure tone.

Note that the characteristics described do not specify the physiological loci of the neural modifications incurred by training. Specificity of perceptual learning to stimulus features has often been taken as evidence that modifications occur in the early stages of sensory processing (Karni and Sagi, 1991; Poggio *et al.*, 1992; Ahissar and Hochstein, 2004; Fahle, 2004). Supporting this idea, neural changes accompanying learning have been identified in primary sensory cortices (Furmanski *et al.*, 2004; Clapp *et al.*, 2005; Li *et al.*, 2008; Pourtois *et al.*, 2008). However, it has also been proposed that changes in later stages of neural processing could also result in perceptual improvements that are specific to certain stimulus features through the reweighting of sensory information from different channels based on the task demand (Mollon and Danilova, 1996). Indeed, neural changes during perceptual learning have also been identified beyond the sensory cortices, including the associative (Law and Gold, 2008) and frontal (Krigolson *et al.*, 2009) cortices. Uniting the two lines of thought, there is increasing evidence that perceptual learning involves synergistic, dynamic responses of multiple systems. For example, perceptual training has been reported to induce changes spanning sensory, motor, associative, and cognitive systems (Vaina *et al.*, 1998; Schiltz *et al.*, 2001; Sigman *et al.*, 2005; van Wassenhove and Nagarajan, 2007). Further, the neural modifications are dynamic in that different sites can be affected at different time points in training (Karni *et al.*, 1998; Petersen *et al.*, 1998; Atienza *et al.*, 2002; Gottselig *et al.*, 2004) and changes at some loci can be reversed when learning is complete (Vaina *et al.*, 1998; Yotsumoto *et al.*, 2008).

2. Possible explanations

The differences in ILD-discrimination learning observed in the pure- and SAM-tone experiments appear to be attributable to the differences in the amplitude envelope shape, rather than in the pitch, overall stimulus level, or incongruence between the two cues in these two stimulus types. Though the pure and SAM tones used in training differed in perceived pitch, this difference does not appear to account for the different ILD-discrimination learning patterns with these two stimulus types. The pitch of a pure tone corresponds well with the stimulus frequency, while that for a

SAM tone is close to the modulation rate and changes little with the carrier frequency (for an overview, see Moore, 1997). Therefore, the two trained stimuli used in the present and previous ILD-discrimination training experiments, though they shared the same central frequency (4 kHz), differed widely in pitch (4 kHz for the pure tone and ~0.3 kHz for the SAM tone). Thus, the different learning patterns for the pure and SAM tones might be attributed to the differences in their pitches. In the pure-tone experiment, the frequency specificity of ILD-discrimination learning can be readily translated into pitch specificity. However, in the SAM-tone experiment, learning was generalized to both an untrained carrier frequency (with a pitch similar to the trained one) and an untrained SAM rate (and hence an untrained pitch), but not to pure tones at either the trained carrier frequency (with an untrained pitch) or the trained SAM rate (with a pitch similar to the trained one). Thus, learning patterns differed between the two stimulus types even when the pitches were taken into consideration.

Similarly, it appears that the different learning patterns did not result from the different stimulus levels that were used in the pure- and SAM-tone experiments. The nominal stimulus level was 70 dB SPL (sound pressure level) in the previous pure-tone ILD training experiment (Wright and Fitzgerald, 2001) but was only 50 dB SPL in the present SAM-ILD training experiment (to avoid combination products, see Sec. II). This difference conceivably could result in different patterns of modification by the same training paradigm because overall stimulus level has been reported to influence the responses of ILD sensitive neurons along the ascending ILD pathway (Semple and Kitzes, 1987; Irvine and Gago, 1990; Irvine *et al.*, 1996; Park *et al.*, 2004). However, this account cannot readily explain the lack of generalization from the trained SAM tone to the pure tones in the current experiment, in which all of the stimuli were presented at the same level.

Another difference between the two stimulus types, the potential conflict between the ILD and ITD cues, also seems unlikely to be the cause of the different learning patterns observed in the two experiments. In natural listening environments, ILD and ITD values co-vary with the sound source location and thus are congruent with each other. However, in the ILD- and ITD-discrimination tasks that have been used to separately investigate ILD and ITD sensitivity, the ILD and ITD values are manipulated independently, creating unnatural situations in which the two cues may not indicate the same sound source location. While there was cue incongruence in the ILD-discrimination training with both the SAM and the pure tones, the influence of this incongruence on performance may have differed due to the difference in human sensitivity to ITDs in the two types of stimuli. For the 4-kHz pure tone used in the previous training experiment, cue incongruence was unlikely to have influenced performance because listeners are not sensitive to ITDs in high-frequency pure tones and therefore might have solved the task using only the ILD cue. Thus, the performance improvement may faithfully reflect improved ILD sensitivity. In contrast, for the 4-kHz SAM tone in the current experiment, in which the perceived sound position varies with both ITDs and ILDs, incongruence between the two cues might have caused the perceived sound image to be diffused or sometimes even split (e.g., Hafter and Jeffress, 1968). Thus, multiple strategies of solving the task might have been available to the listeners during training with the SAM tone and consequently could have influenced the learning pattern. For example, the non-learners might have been prevented from learning by confusion caused by the different locations indicated by the two cues, and the learners, instead of having learned the ILD cue itself, might have learned to ignore the ITD cue or even to use image diffuseness to solve the task. If so, the performance improvement with the SAM tone would reflect an improved ability of the neural system to exclude conflicting information or to make use of the diffuseness of the sound image rather than reflect better ILD sensitivity. However, our previous training on ITD discrimination with the same SAM tone, which presumably presented the same level of cue conflict, yielded no learning in any of the nine trained listeners (Zhang and Wright, 2007). Thus, the current learning pattern for ILD discrimination with the SAM tone differs both from that of ILD discrimination with a pure tone at the same frequency (a stimulus for which listeners are sensitive only to ILDs and hence are unlikely to be influenced by cue incongruence) and from that of ITD discrimination with the same SAM tone (a stimulus for which listeners are sensitive to both cues and hence are likely subject to possible influences of cue incongruence). This observation suggests that the effect of the current training regimen varies both with the cue manipulated (ILD vs ITD) and with the stimulus type employed (SAM vs pure tone), instead of with the presence or absence of conflicting cues.

We instead suggest that the different learning patterns of ILD discrimination with pure and SAM tones reflect the influence of AM on ILD sensitivity. While it is possible to explain these different patterns by assuming that there are separate channels of ILD processing for pure tones and AM sounds, we do not favor this view for two reasons. First, to date, we are aware of no report of ILD sensitive neurons that are activated by AM stimuli but not by pure tones, or vice versa. Second, because most naturally occurring sounds are amplitude modulated, it seems unlikely that separate neural resources should be dedicated to pure tones. Rather, we propose that the temporal fluctuations in the amplitude envelope of high-frequency stimuli play an active role in ILD processing. Current ILD-processing models typically imply that these fluctuations are smoothed out and have little influence on ILD sensitivity (e.g., the level-meter model by Hartmann and Constan, 2002). However, this assumption has not been thoroughly tested, leaving open the possibility that envelope fluctuations do influence ILD encoding. Given this possibility, the improvements on ILD discrimination in the SAM-tone experiment may have resulted from modifications in the extraction of the amplitude envelope itself and/or in the transmission of the extracted envelope to ILD encoding. This proposal appears feasible based on neurophysiological data. In many high-frequency, ILD-responsive neurons in the brainstem, fluctuations in the stimulus amplitude envelope are faithfully preserved, or even enhanced (Joris and Yin, 1995), and the brainstem nuclei that are typically thought to

play a crucial role in ILD processing are also regarded as contributing to AM processing (Brugge *et al.*, 1993; Joris *et al.*, 2004). In contrast, we suggest that the improvements in the pure-tone experiment resulted from modifications in processes unrelated to AM processing. Supporting this idea, the specificity of the pure-tone learning to the stimulus frequency but not to the standard ILD value is consistent with reports that ILD sensitive neurons at several stages of ILD processing, including the lateral superior olive, inferior colliculus, and primary auditory cortex, respond in a frequency-specific manner to a broad range of ILD values (for review, see Ehret and Romand, 1997; Park *et al.*, 2004). Note that the proposed modifications for both stimulus types could have occurred either in quite early stages of ILD processing or in later stages that adjust or interpret the output of the initial encoding stages. The latter possibility is particularly plausible for the SAM-tone case because the affected neural process distinguishes between SAM tones and pure tones even at the same frequency, a feature that is more often demonstrated by late than early stages of auditory processing.

One specific manner in which AM processing might affect ILD sensitivity is through the adjustment of the strength of envelope phase-locking in ILD encoding. High-frequency ILDs are initially computed by brainstem neurons that cannot phase lock to the fine structure of high-frequency sounds but can do so to low-frequency AM of these sounds (up to around 0.5 kHz AM, Joris and Yin, 1995). If training with the SAM tone, through either bottom-up or top-down mechanisms, were to strengthen phase-locking to the peaks of the fluctuating amplitude envelope, where the difference in the sound level from the two ears is largest, ILD sensitivity would be enhanced. This benefit would not be present for pure tones. According to this scenario, though the modifications induced by ILD-discrimination training with the SAM tone may be present in neural circuitry that is activated by both AM sounds and pure tones, behavioral benefits would occur only for AM sounds, as we observed in the present experiment.

C. Differences between ILD- and ITD-discrimination learning

The current results, when added to the previous ones obtained with the same training regimen for ILD and ITD discriminations with different stimuli (Wright and Fitzgerald, 2001; Zhang and Wright, 2007), confirm the notion that learning of lateralization was determined more by the cue manipulated than by any other aspect of the stimulus. For ITD discrimination, after the ~2 h pretest, listeners did not benefit from further multi-hour training with either a pure tone or a SAM tone (0.5-kHz tone, Wright and Fitzgerald, 2001; 4-kHz tone SAM at 0.3 kHz, Zhang and Wright, 2007). In contrast, the multi-hour training on ILD discrimination yielded significant additional learning in the majority of the trained listeners with both stimulus types (4-kHz tone, Wright and Fitzgerald, 2001; 4-kHz tone SAM at 0.3 kHz, the present experiment). In other words, under the current training paradigm, with both pure and SAM tones, ITD discrimination performance reached asymptote by the end of the ~2 h pretest (Wright and Fitzgerald, 2001; Zhang and Wright, 2007), while ILD performance continued to improve in most listeners (Wright and Fitzgerald, 2001 and the current experiment), though with a longer time course for the SAM than the pure tone. Notably, this distinction between ILD and ITD discriminations held even when the same standard stimulus (the SAM tone) was used during training. It remains to be resolved whether the lack of learning in a subset of the listeners trained on ILD discrimination with the SAM tone (the current experiment) resulted from the same cause as the lack of learning on ITD discrimination with both stimulus types (Wright and Fitzgerald, 2001; Zhang and Wright, 2007). The distinct effects of multi-hour training on ILD and ITD discriminations are consistent with the idea of differential neural processing of ILDs and ITDs in humans. It appears that under the current training regimen, the neural mechanisms that determine ITD discrimination sensitivity are less modifiable in long term than those underlying ILD-discrimination sensitivity and that this difference in modifiability lies between the two cues rather than between different frequency regions or different amplitude envelope shapes.

V. CONCLUSIONS

We examined the effect of multi-hour training on ILD discrimination with a 4-kHz tone SAM at 0.3 kHz. Ten out of the 16 trained listeners improved more than untrained controls. Compared to previous results of ILD-discrimination training with a 4-kHz pure tone (Wright and Fitzgerald, 2001), learning with the SAM tone showed less predictability of occurrence based on starting performance, a longer time course, and specificity to stimulus type (amplitude modulated tones vs pure tones) rather than stimulus frequency. Among several other possibilities, we suggest that the differences between the two ILD-discrimination training results indicate that the sound-localization system has the ability to access amplitude envelope information in a sound and use that information to improve ILD representation. This ability, if confirmed, would represent an under investigated type of interaction between temporal and level processing in the binaural system.

ACKNOWLEDGMENTS

We thank Rodrigo Cadiz for technical support and Karen Banai, Julia Huyck, Nicole Marrone, Julia Mossbridge, Jeanette Ortiz, and Andrew Sabin for helpful comments on previous drafts of this paper. This work is supported by NIH/NIDCD and The Hugh Knowles Center for Clinical and Basic Science in Hearing and Its Disorders.

Ahissar, M., and Hochstein, S. (**2004**). "The reverse hierarchy theory of visual perceptual learning," Trends Cogn. Sci. **8**, 457–464.

Atienza, M., Cantero, J. L., and Dominguez-Marin, E. (**2002**). "The time course of neural changes underlying auditory perceptual learning," Learn. Memory **9**, 138–150.

Brugge, J. F., Blatchley, B., and Kudoh, M. (**1993**). "Encoding of amplitude-modulated tones by neurons of the inferior colliculus of the kitten," Brain Res. **615**, 199–217.

Clapp, W. C., Kirk, I. J., Hamm, J. P., Shepherd, D., and Teyler, T. J. (**2005**). "Induction of LTP in the human auditory cortex by sensory stimulation," Eur. J. Neurosci. **22**, 1135–1140.

Delhommeau, K., Micheyl, C., Jouvent, R., and Collet, L. (**2002**). "Transfer of learning across durations and ears in auditory frequency discrimination," Percept. Psychophys. **64**, 426–436.
Ehret, G., and Romand, R. (**1997**). *The Central Auditory System* (Oxford University Press, New York).
Fahle, M. (**2004**). "Perceptual learning: A case for early selection," J. Visualization **4**, 879–890.
Feddersen, W. E., Sandel, T. T., Teas, D. C., and Jeffress, L. A. (**1957**). "Localization of high-frequency tones," J. Acoust. Soc. Am. **29**, 988–991.
Fitzgerald, M. B., and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise," J. Acoust. Soc. Am. **118**, 3794–3803.
Furmanski, C. S., Schluppeck, D., and Engel, S. A. (**2004**). "Learning strengthens the response of primary visual cortex to simple patterns," Curr. Biol. **14**, 573–578.
Gottselig, J. M., Brandeis, D., Hofer-Tinguely, G., Borbely, A. A., and Achermann, P. (**2004**). "Human central auditory plasticity associated with tone sequence learning," Learn. Memory **11**, 162–171.
Hafter, E. R. (**1984**). "Spatial hearing and the duplex theory: How viable is the model?," in *Dynamic Aspects of Neucortical Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Neurosciences Research Foundation, John Wiley and Sons, New York), pp. 425–448.
Hafter, E. R., and Jeffress, L. A. (**1968**). "Two-image lateralization of tones and clicks," J. Acoust. Soc. Am. **44**, 563–569.
Hartmann, W. M., and Constan, Z. A. (**2002**). "Interaural level differences and the level-meter model," J. Acoust. Soc. Am. **112**, 1037–1045.
Irvine, D. R., and Gago, G. (**1990**). "Binaural interaction in high-frequency neurons in inferior colliculus of the cat: Effects of variations in sound pressure level on sensitivity to interaural intensity differences," J. Neurophysiol. **63**, 570–591.
Irvine, D. R., Martin, R. L., Klimkeit, E., and Smith, R. (**2000**). "Specificity of perceptual learning in a frequency discrimination task," J. Acoust. Soc. Am. **108**, 2964–2968.
Irvine, D. R., Rajan, R., and Aitkin, L. M. (**1996**). "Sensitivity to interaural intensity differences of neurons in primary auditory cortex of the cat. I. Types of sensitivity and effects of variations in sound pressure level," J. Neurophysiol. **75**, 75–96.
Irvine, D. R., and Wright, B. A. (**2005**). "Plasticity of spectral processing," Int. Rev. Neurobiol. **70**, 435–472.
Joris, P. X., Schreiner, C. E., and Rees, A. (**2004**). "Neural processing of amplitude-modulated sounds," Physiol. Rev. **84**, 541–577.
Joris, P. X., and Yin, T. C. T. (**1995**). "Envelope coding in the lateral superior olive. I. Sensitivity to interaural time differences," J. Neurophysiol. **73**, 1043–1062.
Karmarkar, U. R., and Buonomano, D. V. (**2003**). "Temporal specificity of perceptual learning in an auditory discrimination task," Learn. Memory **10**, 141–147.
Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., and Ungerleider, L. G. (**1998**). "The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex," Proc. Natl. Acad. Sci. U.S.A. **95**, 861–868.
Karni, A., and Sagi, D. (**1991**). "Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity," Proc. Natl. Acad. Sci. U.S.A. **88**, 4966–4970.
Koehnke, J., Colburn, H. S., and Durlach, N. I. (**1986**). "Performance in several binaural-interaction experiments," J. Acoust. Soc. Am. **79**, 1558–1562.
Krigolson, O. E., Pierce, L. J., Holroyd, C. B., and Tanaka, J. W. (**2009**). "Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise," J. Cogn. Neurosci. **21**, 1834–1841
Kuhn, G. F., and Guernsey, R. M. (**1983**). "Sound pressure distribution about the human head and torso," J. Acoust. Soc. Am. **73**, 95–105.
Law, C. T., and Gold, J. I. (**2008**). "Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area," Nat. Neurosci. **11**, 505–513.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Li, W., Piech, V., and Gilbert, C. D. (**2008**). "Learning to link visual contours," Neuron **57**, 442–451.
Mollon, J. D., and Danilova, M. V. (**1996**). "Three remarks on perceptual learning," Spatial Vis. **10**, 51–58.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, London).
Mossbridge, J. A., Fitzgerald, M. B., O'Connor, E. S., and Wright, B. A. (**2006**). "Perceptual-learning evidence for separate processing of asynchrony and order tasks," J. Neurosci. **26**, 12708–12716.
Mossbridge, J. A., Scissors, B. N., and Wright, B. A. (**2008**). "Learning and generalization on asynchrony and order tasks at sound offset: Implications for underlying neural circuitry," Learn. Memory **15**, 13–20.
Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., and Merzenich, M. M. (**1998**). "Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality," J. Neurosci. **18**, 1559–1570.
Park, T. J., Klug, A., Holinstat, M., and Grothe, B. "Interaural level difference processing in the lateral superior olive and the inferior colliculus," J. Neurophysiol. **92**, 289–301 (**2004**).
Petersen, S. E., van Mier, H., Fiez, J. A., and Raichle, M. E. (**1998**). "The effects of practice on the functional anatomy of task performance," Proc. Natl. Acad. Sci. U.S.A. **95**, 853–860.
Plomp, R. (**1965**). "Detectability threshold for combination tones," J. Acoust. Soc. Am. **37**, 1110–1123.
Pourtois, G., Rauss, K. S., Vuilleumier, P., and Schwartz, S. (**2008**). "Effects of perceptual learning on primary visual cortex activity in humans," Vision Res. **48**, 55–62.
Rowan, D., and Lutman, M. E. (**2006**). "Learning to discriminate interaural time differences: An exploratory study with amplitude-modulated stimuli," Int. J. Audiol. **45**, 513–520.
Rowan, D., and Lutman, M. E. (**2007**). "Learning to discriminate interaural time differences at low and high frequencies," Int. J. Audiol. **46**, 585–594.
Saberi, K. (**1995**). "Some considerations on the use of adaptive methods for estimating interaural-delay thresholds," J. Acoust. Soc. Am. **98**, 1803–1806.
Schiltz, C., Bodart, J. M., Michel, C., and Crommelinck, M. (**2001**). "A pet study of human skill learning: Changes in brain activity related to learning an orientation discrimination task," Cortex **37**, 243–265.
Semple, M. N., and Kitzes, L. M. (**1987**). "Binaural processing of sound pressure level in the inferior colliculus," J. Neurophysiol. **57**, 1130–1147.
Sigman, M., Pan, H., Yang, Y., Stern, E., Silbersweig, D., and Gilbert, C. D. (**2005**). "Top-down reorganization of activity in the visual pathway after learning a shape identification task," Neuron **46**, 823–835.
Stern, R. M. Jr., Slocum, J. E., and Phillips, M. S. (**1983**). "Interaural time and amplitude discrimination in noise," J. Acoust. Soc. Am. **73**, 1714–1722.
Tollin, D. J. (**2003**). "The lateral superior olive: A functional role in sound source localization," Neuroscientist **9**, 127–143.
Vaina, L. M., Belliveau, J. W., des Roziers, E. B., and Zeffiro, T. A. (**1998**). "Neural systems underlying learning and representation of global motion," Proc. Natl. Acad. Sci. U.S.A. **95**, 12657–12662.
van Wassenhove, V., and Nagarajan, S. S. (**2007**). "Auditory cortical plasticity in learning to discriminate modulation rate," J. Neurosci. **27**, 2663–2672.
Watson, C. S. (**1980**). "Time course of auditory perceptual learning," Ann. Otol. Rhinol. Laryngol. Suppl. **89**, 96–102.
Wright, B. A., Buonomano, D. V., Mahncke, H. W., and Merzenich, M. M. (**1997**). "Learning and generalization of auditory temporal-interval discrimination in humans," J. Neurosci. **17**, 3956–3963.
Wright, B. A., and Fitzgerald, M. B. (**2001**). "Different patterns of human discrimination learning for two interaural cues to sound-source location," Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312.
Wright, B. A., and Zhang, Y. (**2006**). "A review of learning with normal and altered sound-localization cues in human adults," Int. J. Audiol. **45**, 92–98.
Wright, B. A., and Zhang, Y. (**2009**). "Insights into human auditory processing gained from perceptual learning," in *The Cognitive Neurosciences IV*, edited by M. S. Gazzaniga (The MIT Press, Cambridge, Mass.).
Yost, W. A., and Dye, R. H. Jr., (**1988**). "Discrimination of interaural differences of level as a function of frequency," J. Acoust. Soc. Am. **83**, 1846–1851.
Yotsumoto, Y., Watanabe, T., and Sasaki, Y. (**2008**). "Different dynamics of performance and brain activation in the time course of perceptual learning," Neuron **57**, 827–833.
Zhang, Y., and Wright, B. A. (**2007**). "Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies," J. Acoust. Soc. Am. **121**, 2207–2216.

Prediction of binaural speech intelligibility with frequency-dependent interaural phase differences

Rainer Beutelmann,[a] Thomas Brand, and Birger Kollmeier
Medizinische Physik, Carl-von-Ossietzky-Universität Oldenburg, 26111 Oldenburg, Germany

(Received 17 May 2009; revised 12 June 2009; accepted 16 June 2009)

The aim of this study was to test the hypothesis of independent processing strategies in adjacent binaural frequency bands underlying current models for binaural speech intelligibility in complex configurations and to investigate the effective binaural auditory bandwidth in broad-band signals. Speech reception thresholds (SRTs) were measured for binaural conditions with frequency-dependent interaural phase differences (IPDs) of speech and noise. SRT predictions with the binaural speech intelligibility model by Beutelmann and Brand (2006, J. Acoust. Soc. Am. **120**, 331–342) were compared with the observed data. The IPDs of speech and noise had a sinusoidal shape on a logarithmic frequency scale. The bandwidth between zeros of the IPD function was varied from 1/8 to 4 octaves. Speech and noise had either the same IPD function (reference condition) or opposite signs of the IPD function (binaural condition). Each condition had two subconditions with alternating and non-alternating signs, respectively, of the IPD function. The binaural unmasking with respect to the reference condition decreased from 6 dB to zero with decreasing IPD bandwidth for the alternating condition while it stayed significantly larger than zero for the non-alternating condition. The observed results were well predicted by the model with an analysis filter bandwidth of 2.3 equivalent rectangular bandwidths (ERBs).
 [DOI: 10.1121/1.3177266]

PACS number(s): 43.66.Pn, 43.71.An, 43.66.Ba [RLF] Pages: 1359–1368

I. INTRODUCTION

Binaural hearing plays an important role in solving the "cocktail party problem" (Cherry, 1953), a term used for the task of understanding speech in complex environments. The classical way of quantifying this effect is the binaural masking level difference (BMLD) which is the difference in threshold level for the detection of a pure tone target in noise between a binaural condition and a (typically diotic) reference condition. When trying to predict binaural speech intelligibility on the basis of binaural tone detection experiments, however, it is noteworthy that the latter typically employ a narrow-band target signal, while in most speech intelligibility experiments both the target and the interferer are broadband. It has been shown that BMLDs at different frequencies are a good predictor for the frequency-dependent effective signal-to-noise-ratio (SNR) enhancement for speech in noise (vom Hövel, 1984; Zurek, 1990; Culling *et al.*, 2004; Beutelmann and Brand, 2006). However, it is not entirely clear whether concurrent binaural cues at different frequencies are integrated or processed separately in speech intelligibility tests.

This study is concerned with questions that arise from the division of broad-band speech and noise signals into narrow frequency bands. It is well known that the masked threshold of a tone centered in a noise band is primarily affected by the masker energy within a certain frequency range around the target tone (Fletcher, 1940), which is referred to as the critical bandwidth. If the overall noise level is kept constant, the threshold remains constant for noise bandwidths below the critical bandwidth and decreases if the noise bandwidth exceeds the critical bandwidth. If the power spectral density of the noise is kept constant, the threshold increases for increasing noise bandwidth up to the critical bandwidth and stays constant for higher bandwidths. These effects of the critical bandwidth are usually modeled in form of band pass filters on the signals, the "auditory" filters (e.g., Patterson, 1976; Glasberg and Moore, 1990). Although this concept of the critical bandwidth and integration of auditory processes within a certain, finite frequency range is generally accepted, it is not clear whether (1) the effective critical bandwidths for binaural processes are different from their monaural counterparts and (2) the hypothesis of independent binaural processing in adjacent or remote auditory frequency bands is true. This study has the intention of assessing the relevance of these questions for binaural speech intelligibility prediction models. Both items are related because a large effective bandwidth also affects the independence of adjacent frequency regions.

A. Binaural bandwidth using narrow-band targets

The effective bandwidth in binaural tone detection has been measured using various methods, resulting in different estimates of binaural bandwidth. The most common method was to measure tone detection thresholds in noises with different bandwidths centered around the target tone (Wightman, 1971; Sever and Small, 1979; Hall *et al.*, 1983; Cokely and Hall, 1991). The effective critical bandwidths measured for an antiphasic tone in homophasic noise (dichotic condition) appear to be 1.5–4 times larger than the effective criti-

[a]Author to whom correspondence should be addressed. Electronic mail: rainer.beutelmann@uni-oldenburg.de

cal bandwidths measured with homophasic tone and noise (diotic condition). In the dichotic conditions, the effective critical bandwidth is furthermore dependent on the noise power spectral density and increases with increasing level (Hall *et al.*, 1983). In broad-band conditions, however, the differences between the effective monaural and binaural bandwidths are much smaller.

Hall *et al.* (1983) also measured the critical bandwidth with a notch of variable width in broad-band noise, centered on the target tone. In the diotic case, the notched-noise and bandlimiting critical bandwidths are about the same, but in the dichotic case, the effective critical bandwidth measured with the notched-noise paradigm is considerably lower than with the bandlimiting paradigm. Nitschmann and Verhey (2007) presented a successful approach which was able to model the different results of the notched-noise and the bandlimiting paradigms using weighted sums of neighboring auditory filters and thus increasing the effective binaural bandwidth.

Sondhi and Guttman (1966) and Holube *et al.* (1998) used noise spectra that were broad-band and flat, but the interaural phase was inverted in a rectangular region of variable width centered around the target tone. The interaural phase was either the same as the interaural phase of the target tone (0 or π), or it was the opposite. The estimated effective binaural bandwidth depended on the assumed filter shape and on the fitting method, but it was significantly larger for conditions with a phase difference between target and on-frequency noise band than in the conditions in which the target and on-frequency noise band had the same interaural phase difference (IPD).

Another method for measuring the binaural bandwidth involves presenting a single, sharp transition of the IPD between 0 and π in an otherwise flat-spectrum, broad-band noise. Kohlrausch (1988) varied the target tone frequency and thus the influence of the interaural phase edge on the detection of the target tone. Kollmeier and Holube (1992) varied the edge frequency while keeping the target tone frequency fixed.

Holube *et al.* (1998) used another paradigm similar to the one used by Houtgast (1977) for monaural auditory filters. The interaural correlation was changed sinusoidally with frequency, and the detection thresholds were measured as a function of the periodicity in the (linear) frequency domain. The estimated effective binaural critical bandwidths were larger than the ones estimated from the rectangular and stepwise interaural correlation changes in the same study.

Kohlrausch (1988) concluded that the effective peripheral critical bandwidth for binaural processes might not be larger than the monaural critical bandwidth, but that the effects found in binaural bandwidth experiments are a consequence of different detection mechanisms for monaural and binaural hearing. A similar conclusion was drawn by Kollmeier and Holube (1992), although in this study there was a significant difference in binaural and monaural bandwidth by a factor of 1.2. They furthermore pointed out that the estimate of the bandwidth is critically dependent on the filter shape and in which way the bandwidth of the respective filter shape is calculated. The hypothesis that monaural and binaural processing are not subject to different underlying critical bandwidths is supported by van der van der Heijden and Trahiotis (1998). They performed antiphasic pure tone detection in noise of different bandwidths and interaural correlations and concluded that the different apparent critical bandwidths arise from differences in the tasks, but are based on the same underlying critical bandwidth. The same reasoning may also be applied to the differences in the tasks of tone detection and speech intelligibility. However, no need for different bandwidths was found so far.

There have been different approaches toward modeling the potential increased binaural critical bandwidth. Metz *et al.* (1968) included a bandwidth dependence in the binaural processing errors of the equalization-cancellation (EC) model (Durlach, 1963) in order to accommodate the noise band-width dependence of binaural detection thresholds. Sondhi and Guttman (1966) found that a different relation between the interaural cross-correlation function and the BMLD than in the original EC theory would be needed in order to predict the data from experiments in their study with rectangularly inverted spectral phase. The binaural model of Breebaart *et al.* (2001), however, was able to explain the wider binaural bandwidth of bandlimiting experimental paradigms without explicit adjustment of the model parameters quite well.

B. Binaural bandwidth using broad-band/speech targets

While the so far mentioned studies concern the effective binaural bandwidth, the hypothesis of independent binaural processing channels has been examined in studies with multiple target tones or speech with frequency-dependent interaural phase or time differences: Akeroyd (2004) showed that detection thresholds of multi-component tone complexes of up to 17 components stretching from 200 Hz to 1 kHz in broad-band, white noise were the same for S_0N_{180}, $S_{180}N_0$, and $S_{270}N_{90}$, where the index of S denotes the target IPD in degrees and the index of N denotes the noise IPD. If the binaural system was constrained to eliminate only noise with a single interaural time difference (ITD) across all frequencies, the thresholds would have been different.

There are also studies which use speech or speech-like sounds as targets in binaural experiments. While the processing of interaural level differences (ILDs) generated by swapping high- and low-frequency bands of target and interferer between the ears seems to be dominated by the ear with the better SNR (Edmonds and Culling, 2006), there is evidence from speech intelligibility experiments that it is possible to process different ITDs and IPDs in high- and low-frequency regions separately (Culling and Summerfield, 1995; Edmonds and Culling, 2005). This is apparently true as long as the binaural cues are not needed for localization and subsequent streaming of different auditory objects (Best *et al.*, 2007). For speech in stationary noise without further speech-like distractors, this should be the case because the harmonicity of speech sounds is a stronger cue than spatial location (Buell and Hafter, 1991).

The present experiments explored a binaural speech intelligibility model developed by Beutelmann and Brand

(2006). It uses a gammatone filter bank (Hohmann, 2002) to split the input signals into auditory equivalent rectangular bandwidth (ERB)[1] wide frequency bands. In each frequency band, the maximally possible SNR enhancement due to interaural differences is calculated using the EC principle proposed by Durlach (1963) with error parameters adapted from vom Hövel (1984). The equalization parameters are independent of each frequency band, but the gammatone filters overlap to a large extent, and so the processing is not completely independent between frequency bands. Finally, a speech reception threshold (SRT) is calculated from the band-wise speech and noise levels with the help of the speech intelligibility index (SII, ANSI, 1997). This model has yielded good SRT predictions for a single, stationary noise source at various azimuths and in different room acoustics, and a simple extension for modulated interferers also yielded promising results (Beutelmann and Brand, 2006).

The model of Beutelmann and Brand (2006) used gammatone filters whose bandwidth was set to the standard values from monaural experiments. However, given the abovementioned discrepancies as to what the binaural bandwidth actually is, we made new experimental measurements using conditions with strongly frequency-dependent IPDs and compared them to model predictions calculated for a large range of bandwidths. The paradigm chosen was based on Houtgast (1977) and Holube *et al.* (1998), in which both the IPDs of the speech signal and the masking noise varied sinusoidally across frequency. Assuming that the hypothesis of independent binaural channels is true, a variation in the spectral spacing of the conflicting binaural cues was introduced as a parameter. If the spectral distance between conflicting binaural cues is smaller than the effective binaural integration bandwidth, a decrease in binaural unmasking is expected. This allows for an estimate of a reasonable filter bandwidth to be used within the binaural speech intelligibility model.

In this study, the sinusoidal variation in the IPD was defined on a logarithmic frequency axis in order to be consistent with the roughly constant ratio of auditory bandwidth and center frequency (as expressed, e.g., in the ERB scale). SRTs were measured in the described binaural condition with opposite IPD signs for speech and noise, varying the IPD periodicity. As reference, conditions with the same IPD function but equal IPD signs for speech and noise were measured. In these conditions, no effect of binaural unmasking was expected. The same conditions, but with one channel switched off, were included in order to assess any effects of the monaural phase distortion.

II. METHODS

A. Sentence test procedure

The speech intelligibility measurements were carried out using the HörTech Oldenburg measurement applications (OMA), version 1.2. The Oldenburg sentence test in noise (Wagener *et al.*, 1999a, 1996b, 1996c) was used as speech material. Except for the convolution with the filters that produced the binaural conditions as described in Sec. II B, the signals were the same as in the commercially available version. Each sentence of the Oldenburg sentence test consisted of five words with the syntactic structure "name verb numeral adjective object." For each part of the sentence, ten alternatives were available, each of which occurred exactly twice in a list of 20 sentences, but in random combination. This resulted in syntactically correct, but semantically unpredictable sentences. The lists were completely interchangeable, as the difference between mean SRTs across lists is about 0.2 dB (Wagener *et al.*, 1999b). The subjects' task was to repeat each word they recognized as closely as possible. An instructor marked the correctly repeated words on a touch screen display connected to a computer, which adaptively adjusted the speech level after each sentence to measure the SRT level of 50% intelligibility, according to the "A1" procedure published by Brand and Kollmeier (2002). The step size of each level change depended on the number of correctly repeated words of the previous sentence and on a "convergence factor" that decreased exponentially after each reversal of presentation level. The intelligibility function was represented by the logistic function, which was fitted to the data using a maximum-likelihood method. A test list of 20 sentences was selected from 45 such lists to obtain each observed SRT value. Two sentence lists with 20 sentences each were presented to the subjects prior to each measurement session for training purposes. The test lists were balanced across subjects and conditions, and all measurements except for the training lists were performed in random order.

The noise used in the speech tests was generated by randomly superimposing the speech material of the Oldenburg sentence test (Wagener *et al.*, 1999a). Therefore, the long-term spectrum of this noise was very similar to the mean long-term spectrum of the speech material. The noise token was presented simultaneously with the sentences. It started 500 ms before and stopped 500 ms after each sentence. The starting point of the noise token was randomly selected within the whole noise signal of about 3.7 s which was looped to its beginning if necessary. The noise level was kept fixed at 65 dB SPL.

The headphones (Sennheiser HDA 200) were free-field equalized according to international standard (ISO 389-8), using an finite impulse response filter with 801 coefficients. This free-field equalization is already inherent in the standard signals of the Oldenburg sentence test for the HDA 200. The measurement setup was calibrated to dB SPL using a Brüel & Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 in. microphone, a B&K 2669 preamplifier, and a B&K 2610 measuring amplifier.

B. Stimuli

Both speech and noise signals were presented to the listeners with frequency-dependent IPDs. The IPD $\phi(f)$ as a function of frequency was given by

$$\phi(f) = \phi_0 \sin\left[4\pi\left(B \log \frac{f_h}{f_l}\right)^{-1} \log \frac{f}{f_l}\right]. \quad (1)$$

The value of $|\phi_0|$ was always $\pi/2$, whereas the sign of ϕ_0 was varied according to the condition and the respective signal. The speech and noise signals were bandpass filtered be-

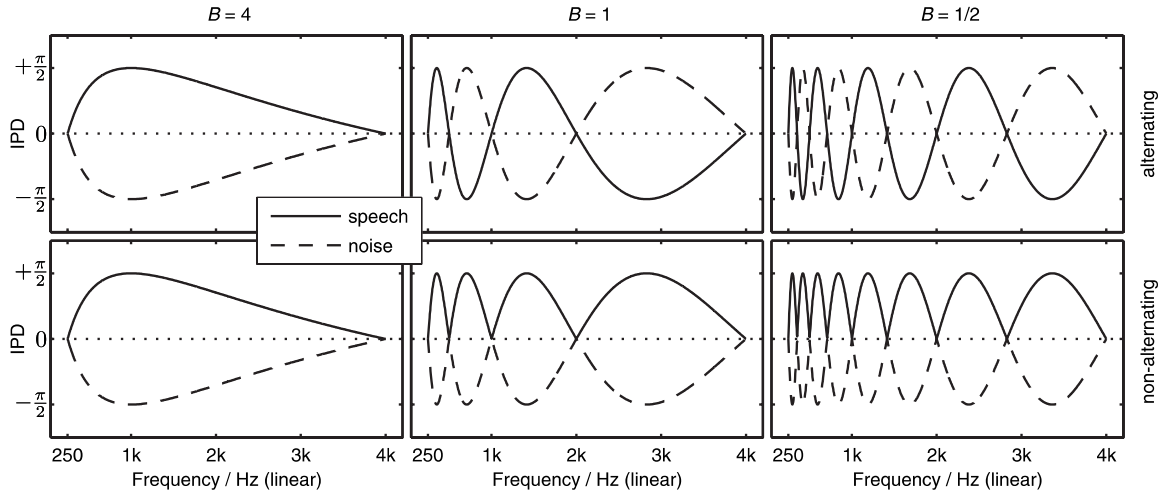


FIG. 1. Schematic display of the IPD function $\phi(f)$ for three different examples of the parameter B which corresponds to the IPD bandwidth in octaves. Solid lines show the IPD of the speech signal; dashed lines the IPD of the noise signal. The upper panels show the IPD functions used in the conditions, in which periods of positive and negative IPD signs alternate. The lower panels show IPD functions of the non-alternating conditions, in which the speech signal has always a positive IPD and the noise signal always a negative IPD. The signals were bandpass filtered between 250 Hz and 4 kHz; IPDs outside this range (considering finite filter slopes) were always zero.

tween f_l=250 Hz and f_h=4000 Hz and $\phi(f)$ was set to zero below f_l and above f_h in order to avoid edge effects because of the finite filter slopes. The parameter B was used to control the bandwidth of the half periods of $\phi(f)$ or, in other words, the distance between zeros of the IPD. B corresponds to the frequency ratio between zeros measured in octaves. The values used for B were 0.125, 0.25, 0.5, 1, 2, and 4. At these values of B, $\phi(f)$ is equal to zero at f_l and f_h. Examples of $\phi(f)$ for different values of B are illustrated in Fig. 1. Note that an important distinction is made in the following between the "IPD bandwidth," which is controlled by the parameter B, and the "filter bandwidth," which denotes the filter bandwidth of the model.

The IPDs were realized by fast convolution of the speech and noise signals with finite impulse response filters. The filters were digitally generated in the frequency domain and had a length of 65 536 samples (≈1.49 s at a sampling rate of 44 100 Hz). The phase shift creating the IPD was divided symmetrically among both ears in order to reduce the monaural phase distortions. Thus, the frequencies at a maximum or minimum of $\phi(f)$ were shifted by $\pm\pi/4$ in the left ear and $\mp\pi/4$ in the right ear with respect to the frequencies at which $\phi(f)$ was zero. The amplitude function of the filter was flat between 250 and 4000 Hz and decreased linearly to zero within a third octave below and above this region. The actual IPD (with respect to phase distortions in the headphones and due to the headphones' placement) was controlled by recording the output of the headphones with an artificial head (B&K HATS 4128C) several times and removing and repositioning the headphones after each recording. A frequency-dependent IPD deviation from the desired value was measured, which is mostly due to an asymmetry of the artificial head in combination with the limited reproducibility of headphone's placement. The maximal absolute IPD deviation in the frequency range used in the stimuli (250 Hz–4 kHz) was $\pi/6$. This deviation is not expected to affect the results substantially, because the exact IPD is not critical in the design of this experiment, as long as the IPD between speech and noise as well as the alternating IPD signs are reproduced correctly. This was checked by measuring the difference between adjacent IPD maxima and minima in the recordings. The deviation from the desired value of π was below $\pi/50$.

SRTs were measured in six conditions for each value of B. The conditions were a combination of three levels of binaural cues for segregation of speech and noise present in the stimuli (monaural, reference, and binaural) and the characteristics of the IPD function (alternating and non-alternating). The IPD functions used for each condition are listed in Table I. In the binaural conditions, the IPDs of speech and noise exhibit differences of up to $\pm\pi$ which may be used as a cue for binaural unmasking. In the non-alternating binaural condition, the difference was always positive, while in the alternating binaural condition, the sign of the IPD between speech and noise changed at each zero of the IPD function. The alternating binaural condition requires independent binaural processing in different frequency bands for maximal binaural unmasking. In the reference conditions, speech and noise had the same IPD at all frequencies and thus no binaural unmasking could be expected, neither for

TABLE I. Conditions and their respective IPD functions used for the speech and noise signals, where $\varphi(f)$ is given in Eq. (1). The monaural conditions are equivalent to the binaural conditions except that the right headphone was switched off and only the monaural phase distortion due to the IPD filter was present in the left ear. In the reference and the binaural conditions, stimuli were presented to both ears.

		Monaural	Reference	Binaural
Alternating	Speech IPD	$+\varphi(f)$	$+\varphi(f)$	$+\varphi(f)$
	Noise IPD	$-\varphi(f)$	$+\varphi(f)$	$-\varphi(f)$
Non-alternating	Speech IPD	$+\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$
	Noise IPD	$-\lvert\varphi(f)\rvert$	$+\lvert\varphi(f)\rvert$	$-\lvert\varphi(f)\rvert$

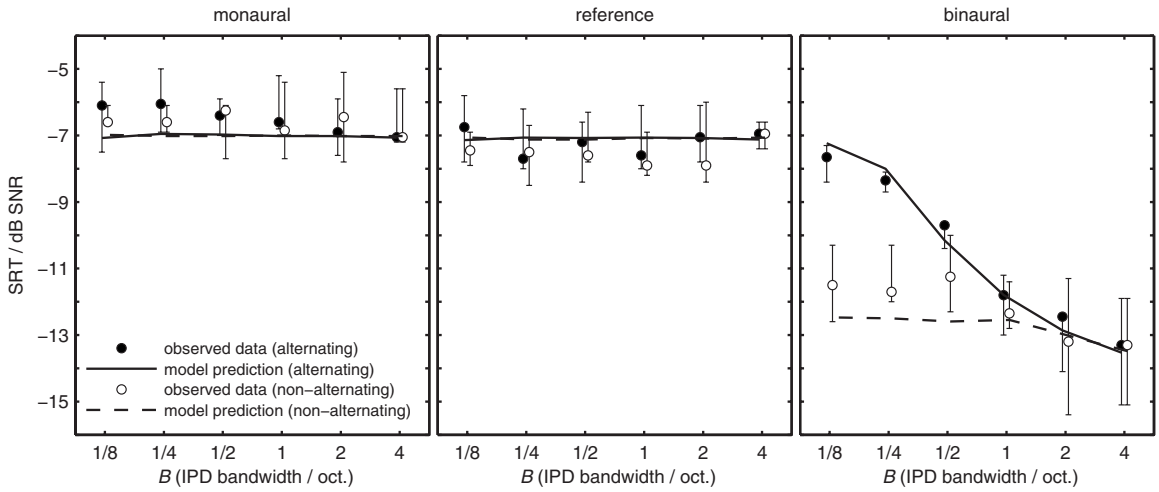


FIG. 2. Observed SRTs of six subjects (circles, median with upper and lower quartiles) and model predictions (lines). Filled symbols and solid lines represent alternating conditions; open symbols and dashed lines represent non-alternating conditions. The leftmost panel shows SRTs for the monaural conditions, the middle panel for the reference conditions with equal IPD for speech and noise, and the rightmost panel for the binaural conditions with opposite IPD signs for speech and noise.

the alternating reference condition nor for the non-alternating reference condition. In the monaural conditions, the stimuli were only presented to the left ear; the right ear channel was switched off. The presented left channel contained the same phase shifts that were necessary to generate the IPDs in the binaural conditions in order to assess the effect of monaural phase distortions on the SRT.

C. Subjects

Six subjects with normal hearing participated in the measurements. Their ages ranged from 24 to 32 years. Their hearing levels did not exceed 15 dB HL (measured at 11 audiometric frequencies between 125 Hz and 8 kHz). All subjects had little or no prior experience in sentence tests. Three of them were members of the research group; the other three subjects were paid for their participation.

D. Model

1. Model structure

A detailed description of the binaural speech intelligibility model used to predict the measurement data of this study can be found in Beutelmann and Brand (2006). Here, only a short overview of the important features is given. The binaural speech intelligibility model processes binaural speech and noise input signals separately. The signals are split into 30-ERB-wide frequency bands (Glasberg and Moore, 1990) between 140 Hz and 9 kHz with a gammatone filter bank (Hohmann, 2002). In each frequency band, an EC (Durlach, 1963) process is used to estimate the best SNR achievable by binaural interaction. The performance of the process is limited by both an additional internal noise that represents the hearing threshold and artificial inaccuracies of the EC process (cf. Durlach, 1963; vom Hövel, 1984) that constrain the maximum SNR benefit due to binaural interaction. The bandwise SNRs are then used as input into the SII (ANSI, 1997), from which a speech intelligibility and finally a SRT are computed. A non-standard SII frequency band scheme was employed in order to match the center frequencies of the SII with the gammatone filter bank. The SII calculation procedure was left unchanged except for the computation of the spread of masking between the frequency bands, which was skipped. This was done because the gammatone filters used in the binaural model are overlapping and already incorporate the spread of masking as it is computed explicitly in the standard SII for non-overlapping bands. The importance function was adapted from the standard importance function for speech in noise by interpolating the bandwidth-weight product, which is practically constant across all standardized SII frequency band schemes.

2. Modifications and tests

In order to examine the influence of the model's filter bandwidth, the model calculations were repeated with different bandwidths of the gammatone filters. The filter bandwidths were varied in steps of 0.1 ERB between the original value of 1 and 4 ERBs, while the center frequencies remained unchanged. Furthermore, the calculations were repeated with the model forced to use a constant time delay (τ=const) or a constant phase delay ($\varphi=\omega_k\tau_k$=const, where ω_k is the center frequency and τ_k the equalization delay of the kth band) across all bands, while calculating the best possible SRTs for each type of delay. This was done in order to simulate the extreme case in which binaural processing is fixed across all frequency bands.

III. RESULTS

A. Measurement data

Figure 2 shows the SRTs of all conditions. The observed SRTs are displayed as medians of six subjects with error bars showing the respective upper and lower quartiles of the data. Filled symbols represent the alternating conditions and open symbols represent the non-alternating conditions. The data at an IPD bandwidth of 4 octaves are the same in alternating and non-alternating conditions because one single IPD half period spans the complete frequency range used in this ex-

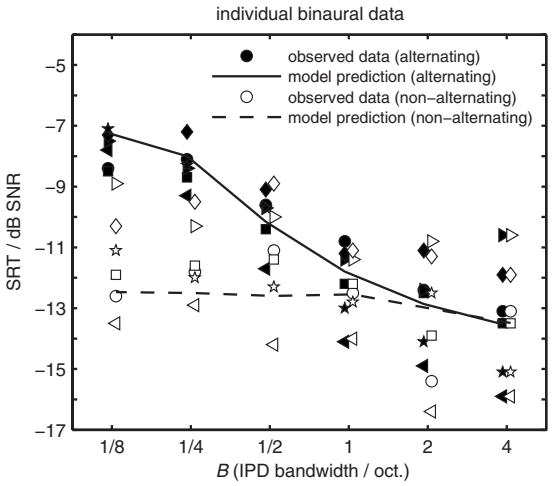


FIG. 3. Individual observed SRTs of six normal-hearing listeners in binaural alternating (filled symbols) and non-alternating (open symbols) conditions and model predictions (lines). Each symbol identifies data of one individual subject.

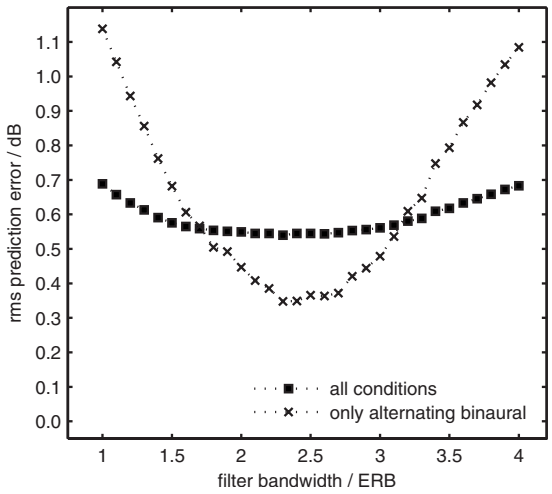


FIG. 4. rms errors between model predictions and observed data (median across subjects) as a function of filter bandwidth for all conditions (square symbols) and for the alternating binaural conditions (cross symbols).

periment, and there is no difference between the alternating and non-alternating conditions at this IPD bandwidth. As expected, no dependence on IPD bandwidth was found in all monaural and reference conditions (leftmost and middle panels in Fig. 2). The SRTs in the alternating binaural condition were strongly dependent on the IPD bandwidth as opposed to the SRTs in the non-alternating binaural condition.

An analysis of variance (ANOVA) of the observed SRTs with the three factors, IPD bandwidth (B), condition, and subject, showed a significant (at the 5% level) main effect of the factor "subject" and no significant effect of the other factors and of any two-way interaction for the conditions in the two left panels in Fig. 2. The effect of the factor subject can be seen in Fig. 3. The spread of individual data was quite large, but was mostly caused by an overall offset between the subjects. An additional difference between subjects was found in the individual maximal amount of binaural unmasking and the separation between the binaural alternating and non-alternating conditions at low IPD bandwidths. *Post-hoc* comparisons with Bonferroni adjustments for multiple comparisons showed the following results (all significances at the 5% level): The amount of binaural unmasking, defined by the difference between corresponding binaural and reference conditions, was significantly larger than zero for all non-alternating conditions and for all alternating conditions except for IPD bandwidths of 0.125 and 0.25 octaves. The difference between alternating and non-alternating conditions was significant only in binaural conditions with IPD bandwidths of 0.125, 0.25, and 0.5 octaves.

Monaural and reference conditions were significantly different at 0.25 octave IPD bandwidth with alternating sign (at the 5% level).

B. Model predictions

In Fig. 2, the predicted SRTs are shown with solid (alternating) and dashed (non-alternating) lines. The filter bandwidth used in the simulations was set to 2.3 ERBs. This filter bandwidth resulted in the lowest overall root mean squared (rms) error between predicted and observed data (median across subjects) of 0.5 dB. This was the minimum of the overall rms prediction error as a function of filter bandwith shown with square symbols in Fig. 4. Model predictions at filter bandwidths of 1, 2, 3, and 4 ERBs are shown in Fig. 5.

The rms errors at filter bandwidths of 1 and 4 ERBs, respectively, were both 0.2 dB higher than at 2.3 ERBs, and the error values increased monotonically from the minimum as a function of filter bandwidth (square symbols in Fig. 4). The range of the error values was small because it was dominated by the errors in the monaural and reference conditions and by the non-alternating binaural conditions (dashed line, rightmost panel in Fig. 2). The difference between predictions and observed data in the non-alternating binaural conditions was larger than for most other conditions, especially at low IPD bandwidths.

Varying the filter bandwidth had practically only an effect on the predictions in the alternating binaural conditions

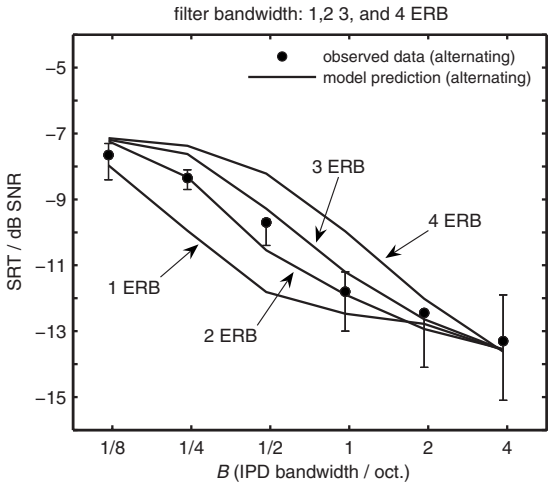


FIG. 5. Model predictions with filter bandwidths of 1, 2, 3, and 4 ERBs (lines, from left to right) and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the alternating binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 2.

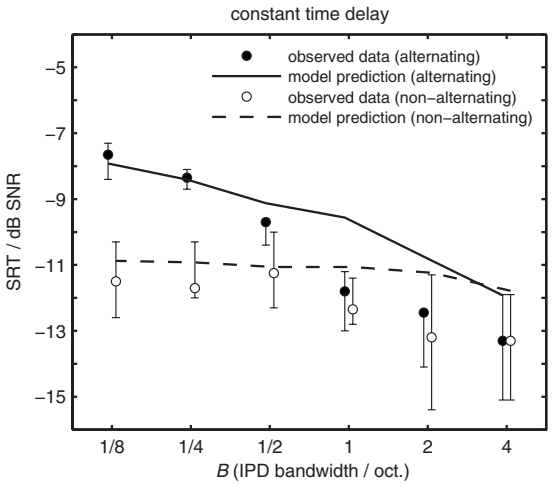


FIG. 6. Model predictions (lines) with constant time delay across all frequency bands and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 2.

(solid line, rightmost panel in Fig. 2) because the maximal difference between corresponding model predictions with different filter bandwidths was below 0.3 dB in all other conditions. In the alternating binaural conditions, however, the maximal difference was up to 3.6 dB (cf. Fig. 5). The effect of filter bandwidth becomes also more apparent if only the rms error across the alternating binaural conditions is shown (cross symbols in Fig. 4).

The predictions with forced constant time or phase delay are shown in Figs. 6 and 7, respectively. Both predictions underestimated the binaural unmasking at high IPD bandwidths in the alternating conditions (solid lines). The predictions using constant phase delay underestimated the binaural unmasking slightly more than the predictions using constant time delay. While the predictions of the non-alternating conditions (dashed lines) with constant phase delay differed only negligibly from the predictions with independent frequency

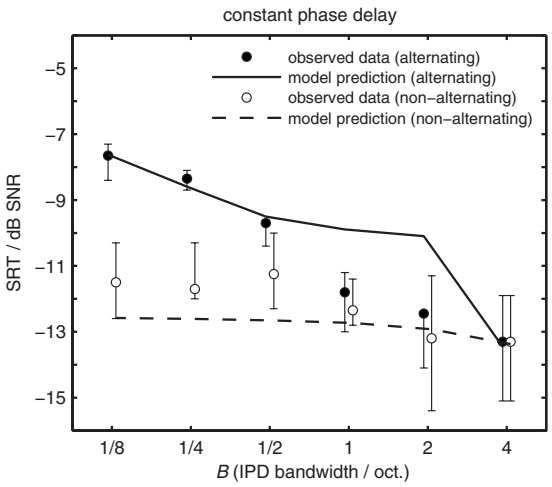


FIG. 7. Model predictions (lines) with constant phase delay ($\varphi = \omega_k \tau_k = \text{const}$, where ω_k is the center frequency and τ_k the equalization delay of the kth band) across all frequency bands and observed SRTs of six subjects (circles, median with upper and lower quartiles) for the binaural conditions. The symbols and line styles and the observed data correspond to the rightmost panel in Fig. 2.

bands, the predicted SRTs with constant time delay were about 1.7 dB higher than all other predicted SRTs in these conditions. This was especially striking at an IPD bandwidth of 4 octaves. None of the other predictions appeared to be dependent on filter bandwidth. The prediction error relative to the median observed data was very low.

IV. DISCUSSION

A. Measurement results

The most striking result of this study is the dependence of the binaural unmasking on the IPD bandwidth. If the IPD bandwidth in the alternating binaural conditions was smaller than 0.5 octaves, no significant binaural unmasking was found, while at IPD bandwidths above this threshold, binaural unmasking occurred up to 6 dB. When the IPD was consistent across frequency bands, as in the non-alternating binaural conditions, the binaural unmasking was not strongly dependent on the IPD bandwidth. The results in the alternating and non-alternating binaural conditions were not significantly different for IPD bandwidths above 0.5 octaves. This suggests that IPDs can be utilized independent of different frequency bands provided the distance between frequency bands having incongruent IPDs is sufficiently large.

The results in the monaural control conditions, which are not significantly dependent on the IPD bandwidth, show that the effect is truly binaural and not mainly caused by monaural phase distortions. There are, nevertheless, a few conditions showing significantly different SRTs (of about 1.5 dB) between monaural and reference conditions, which was not expected. This may be due to the relatively small number of subjects. Some trends that can already be found may become significant with a larger number of subjects. One of those trends is the overall slightly higher SRT in the monaural conditions, which is not reflected in the model predictions. Even in situations without binaural differences between target speech and interfering noise there may be a SRT improvement of about 1 dB from monaural to binaural presentation of the signals (Bronkhorst and Plomp, 1988). The observed difference between monaural and reference conditions may be explained by such a SRT improvement, even though the signals in the monaural conditions and the respective ear of the reference conditions are not the same (the monaural conditions correspond to the binaural conditions with opposite IPD of speech and noise, but the right channel was switched off).

B. Model predictions

The general trend of the observed data—the break-down of the binaural unmasking at small IPD bandwidths for the alternating binaural conditions and the roughly constant binaural unmasking for the corresponding non-alternating binaural conditions—is qualitatively predicted well at all tested filter bandwidths by the binaural speech intelligibility model. However, in order to achieve the best prediction of the exact relation between IPD bandwidth and binaural unmasking, a filter bandwidth of 2.3 ERBs had to be used instead of the filter bandwidth of 1 ERB used in the original implementation of the model. Given the relatively large spread of indi-

vidual observed SRTs, the value of 2.3 ERBs may need to be adjusted, if data from more subjects are added, but it can be expected to stay within the range of more than 1 ERB and less than 4 ERBs. The prediction error (Fig. 4) differs only slightly from the minimum within a range of about 0.5 ERB. The filter bandwidths calculated for each individual subject range from 1.2 ERBs to 3.4 ERBs.

By far the largest prediction error occurs in the non-alternating binaural conditions. This may be attributed to the fact that the steep slopes in the vicinity of the zeros of the ideal non-alternating IPD function cannot be reproduced exactly by the measurement equipment and thus provide less useful binaural information to the listener than to the model. In the monaural and reference conditions, this is not relevant and therefore the model predictions are more accurate.

Forcing the model to use only one time delay which is constant across all frequency bands can be regarded as a case with extremely wide filters. Thus it is not surprising that the predictions with constant time delay (Fig. 6) are similar to the predictions with 4 ERB filter bandwidth (Fig. 5, rightmost line) and underestimate the binaural unmasking even at higher IPD bandwidths. The predictions of the model for an IPD bandwidth of 4 octaves are not as accurate as those achieved using the independent-band model. This indicates that a constant time delay across all frequency bands is not sufficient for the correct prediction of even the condition with the least variation in IPD. The predictions with constant phase delay, that is, with a constant $\varphi=\omega_k\tau_k$ in each frequency band with the center frequencies ω_k, are as good as the independent-band model for the IPD bandwidth of 4 octaves and for all non-alternating binaural conditions, but they underestimate the binaural unmasking in the alternating binaural conditions for high IPD bandwidths. This is especially remarkable at an IPD bandwidth of 2 octaves because in this case, the first zero of the IPD function is at 1 kHz, and it is usually expected that the contribution of IPDs to binaural unmasking is by far more important in the frequency range below 1 kHz than above. Thus, the optimal strategy would be to choose the phase delay for equalization that yields good binaural unmasking in the low-frequency range. The error made by this strategy in the high-frequency range should be negligible if the contribution of binaural unmasking due to IPDs between speech and noise in the frequency range above 1 kHz was small compared to the contribution at frequencies below 1 kHz. The fact that the predictions with constant phase delay and the observed data at an IPD bandwidth of 2 octaves differ significantly shows that the contribution of high frequencies has to be taken into account.

In this study, no long-term ILDs were present in the stimuli. The relation between model filter bandwidths and IPD bandwidths is therefore mainly based on the processing of IPDs and may be different for similar experiments, which employ frequency-dependent ILDs or combined IPDs and ILDs. While there is evidence for less independent processing of ILDs in adjacent frequency bands (Edmonds and Culling, 2006), the combination of ILDs and IPDs should be examined in further studies and may be crucial for the development of broad-band binaural models like the binaural speech intelligibility model presented here. Related to this is the question of how the larger binaural filter bandwidths should be combined with the smaller monaural bandwidths in those models. The filter bandwidth has apparently only an influence on the prediction of conditions with very extreme spectral changes in the IPD, but not on the predictions of the monaural and reference conditions. Nevertheless, it is worthwhile examining more closely if there is a need for multiple bandwidths in binaural speech intelligibility models for monaural and binaural conditions, particularly with regard to potentially different auditory bandwidths of hearing-impaired subjects. Nitschmann and Verhey (2007) approached this issue, for example, by using the monaural filter bandwidth for signal analysis, but combining the information of the target-centered band with neighboring bands for binaural processing. For a model using linear signal processing, such as the binaural speech intelligibility model in its present form, summing up the output of adjacent auditory filters is mathematically equivalent to using a wider filter.

The auditory bandwidth factor of 2.3 for binaural processing relative to monaural processing estimated in this study is generally within the range of other results from the literature. it matches very well the factor of about 2.5 found by Hall *et al.* (1983) for binaural tone detection in band-limited noise and with a spectral level of 30 dB/Hz, which is close to the average noise spectral level used in this study. Sondhi and Guttman (1966) found a factor of about 2 for the frequency band centered on 500 Hz, with a paradigm of noise bands with binaural cues closely embedded in noise bands without binaural cues, which is similar to the paradigm used in this study. In the study of Holube *et al.* (1998), the similar periodic variation in binaural cues on a linear frequency scale resulted in binaural bandwidth factors of about 1.6, which is smaller than the value from this study, but would still lead to tolerable predictions with the binaural speech intelligibility model.

Exact comparisons of the binaural filter bandwidth would need to consider not only the bandwidth but also the filter shape, as described in Kollmeier and Holube (1992). As a compromise, the −10 dB-bandwidth or even better, the bandwidth which encompasses 90% of the integrated filter function, was suggested instead of the −3 dB-bandwidth. This is reflected in the comparison of this study and the model of Nitschmann and Verhey (2007). The latter is aimed at predicting differences between effective binaural bandwidths calculated from bandlimiting and notched-noise experiments as performed by Hall *et al.* (1983). The −3 dB-bandwidth is nearly the same in this study and in Nitschmann and Verhey (2007), while the −10 dB-bandwidth of the 2.3-ERB-wide fourth order gammatone filters used in this study is about 30% larger than the weighted combination of three adjacent 1-ERB-wide third order gammatone filters used by Nitschmann and Verhey (2007).

Another question concerns the difference between ITD and IPD. Would the results of this study be similar if the frequency-dependent IPDs were replaced by frequency-dependent ITDs? This question applies to the currently ongoing discussion about the way how binaural timing disparities are represented in the brain. The assumption of the very popular and successful model by Jeffress (1948) was that

ITD is coded by the activation of neurons, which are tuned to a certain best ITD due to the difference of axonal propagation time between the left and the right ear. ITDs are displayed by coincident arrival of spikes at certain neurons, each of which represents a certain ITD. Although there is anatomical evidence for this kind of structure in birds (Carr and Konishi, 1990), recent studies (McAlpine *et al.*, 2001; McAlpine and Grothe, 2003) have cast doubt on this "delay line" hypothesis in mammals. Appendix supplies some detailed arguments for why the results of this study may have been different if ITDs instead of IPDs were perceived by the subjects.

The consequences for binaural modeling that can be drawn are that (1) the binaural processing of broad-band target and interferer signals with frequency-dependent IPDs is subject to a larger auditory integration bandwidth than typically used in monaural detection models and (2) the hypothesis of independent processing in different auditory frequency bands is supported by the results of this study.

V. CONCLUSIONS

A sinusoidally frequency-dependent IPD between speech and noise resulted in binaural unmasking of up to 6 dB relative to the SRT for equal IPDs of speech and noise. If the IPD had alternating signs in adjacent frequency bands, the binaural unmasking strongly depended on the bandwidth of the IPD periods and was only significantly larger than zero for IPD bandwidths larger than a third octave. If the sign of the IPD was non-alternating, that is, consistent across all frequencies, the binaural unmasking showed only little variation and was significantly larger than zero even for IPD bandwidths below 0.5 octaves.

The binaural speech intelligibility model by Beutelmann and Brand (2006) predicted the binaural unmasking due to the IPDs between speech and noise very well. The predictions correctly exhibit the decrease in unmasking with decreasing IPD bandwidth for alternating sign of the IPD and the stable unmasking for non-alternating sign of the IPD. The lowest prediction error was achieved by assuming a binaural filter bandwidth of 2.3 ERBs in the model.

The assumption of constant equalization parameters across all frequency bands is not sufficient for good predictions. Provided that the filter bandwidth is within the limits mentioned above, it appears reasonable that binaural processing in each frequency band is virtually independent of the adjacent bands (i.e., the equalization parameters can be chosen independently). Thus, the "independent binaural processing channel" hypothesis is supported.

ACKNOWLEDGMENTS

We would like to thank the editor, Richard Freyman, and three anonymous reviewers for their thorough and helpful reviews. This study was supported by the Deutsche Forschungsgemeinschaft within the SFB TRR 31 "The active auditory system."

APPENDIX: IPD AND ITD OF BROAD-BAND SIGNALS

In tone detection experiments, IPD and ITD of the target tone are virtually indistinguishable, but for the interferer (apart from sine tones used as interferers), a constant IPD leads to a frequency-dependent ITD and vice versa. If the frequency band that needs to be considered is sufficiently small, Breebaart *et al.*(1998) showed that the difference between the effect of constant ITD and IPD, respectively, on binaural unmasking is rather small. For broad-band target signals as in binaural speech intelligibility experiments, however, it is certainly necessary to distinguish between IPD and ITD. Whereas the IPD is unambiguously defined as a function of frequency, the ITD as a function of frequency can be either defined as a *phase* delay, $\varphi(\omega)/\omega$, where $\varphi(\omega)$ is the IPD as a function of angular frequency ω, or as a *group* delay, $d\varphi(\omega)/d\omega$. The values of the ITD according to these two definitions are only equal if the ITD is constant across all considered frequencies. Phase delay generally acts on the fine structure of a signal, while group delay affects its envelope. Using a windowed sinusoid as the signal, for example, a constant phase delay shifts the zeros of the sinusoid without changing the window position, while a constant group delay shifts the maximum of the window. The phase delay ITDs calculated from the IPD functions used in this study do not change the functional form and the sign of the IPD functions; they are only multiplied by a factor of $1/\omega$. The group delay ITDs in the alternating conditions have different zero-crossing frequencies than the IPD functions [due to the derivative of the sine-function in Eq. (1)], but the general periodic form of the function is similar between IPD and group delay ITD function. Most interesting is the group delay ITD in the non-alternating conditions because from the mathematical point of view, the group delay ITD in these conditions is still alternating between positive and negative signs across frequencies. Binaural processing exclusively based on group delay would not be expected to result in the different dependence of binaural SRTs on IPD bandwidth in the alternating and the non-alternating conditions observed in this study because the distinction between alternating and non-alternating signs is not given in the group delay ITDs calculated from the IPD function in Eq. (1).

[1]The equivalent rectangular bandwidth (ERB) in hertz of an auditory filter is defined by $\mathrm{ERB}=24.7(4.37f+1)$, where f is the center frequency of the filter in kilohertz. It was derived from monaural data (Glasberg and Moore, 1990).

Akeroyd, M. A. (**2004**). "The across frequency independence of equalization of interaural time delay in the equalization-cancellation model of binaural unmasking," J. Acoust. Soc. Am. **116**, 1135–1148.

ANSI (**1997**). "Methods for the calculation of the speech intelligibility index," American National Standard S3.5 (Acoustical Society of America, Melville, NY).

Best, V., Gallun, F. J., Carlile, S., and Shinn-Cunningham, B. G. (**2007**). "Binaural interference and auditory grouping," J. Acoust. Soc. Am. **121**, 1070–1076.

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Breebaart, J., van de Par, S., and Kohlrausch, A. (**1998**). "Binaural signal detection with phase-shifted and time-delayed noise maskers," J. Acoust. Soc. Am. **103**, 2079–2083.
Breebaart, J., van de Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.
Bronkhorst, A. W., and Plomp, R. (**1988**). "The effect of head-induced interaural time and level differences on speech intelligibility in noise," J. Acoust. Soc. Am. **83**, 1508–1516.
Buell, T. N., and Hafter, E. R. (**1991**). "Combination of binaural information across frequency bands," J. Acoust. Soc. Am. **90**, 1894–1900.
Carr, C. E., and Konishi, M. A. (**1990**). "A circuit for detection of interaural time differences in the brainstem of the barn owl," J. Neurosci. **10**, 3227–3246.
Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.
Cokely, J. A., and Hall, J. W. (**1991**). "Frequency resolution for diotic and dichotic listening conditions compared using the bandlimiting measure and a modified bandlimiting measure," J. Acoust. Soc. Am. **89**, 1331–1339.
Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (**2004**). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," J. Acoust. Soc. Am. **116**, 1057–1065.
Culling, J. F., and Summerfield, Q. (**1995**). "Perceptual separation of concurrent speech sounds—Absence of across-frequency grouping by common interaural delay," J. Acoust. Soc. Am. **98**, 785–797.
Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.
Edmonds, B. A., and Culling, J. F. (**2005**). "The spatial unmasking of speech: Evidence for within- channel processing of interaural time delay," J. Acoust. Soc. Am. **117**, 3069–3078.
Edmonds, B. A., and Culling, J. F. (**2006**). "The spatial unmasking of speech: Evidence for better-ear listening," J. Acoust. Soc. Am. **120**, 1539–1545.
Fletcher, H. (**1940**). "Auditory Patterns," Rev. Mod. Phys. **12**, 47–65.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.
Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1983**). "Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking," J. Acoust. Soc. Am. **73**, 894–898.
Hohmann, V. (**2002**). "Frequency analysis and synthesis using a gammatone filterbank," Acust. Acta Acust. **88**, 433–442.
Holube, I., Kinkel, M., and Kollmeier, B. (**1998**). "Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments," J. Acoust. Soc. Am. **104**, 2412–2425.
Houtgast, T. (**1977**). "Auditory-filter characteristics derived from direct-masking data and pulsation-threshold data with a rippled-noise masker," J. Acoust. Soc. Am. **62**, 409–415.
Jeffress, L. (**1948**). "A place theory of sound localization," J. Comp. Physiol. Psychol. **41**, 35–39.
Kohlrausch, A. (**1988**). "Auditory filter shape derived from binaural masking experiments," J. Acoust. Soc. Am. **84**, 573–583.
Kollmeier, B., and Holube, I. (**1992**). "Auditory filter bandwidths in binaural and monaural listening conditions," J. Acoust. Soc. Am. **92**, 1889–1901.
McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—Do mammals fit the model?," Trends Neurosci. **26**, 347–350.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
Metz, P. J., von Bismarck, G., and Durlach, N. I. (**1968**). "Further results on binaural unmasking and the EC model. II. Noise bandwidth and interaural phase," J. Acoust. Soc. Am. **43**, 1085–1091.
Nitschmann, M., and Verhey, J. L. (**2007**). "Experimente und Modellrechnungen zur binauralen spektralen Selektivität (Experiments and model calculations on binaural spectral selectivity)," in *Fortschritte der Akustik, DAGA 2007* (Deutsche Gesellschaft für Akustik e.V., Berlin), pp. 371–372.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Sever, J. C., and Small, A. M. (**1979**). "Binaural critical masking bands," J. Acoust. Soc. Am. **66**, 1343–1350.
Sondhi, M. M., and Guttman, N. (**1966**). "Width of the spectrum effective in the binarual release of masking," J. Acoust. Soc. Am. **40**, 600–606.
van der Heijden, M., and Trahiotis, C. (**1998**). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," J. Acoust. Soc. Am. **103**, 1609–1614.
vom Hövel, H. (**1984**). "Zur Bedeutung der Übertragungseigenschaften des Außenohrs sowie des binauralen Hörsystems bei gestörter Sprachübertragung (On the importance of the transmission properties of the outer ear and the binaural auditory system in disturbed speech transmission)," Ph.D. thesis, RTWH Aachen, Aachen, Germany.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999a**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a sentence test for the German language I: Design of the Oldenburg sentence test)," Z. Fuer Audiologie, Audiological Acoust. **38**, 4–15.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999b**). "Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a sentence test for the German language II: Optimization of the Oldenburg sentence test)," Z. Fuer Audiologie, Audiological Acoust. **38**, 44–56.
Wagener, K., Brand, T., Kühnel, V., and Kollmeier, B. (**1999c**). "Entwicklung und Evaluation eines Satztests für die Deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a sentence test for the German language III: Evaluation of the Oldenburg sentence test)," Z. Fuer Audiologie, Audiological Acoust. **38**, 86–95.
Wightman, F. L. (**1971**). "Detection of binaural tones as a function of masker bandwidth," J. Acoust. Soc. Am. **50**, 623–636.
Zurek, P. M. (**1990**). "Binaural advantages and directional effects in speech intelligibility," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hockberg, 2nd ed. (Allyn and Bacon, Boston), Chap. 15, pp. 255–276.

Acoustic and spectral patterns in young children's stop consonant productions[a)]

Shawn L. Nissen
Department of Communication Disorders (138 TLRB), Brigham Young University, Provo, Utah 84602

Robert Allen Fox
Department of Speech and Hearing Science, The Ohio State University, Columbus, Ohio 43210

(Received 15 April 2009; revised 6 July 2009; accepted 7 July 2009)

The aim of this study was to examine the acoustic and spectral patterns of stop articulation in the speech of pre-pubescent children. A set of voiceless stop consonants, /ptk/, produced by a group of adults and typically developing children 3–5 years of age were examined in terms of multiple acoustic and spectral parameters. Findings indicated that, with the exception of spectral kurtosis, the acoustic and spectral characteristics of the stop productions varied significantly as a function of place of articulation and vowel context. Sex-specific differences in spectral slope, mean, and skewness were found for the 5-year-old and adult speakers. Such differences in adult speakers can be explained in part by variation in vocal tract size across the sex of the speaker; however, vocal tract dimorphism is typically not present in pre-pubescent children. Thus, the findings of this study provide some support that sex-specific differences in the speech patterns of young children may be associated with learned or behavioral factors, such as patterns of obstruent articulation that depend in part on a culturally determined male-female archetype.
 [DOI: 10.1121/1.3192350]

PACS number(s): 43.70.Ep, 43.70.Bk, 43.70.Fq, 43.70.Aj [BHS] Pages: 1369–1378

I. INTRODUCTION

Researchers continue to search for more accurate models and descriptions of the complex process by which humans develop the ability to produce and perceive speech in a remarkably efficient and accurate manner. Advances in instrumentation and computer technology have allowed scientists to utilize more specific types of acoustic analysis to investigate speech communication. Spectral moments analysis (e.g., Forrest *et al.*, 1988; Jongman *et al.*, 2000; Nittrouer, 1995) is an analysis method which allows the extraction of a number of spectral characteristics from discrete time segments of the speech signal (including spectral mean, variance, skewness, and kurtosis), providing researchers with the means to identify unique patterns of acoustic energy.

Utilizing spectral moments analysis, as well as more traditional types of acoustic analysis, researchers have investigated the differences in acoustic structure of obstruents produced by phonologically disordered children (e.g., Forrest *et al.*, 1994; Miccio, 1996). An associated line of research has focused on changes in obstruent production as a function of typical speech development and maturation (e.g., Forrest *et al.*, 1990; Fox and Nissen, 2005; Miccio, 1996; Nissen and Fox, 2005; Nittrouer, 1995; Nittrouer *et al.*, 1989).

Following the methodology employed with adult speakers by Forrest *et al.* (1988), a study by Nittrouer (1995) utilized spectral moments analysis to investigate the nature of obstruent acquisition in children between 3 and 7 years of age and a comparison group of adult speakers. Nittrouer (1995) examined the acoustic nature of voiceless stop (/t/ and /k/) and fricative (/s/ and /ʃ/) productions in terms of spectral mean, skewness, and kurtosis. It was found that the children exhibited significantly fewer distinctions and less contrast in spectral mean between the /s/ and /ʃ/ fricatives than the adult speakers. The author suggested that this age-related difference may indicate that the children were continuing to "fine-tune" their fricative articulations with developmental maturation. Interestingly, it was found that the stop consonant productions differed very little as a function of age in terms of spectral mean, indicating that developmental differences in speech production may vary as a function of both the age of the speaker and the manner of the obstruent contrast.

The second spectral moment of variance has typically not been utilized in earlier spectral studies of obstruent production in children (e.g., Forrest *et al.*, 1990, 1994; Nittrouer, 1995). The exclusion of the spectral variance measure from these studies may be due to the fact that Forrest *et al.* (1988) did not find a significant contribution of spectral variance to the discrimination of a set of voiceless obstruents produced by a group of adult speakers. However, several recent studies have found that the spectral variance, as well as the spectral slope, can be useful in discriminating between fricative contrasts (Jongman *et al.*, 2000; Fox and Nissen, 2005; Nissen and Fox, 2005).

Previous research has not only found that speech development varies as a function of age but also according to the sex of the speaker. For example, research investigating voice onset time (VOT) has found sex differences in the production of stop consonants. In general (see Sweeting and Baken,

[a)] Portions of this work are contained in the unpublished doctoral dissertation, "An acoustic study of voiceless obstruents produced by adults and typically developing children," The Ohio State University, 2003.

1982), adult female speakers of American English have been reported to exhibit significantly longer VOT values than adult males (e.g., Ryalls *et al.*, 1997; Swartz, 1992; Whiteside and Irving, 1997). Sex-related acoustic differences have also been noted in speaking rate (Byrd, 1992, 1994; Klatt and Klatt, 1990) and the occurrence of pauses (Whiteside, 1996). It has been shown that for some consonants, adult female speakers exhibit a greater degree of devoicing when compared to male speakers (Fant *et al.*, 1991; Whiteside, 1996). Moreover, research has revealed significant differences between male and female speakers in prosodic patterns (Fitzsimmons *et al.*, 2001; Graddol, 1986) and voice quality (Klatt and Klatt, 1990; Mendoza *et al.*, 1996).

Sex-related acoustic differences have also been noted in the speech of younger speakers. Several studies involving preadolescent children (Bennett, 1981; Busby and Plant, 1995; Perry *et al.*, 2001; Whiteside and Hodgsen, 2000) have found that formant frequency patterns differed significantly as a function of speaker sex. A significant contribution to this area of research was a large scale study ($N=436$) by Lee *et al.* (1999) and a subsequent reanalysis of the data by Whiteside (2001). Although some researchers have attributed dissimilarities in formant patterns across speaker sex to nonuniform anatomical changes in the vocal tract size (e.g., Bennett, 1981; Fant, 1966), Lee *et al.* (1999) concluded that the acoustic variation in formant patterns across speaker sex could not be solely attributed to differences in vocal tract morphology.

Less emphasis has been placed on examining sex-related developmental differences in the productions of voiceless obstruents. A study examining voiceless fricative productions from children as young as 6 years of age found the presence of sex-related differences in multiple individual acoustic and spectral parameters, with findings from a discriminant analysis also indicating differences in combinations of speech parameters (Fox and Nissen, 2005). In addition, an examination of fricative productions from younger children, 3–5 years of age, also found sex-specific differences in spectral mean and slope (Nissen and Fox, 2005). However, it is unclear if such variation will be found in other types of obstruent productions (i.e., stop consonants) and at what age possible differences might occur. If sex-specific differences are learned phenomena, such differences may occur across phonetic classes of sounds.

Previous research has greatly enhanced the understanding of speech communication; however, the complex relationships between the acoustic structures of speech, as well as the manner and time in which they are acquired by children, have yet to be fully and adequately explained. Research examining stop consonant production in pre-pubescent children as a function of age and sex of the speaker, including several measures (i.e., spectral slope and variance) frequently not included in spectral studies involving children may lead to additional insights into the developmental nature of speech production. In particular, it is of interest to investigate if sex-specific differences noted with fricative production will also occur with stop consonant productions.

Considering recent large-scale magnetic resonance imaging (MRI) studies investigating sexual dimorphism in the oral and pharyngeal portions of the vocal tract (Fitch and Giedd, 1999; Vorperian *et al.*, 2005, 2009), it is of interest to examine sex-specific differences in the speech acoustics of young children to developmental growth patterns of anatomic structures in the vocal tract. Such comparisons are important to establish anatomic-acoustic correlates (Vorperian *et al.*, 2009), as well as provide further insight into the source (anatomic or learned) of male-female speech differences. Thus, this study aims to describe the acoustic patterns (normalized amplitude, slope, mean, variance, skewness, and kurtosis) of the voiceless stop consonants, /p t k/, produced by a group of adults and typically developing children 3–5 years of age, as well as investigate to what extent the individual amplitude and spectral characteristics of the target productions change as a function of age, sex of the speaker, place of articulation, and vowel context.

II. METHODOLOGY

A. Participants

Participants included three groups of children between the ages of 3 and 5 years of age ($N=30$) and one comparison group of adults ($N=10$). Speakers in the 3-year-old group were between 3:0 and 3:11 years of age ($M=3:6$), the 4-year-old group were between 4:0 and 4:11 years of age ($M=4:8$), and the 5-year-old group contained children between 5:0 and 5:11 years of age ($M=5:7$). The adult subjects within the comparison group were between 18 and 40 years of age. Each group was composed of an equal number of male and female subjects. All participants were monolingual speakers of American English, with no diagnosed history of speech, language, or hearing problems. At the time of their participation all of the speakers exhibited pure-tone air-conduction thresholds $\leq$15 dB hearing loss at octave frequencies from 125 to 8000 Hz and had visible front incisors. Prior to recording, the phonemic inventory of each child was evaluated by a certified speech language pathologist using the "Sounds-in-Words" subtest of the Goldman–Fristoe Test of Articulation (GFTA) (Goldman and Fristoe, 1986). All children who participated in the study exhibited target appropriate stop productions, as measured by the GFTA.

B. Stimuli

Target phonemes were elicited from a series of words with an initial syllable containing a combination of one of three voiceless obstruents (/p/, /t/, or /k/) in initial position followed by a monophthongal vowel (/i/, /ɑ/, or /u/). Specifically, the corpus included the following words: *peanut, pocket, Poohbear, teapot, Thomas, toothbrush, key, car*, and *cougar*. Participants produced each word three times while embedded in the carrier phrase "This is a__again." The targeted syllable combinations were in the initial and stressed position of each word to elicit relatively similar vocal emphasis across productions. Occasionally the child participants produced the target word or carrier phrase in a dysfluent manner or incorrectly identified the picture as a different lexical item; in which case the utterance was rerecorded. As expected, the older children displayed fewer instances of dysfluent speech, as well as fewer misidentifications.

C. Elicitation procedures

The speech productions were recorded online to computer in a quiet room environment with a low impedance dynamic microphone (Shure SM10A-CN) and preamplifier (Samson Mixpad-4). The microphone was affixed to a headset and placed approximately 4 cm from the speaker's lips during recording. The productions were digitized at a sampling rate of 44.1 kHz and a quantization of 16 bits, and subsequently low-pass filtered at 22.05 kHz. Target productions were elicited from the participants through the verbal identification of age-appropriate pictures representing the target words. Custom software programmed in MATLAB was utilized to randomly present the elicitation pictures on a 15 in. computer screen and subsequently capture the participants' responses. The participants were familiarized with the names of the pictures and the elicitation procedure prior to the recording session by the test administrator modeling the procedure for the subjects prior to data collection. If a participant incorrectly identified a picture as a different lexical item during the recording session, the correct target word was modeled by the experimenter and the child was instructed to repeat the identification of that particular item.

D. Acoustic and spectral analysis

Segmentation of the onset and offset of the obstruent target segments was conducted through waveform display assisted by spectrographic inspection using ADOBE AUDITION Version 1.3 (Adobe Systems Incorporated, 2003). The onset of the stop burst was characterized by a sharp increase in diffuse noise energy and the rapid increase in zero crossings, with the burst offset defined by a sharp decrease in diffuse noise energy. Segmentation values were then recorded into a text file (in milliseconds) and later checked, corrected, and re-checked using a MATLAB program that displayed the segmentation marks superimposed over a display of the token's waveform. In addition, to test for segmentation accuracy and reliability, 540 tokens (three subjects randomly chosen from each age group) were independently analyzed by a second person and subsequently correlated ($r=0.99$, $p<0.0001$) to the original segmentation of these same tokens, differing by an average of approximately 1 ms.

A measure of normalized amplitude was computed for each stop burst, calculated by subtracting the root-mean-square (rms) amplitude in decibels of the segment of the stop burst from the rms amplitude of the strongest component within the initial 40 ms of the following vowel (Behrens and Blumstein, 1988a, 1988b; Jongman *et al.*, 2000). This amplitude measure served to normalize differences in speaker intensity.

Spectral moments measures (mean, variance, skewness, and kurtosis) were computed for the stop consonants following the approach of Forrest *et al.* (1988) and Nittrouer (1995). Normalized power spectra were derived from a 20 ms Hamming window centered +10 ms from the release of the stop burst, which was then pre-emphasized by first-differencing. Though the need for pre-emphasis is minimal when analyzing voiceless sounds, it was determined that such a procedure was necessary to more effectively compare subsequent results to previously published findings (e.g., Forrest *et al.*, 1988; Jongman *et al.*, 2000; Nittrouer, 1992, 1995). Using a 1024-point fast Fourier transform with zero-padding, the spectral amplitudes of a series of frequency points were derived from the complex acoustic signal within the 20 ms window. The resulting power spectra were considered random distribution probabilities, from which the first (mean), second (variance), third (skewness), and fourth (kurtosis) spectral moments were computed for each of the target stimuli. The third and fourth spectral values were subsequently normalized by the spectral variance for that same token. Measures of spectral slope were derived from the power spectra generated during the spectral moments analysis and calculated from a linear regression line fit to the extracted relative amplitudes of acoustic energy from 1 to 10 kHz. The slope values were reported as a ratio of amplitude to frequency (i.e., dB/kHz). These normalization procedures, as well as the other algorithms utilized in this study, are specifically described in previous spectral moments studies (e.g., Forrest *et al.*, 1988; Fox and Nissen, 2005; Nissen and Fox, 2005; Nittrouer, 1995).

The stimuli in this study were elicited, recorded, and analyzed using custom designed computer programs (MATLAB) created by the authors. A corpus of test tokens comprised of known acoustic components was utilized to evaluate the accuracy and reliability of the spectral analysis. For example, a test token composed of several sinusoidal frequencies (1, 3, and 5 kHz) of equal strength was analyzed by the computer programs and found to have the appropriate values for the various acoustic measures.

E. Statistical analysis

Data were collapsed across repetitions of a given stimulus item and the spectral mean values were transformed to a perceptually normalized scale prior to statistical analysis. Specifically, the equivalent rectangular bandwidth (ERB-2) scale (Glasberg and Moore, 1990; Moore, 1997) was used to normalize the spectral mean measurements, thereby increasing the validity of comparisons across individual speakers. Repeated measures analyses of variance (ANOVAs) were used to examine possible acoustic differences in the stop productions as a function of place of articulation, vowel context, speaker sex, and age group. Results of significant F-tests include a measure of effect size (partial eta squared or η^2), which can be considered a measure of the proportion of variance explained by a dependent variable when controlling for other factors. Greenhouse–Geisser adjustments were utilized to adjust F-tests with regard to degrees of freedom when significant deviations from sphericity were found.

III. RESULTS

Detailed listings of the acoustic and spectral measures for the male and female speakers are found in Tables I and II, respectively. The data are tabularized according to speaker age, stop place of articulation, and vowel context.

TABLE I. Acoustic measures from male speakers for three classes of voiceless American English stop consonants, grouped as a function of speaker age, place of articulation, and vowel context. Normalized amplitude in dB (NAmp), spectral slope in dB/kHz (slope), and the first four spectral moments (mean in ERB, M1, variance in kHz^2, M2; skewness, M3; and kurtosis, M4).

		Bilabial-/p/			Alveolar-/t/			Velar-/k/		
		/i/	/ɑ/	/u/	/i/	/ɑ/	/u/	/i/	/ɑ/	/u/
3 yr. old										
	NAmp	−3.76	−11.96	−7.06	−2.70	−5.75	−3.93	−2.08	−7.38	−3.64
	Slope	5.11	−5.11	−5.57	30.18	22.99	15.56	14.40	−20.90	−20.90
	M1	28.36	28.06	27.29	30.95	30.97	30.07	29.56	25.85	26.46
	M2	6.54	6.65	6.31	5.96	5.53	4.82	4.69	5.69	5.88
	M3	−0.07	0.10	0.11	−1.20	−0.94	−0.25	0.03	0.72	0.96
	M4	−0.03	0.29	1.00	0.92	0.51	0.51	0.05	2.06	1.92
4 yr. old										
	NAmp	−7.26	−17.00	−10.35	−3.28	−5.19	−2.78	−2.83	−8.00	−3.89
	Slope	2.32	−1.16	15.56	28.10	20.20	5.57	−5.34	−19.97	−26.93
	M1	28.99	28.61	30.19	30.01	29.76	29.16	27.93	25.58	26.59
	M2	6.81	7.47	6.75	5.55	4.79	5.26	3.21	6.66	5.18
	M3	−0.36	−0.51	−0.92	−1.46	−0.32	0.03	1.14	0.82	1.12
	M4	−0.45	0.11	0.10	1.44	0.78	1.34	4.78	1.11	2.47
5 yr. old										
	NAmp	−6.48	−10.32	−7.63	−2.99	−5.30	−3.54	−1.44	−7.12	−5.58
	Slope	−16.95	−12.77	−10.45	16.25	4.18	13.93	−3.02	−29.02	−25.77
	M1	26.67	27.53	27.79	30.11	29.88	30.53	28.63	23.79	25.49
	M2	5.39	6.95	7.10	5.37	6.25	4.60	2.86	5.60	6.17
	M3	0.98	0.26	0.29	−0.72	−0.02	−0.28	1.27	1.68	1.47
	M4	1.31	−0.37	−0.19	0.53	−0.33	0.59	3.93	3.85	3.79
Adult										
	NAmp	−9.16	−12.66	−12.14	−4.52	−7.62	−3.66	−4.96	−10.52	−8.53
	Slope	−20.43	−19.04	−18.58	−2.79	−7.20	−20.43	−22.06	−46.44	−33.20
	M1	26.07	26.60	26.74	29.42	28.92	27.33	26.87	22.94	24.24
	M2	6.00	6.99	7.81	5.44	5.92	3.50	4.48	3.36	5.15
	M3	0.86	0.40	0.32	0.23	0.30	1.50	1.51	1.70	1.44
	M4	1.24	0.34	0.13	−0.40	−0.43	3.94	3.41	4.25	2.72

A. Normalized amplitude

For the dependent measure of normalized amplitude (rms amplitude in decibels of the entire stop burst relative to the strongest component in the following vowel), a main effect of place [$F(2,64)=68.99$, $p<0.001$, $\eta^2=0.68$] was obtained. Pairwise comparisons indicated that all three places of articulation were significantly ($p<0.01$) different from each other in terms of normalized amplitude (−10.1 dB for /p/, −4.3 dB for /t/, and −6.1 dB for /k/). A main effect of vowel context was also found to be significant [$F(2,64)=47.86$, $p<0.001$, $\eta^2=0.60$], with differences in the normalized amplitude of the stop depending on the articulation of the following vowel. All three vowel contexts, /i ɑ u/, produced significantly ($p<0.001$) different normalized amplitudes, with mean values of −5.0, −8.9, and −6.7 dB, respectively. Although a place-by-vowel interaction was also found to be significant [$F(4,128)=5.42$, $p<0.01$], the effect size was relatively small ($\eta^2=0.12$).

B. Spectral measures

1. Spectral slope

Significant differences in spectral slopes were found across place of stop articulation [$F(2,64)=116.23$, $p<0.001$, $\eta^2=0.78$]. Subsequent pairwise comparisons ($p<0.001$) demonstrated that the mean spectral slope of all three stops was significantly different from each other (−3.83 dB/kHz for /p/, 18.34 dB/kHz for /t/, and −16.58 dB/kHz for /k/). In addition, a significant effect of vowel context [$F(2,64)=53.56$, $p<0.001$, $\eta^2=0.63$] indicated that the spectral slope values of the stop articulations were different from each other depending on the articulation of the following vowel (the mean slope for /i, ɑ, u/ contexts were 9.03, −5.60, and −5.50 dB/kHz, respectively). The effect of vowel context was mainly due to the significantly increased slope values of stops preceding an /i/ vowel ($p<0.001$). The ANOVA also yielded a significant place-by-vowel interaction [$F(4,128)=18.95$, $p<0.001$, $\eta^2=0.37$], characterized by a significantly elevated mean slope for /t/ and /k/ when immediately followed by an /i/ vowel ($p<0.001$).

Interestingly, a main effect was obtained for both the sex of the speaker [$F(1,32)=9.32$, $p<0.01$, $\eta^2=0.23$] and the age group [$F(3,32)=9.19$, $p<0.001$, $\eta^2=0.46$]. In addition, a significant sex-by-age group interaction was also noted [$F(3,32)=3.10$, $p<0.05$, $\eta^2=0.23$]. As can be seen in Fig. 1,

TABLE II. Acoustic measures from female speakers for three classes of voiceless American English stop consonants, grouped as a function of speaker age, place of articulation, and vowel context. Normalized amplitude in dB (NAmp), spectral slope in dB/kHz (slope), and the first four spectral moments (mean in ERB, M1; variance in kHz^2, M2; skewness, M3; and kurtosis, M4).

		Bilabial-/p/			Alveolar-/t/			Velar-/k/		
		/i/	/ɑ/	/u/	/i/	/ɑ/	/u/	/i/	/ɑ/	/u/
3 yr. old										
	NAmp	−6.89	−13.90	−14.36	−6.01	−7.33	−4.93	−6.42	−8.30	−6.77
	Slope	6.04	0.93	2.32	40.40	20.67	24.85	10.68	−26.01	−30.65
	M1	28.44	28.93	28.95	31.01	29.86	30.29	29.17	24.39	25.92
	M2	7.02	6.31	6.87	4.53	5.53	4.94	3.89	5.17	4.64
	M3	−0.22	−0.14	−0.12	−1.84	−0.72	−0.75	0.44	1.39	1.26
	M4	0.03	0.01	−0.61	3.13	0.05	0.37	0.85	7.26	4.06
4 yr. old										
	NAmp	−6.71	−10.67	−5.56	−1.97	−4.12	−3.92	−1.33	−8.80	−3.61
	Slope	−4.18	−4.18	−0.46	34.59	21.13	24.38	26.70	−16.95	−19.27
	M1	28.20	27.54	28.61	31.17	30.61	29.71	29.91	25.92	27.21
	M2	6.06	7.13	6.45	4.99	5.56	5.39	4.96	6.56	4.98
	M3	0.06	0.15	0.08	−1.80	−0.51	−0.56	−0.75	1.19	1.08
	M4	−0.07	−0.32	−0.27	2.34	0.01	0.22	−0.33	2.80	3.02
5 yr. old										
	NAmp	−6.96	−9.55	−8.99	−3.45	−5.52	−3.45	−5.20	−7.33	−6.64
	Slope	6.97	14.16	17.41	49.92	28.79	32.74	13.70	−20.67	−21.36
	M1	28.58	29.85	30.26	31.58	31.48	30.55	29.11	26.56	25.95
	M2	6.35	7.86	6.78	3.25	4.04	5.03	3.19	9.16	6.91
	M3	−0.24	−1.10	−1.04	−2.11	−0.97	−1.18	0.47	0.15	0.19
	M4	−0.32	0.16	0.12	5.17	1.18	1.34	1.08	−0.68	1.00
Adult										
	NAmp	−9.55	−14.93	−16.12	−4.51	−4.81	−2.24	−6.66	−10.67	−9.35
	Slope	−15.56	−13.24	−15.09	31.58	5.57	−0.46	−9.75	−30.19	−35.06
	M1	26.94	27.16	27.53	32.24	30.21	29.57	27.73	24.37	24.54
	M2	6.49	7.57	6.16	3.56	4.54	3.91	4.10	5.15	4.66
	M3	0.34	−0.06	0.40	−1.32	−0.14	0.53	1.27	1.54	1.83
	M4	−0.15	−0.70	0.12	2.58	0.66	0.57	3.08	4.14	4.73

sex-specific differences ($p<0.001$) in spectral slope began with the 5-year-old speakers and extended to the adults, with male speakers showing a significant decrease in spectral slope when compared to female speakers of similar age.

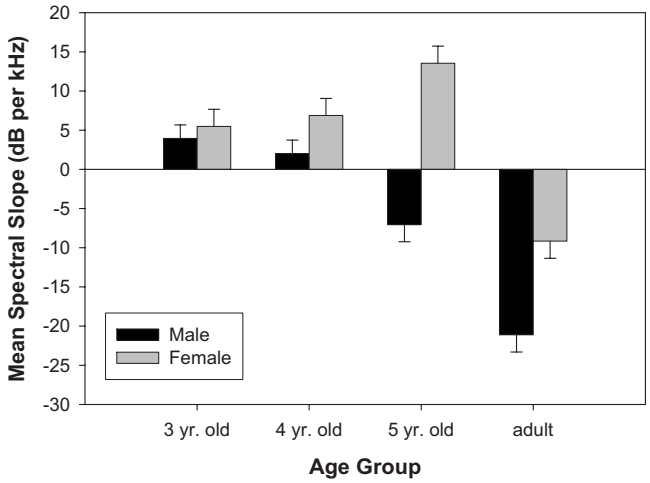
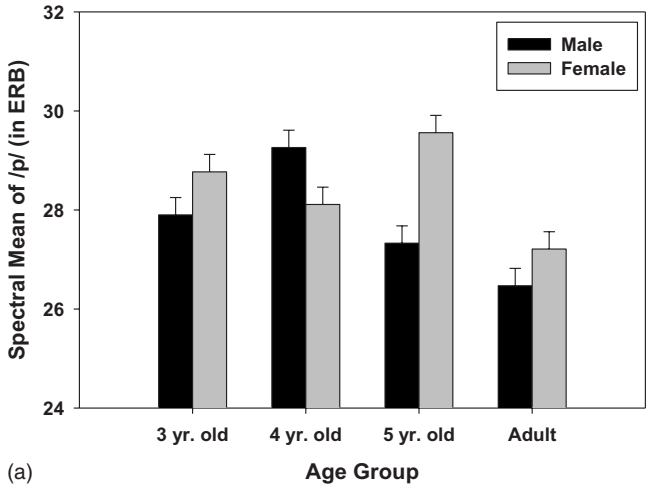


FIG. 1. Spectral slope as a function of the sex of the speaker and age group.

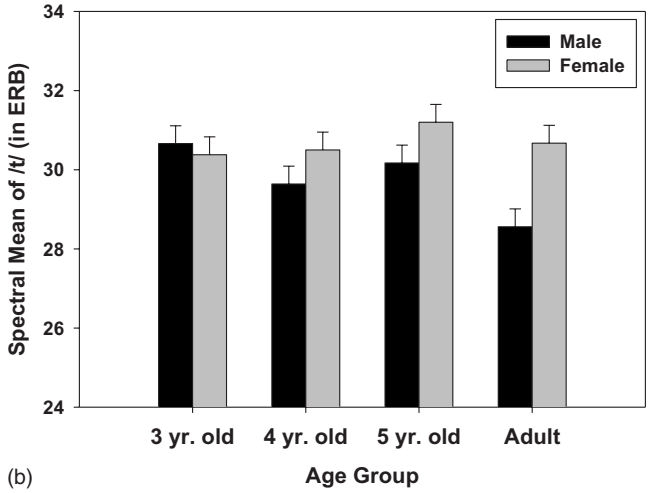
2. Spectral mean

As expected, the statistical analysis revealed a significant main effect for place of articulation [$F(2,64)=191.93$, $p<0.001$], characterized by a strong effect size ($\eta^2=0.86$). *Post hoc* analyses indicated significant differences ($p<0.001$) between all three places of articulation. Collapsed across speaker and vowel context the stop consonants, /p t k/, exhibited spectral means of 28.08, 30.22, and 26.45 ERB, respectively. There was also a significant effect of vowel context [$F(2,64)=36.74$, $p<0.001$, $\eta^2=0.53$] and a significant place-by-vowel interaction effect [$F(4,128)=30.11$, $p<0.001$, $\eta^2=0.48$]. Pairwise comparisons indicated that the main effect of vowel context was attributed primarily to the elevated spectral mean of the stop burst preceding an /i/ vowel ($p<0.001$), whereas the differences between /ɑ/ and /u/ were not found to be statistically significant. The /i/ vowel context effect was increased in velar stops (/k/) and relatively reduced in bilabial (/p/) and alveolar stops (/t/).

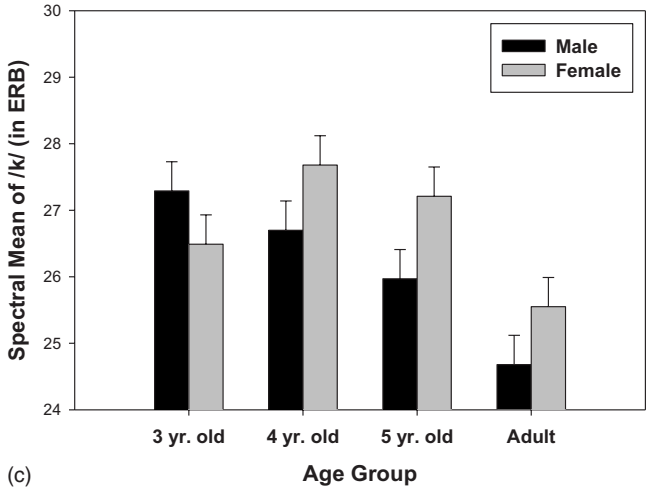
As expected, significant differences in spectral mean were found as a function of both the sex of the speaker [$F(1,32)=8.03$, $p=0.008$, $\eta^2=0.20$] and the age group



(a)



(b)



(c)

FIG. 2. [(a)–(c)] Spectral mean as a function of place of stop articulation, speaker sex, and age group. Linear measures in hertz were converted to an ERB scale (Glasberg and Moore, 1990; Moore, 1997) prior to analysis.

$[F(3,32)=7.65,\ p=0.001,\ \eta^2=0.42]$, with female and child speakers exhibiting higher overall mean values than the male and adult speakers. Interestingly, the data also contained a significant speaker sex-by-age group $[F(3,32)=4.31,\ p=0.012,\ \eta^2=0.29]$ and place-by-speaker sex-by-age group interactions $[F(6,64)=3.35,\ p=0.006,\ \eta^2=0.24]$. As shown in Fig. 2, significant sex-specific differences ($p<0.001$)

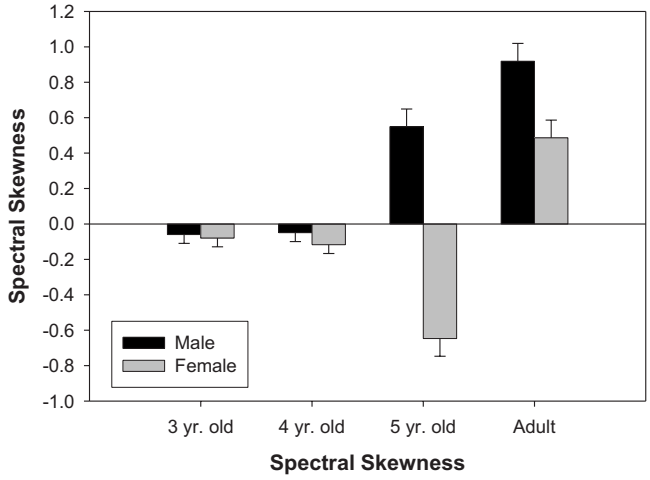


FIG. 3. Spectral skewness as a function of the sex of the speaker and age group.

were exhibited for /p/ and /t/ productions starting at age 4 and by age 5 for /k/. Within these two age groups, stop productions from male speakers were found to have a significant decrease in spectral mean when compared to female productions of the same age group, except for the 4-year old groups' productions of /p/, in which the spectral mean was significantly elevated.

3. Spectral variance

A significant main effect of place of stop articulation $[F(2,64)=31.86,\ p<0.001,\ \eta^2=0.50]$ and vowel context $[F(2,64)=18.17,\ p<0.001,\ \eta^2=0.36]$ was also found for the measure of spectral variance. The spectral variance of the bilabial stops was significantly higher ($p<0.001$) than the other two places of articulation (the variances for /p t k/ were 6.7, 4.9, and 5.1 kHz^2, respectively). In addition, the variance of the stop burst was significantly lower ($p<0.001$) when followed by an /i/ vowel. However, this pattern was reversed for alveolar stops (/t/), which explains a small (an effect size of $\eta^2=0.12$) yet significant place-by-vowel interaction effect $[F(4,128)=4.51,\ p<0.01]$.

4. Spectral skewness

A main effect of place $[F(2,64)=87.47,\ p<0.001]$, with a large effect size ($\eta^2=0.73$) was obtained for spectral skewness. *Post hoc* analyses ($p<0.001$) indicated that the mean spectral skewness of all three stops was significantly different from each other (−0.017 for /p/, −0.640 for /t/, and 0.997 for /k/). In addition, an effect of vowel context $[F(2,64)=15.42,\ p<0.001,\ \eta^2=0.32]$, as well as a significant place-by-vowel interaction $[F(4,128)=14.97,\ p<0.001,\ \eta^2=0.32]$ were obtained from the ANOVA. The effect of vowel context and associated interaction were largely due to the fact that the spectral skewness of the alveolar stop burst decreased significantly ($p<0.001$) when preceding an /i/ vowel.

The ANOVA also revealed main effects for both the sex $[F(1,32)=8.96,\ p<0.01,\ \eta^2=0.22]$ and age $[F(3,32)=7.23,\ p<0.001,\ \eta^2=0.40]$ of the speaker, as well as a significant sex-by-age group interaction $[F(3,32)=3.59,\ p<0.03,\ \eta^2=0.25]$, as shown in Fig. 3. Similar to the pattern of results

obtained for spectral mean, when looking at skewness there were distinct sex-specific differences which began with the 5-year-old age group and were extended to the adults. *Post hoc* tests demonstrated that within these two age groups the sex-specific differences were significant ($p<0.001$).

5. Spectral kurtosis

Analysis revealed only one significant effect for the measure of kurtosis, that being place of articulation [$F(2,64)=20.24$, $p<0.001$, $\eta^2=0.39$], with subsequent comparisons indicating that all three places were significantly different ($p<0.01$). The spectral kurtosis was found to increase as the stop articulation moved posterior in the oral cavity (0.06 for /p/, 1.13 for /t/, and 2.72 for /k/).

IV. DISCUSSION

Findings indicated that the dependent measures of normalized amplitude, spectral slope, and all four spectral moments varied significantly as a function of place of articulation. Subsequent pairwise comparisons revealed that, with the exception of spectral variance, the acoustic and spectral measures varied significantly across all three places of stop articulation. These results are similar to previous findings (i.e., Forrest *et al.*, 1988; Nittrouer, 1995), which indicate that the spectral moments measures of mean and skewness differentiate alveolar and velar stops. In addition, subsequent analysis demonstrated that all three places of stop articulation were significantly different from each other in terms of normalized amplitude and spectral slope; measures not reported by Forrest *et al.* (1988) and Nittrouer (1995). Unlike Forrest *et al.* (1988), this study found significant differences between stops in terms of the second spectral moment (variance), a spectral measure that has not been included in a number of spectral moments studies of child productions (e.g., Forrest *et al.*, 1990; Nittrouer, 1995). The results of this study found that spectral variance was useful in distinguishing /p/ from /t/ and /k/ across all age groups.

Not surprisingly, the results of this study indicate that the acoustic structure of the stop burst also varied as a function of the following vowel. Significant vowel context effects were found for the measures of spectral slope and the first three spectral moments. Nittrouer (1995) postulated that similar vowel context effects for spectral mean were primarily the result of changes in the acoustic parameters of stop productions when followed by a high front vowel. In theory, this anticipatory action would produce a shortening of the anterior resonating cavity and thereby result in an increase in the spectral mean. Due to the relative position of the tongue during production, this theory is a possible explanation for differences in velar productions, but unlikely for alveolar stops.

Findings from locus equation studies (Gibson and Ohde, 2007; Sussman *et al.*, 1992) may provide some evidence regarding this interpretation. These studies examined locus equations as a measure of coarticulation across voiced stop productions in young children. Results indicated that voiced alveolar stops exhibited less coarticulation than voiced bilabial or velar stops. However, comparisons between slope values from locus equations and spectral moments studies should be interpreted with caution considering the differences in analysis procedures.

Relevant to the reported differences in normalized amplitude in this study, previous research has indicated that the amplitude of the stop burst does have an effect on the perception of the place of articulation for both voiceless and voiced stop consonants (e.g., Blumstein and Stevens, 1980; Hedrick and Jesteadt, 1996; Ohde and Stevens, 1983; Repp, 1984). It is unclear if the spectral amplitude and shape of the stop burst provide an invariant and independent cue to the perception of place of articulation (Blumstein and Stevens, 1979) or if the cue of relative burst amplitude is context dependent, determined in part by the acoustic components of the following vowel (Dorman *et al.*, 1977). In support of the latter approach, the findings of the current study found that the normalized amplitude of the stop burst was significantly affected by the following vowel context. This result may provide evidence that children at an earlier stage of development, as young as 3 years of age, exhibit phoneme specific patterns of anticipatory coarticulation (Gibson and Ohde, 2007; Sussman *et al.*, 1996). Within the scope of this study, these results also indicate that male and female children appear to produce normalized amplitude in a similar manner.

As expected, female and child speakers exhibited higher overall spectral mean values. However, significant sex and age group differences were also found for the spectral measures of slope and skewness. Of particular interest, significant speaker sex-by-age interaction effects were found for spectral slope, mean, and skewness. In general, sex-specific differences for these measures began with the 5-year-old speakers and extended to the adults. For some spectral measures, the age at which these differences emerged was also dependent on the specific type of stop articulation. For example, sex-specific differences in spectral mean for bilabial stop productions were noted in the 5-year-old speakers, whereas such differences for alveolar and velar stops were found to emerge at 4 years of age. It is unclear why sex-specific differences in spectral mean were delayed in bilabial stops relative to alveolar and velar stops or why the 4-year old bilabial productions were elevated. Normative research examining the developmental age of stop consonant acquisition has indicated that bilabial stop consonants are typically acquired at the same time or several months earlier than alveolar and velar stops (Arlt and Goodban, 1976; Smit *et al.*, 1990); thus it is unlikely that developmental age is a factor in the differences noted in this study. Sex-differences across stop type may be due to the manner of the articulations required, namely, a bilabial closure as compared to a tongue-to-palate contact patterns.

Acoustic differences in the speech of male and female adult speakers can be largely explained by sex-related variation in characteristics such as fundamental frequency and formant ratios. For example, typical male speakers generally exhibit lower fundamental and formant frequencies than female speakers when producing vowel segments (Hillenbrand *et al.*, 1995; Mattingly, 1966; Peterson and Barney, 1952). In adults, these acoustic differences are in part the result of

sexual dimorphism of anatomical factors such as vocal tract length and shape, as well as vocal fold size (Fitch and Giedd, 1999; Titze, 1989).

However, findings from studies with children indicate that sex-related acoustic differences may be more complex than can be reasonably explained by nonuniform variation in vocal tract morphology. Sexual dimorphism of the vocal tract in children is generally considered to begin at peri- and postpubertal stages of physical maturation. Although anatomical studies of vocal tract morphology have historically involved a relatively small number of adult subjects (Baer *et al.*, 1991; Dang *et al.*, 1994; Moore, 1992; Story *et al.*, 1996; Sulter *et al.*, 1992), the increased availability and decreased health risks associated with MRI have provided the opportunity for more accurate morphometric research on the vocal tract anatomy of larger numbers of children. Recent large-scale MRI studies (Fitch and Giedd, 1999; Vorperian *et al.*, 2005, 2009) have provided a greater understanding of the anatomical development of the vocal tract in children. Findings from these studies indicate that the oral and pharyngeal structures of the vocal tract have ongoing, and at times, accelerated periods of growth through childhood. However, results of both studies indicated no appreciable sexual dimorphism in the vocal tract structures of younger prepubescent children.

In view of these anatomical data, it is reasonable to postulate that sex-related differences in acoustic properties of speech in children are in part the result of factors other than anatomical variation in the vocal tract. Data from this study may support the conclusion that articulatory development for some aspects of stop production follow different patterns in girls as opposed to boys, patterns which may be based on male-female archetypes present in adult production patterns. Evidence from perceptual experiments designed to evaluate listeners ability to identify the sex of a speaker from only auditory information have also provided some support for the notion that the acoustic signatures of children's speech may be the result of learned characteristics. In a perceptual study, Sachs *et al.* (1973) found that adult listeners were able to accurately identify (81%) a young speaker's sex from short passages of speech. Interestingly, an acoustic analysis of the recordings indicated that the male children participating in the study exhibited a higher average F0 than the young female speakers, yet lower formant frequencies. From these results, the authors concluded that the listeners' identifications may have been based in part on sex-related acoustic differences (e.g., formant frequency patterns, voice quality, intonation patterns) which arise from learned articulatory patterns: patterns which adhere to culturally determined articulatory patterns viewed as appropriate for each sex.

The findings of this study indicate that in terms of speech production the spectral slope, mean, and skewness differed as a function of the speaker's sex. However, the perceptual relevance and physiologic mechanisms for these acoustic differences remains unclear. Research has indicated that listeners' are able to identify gender from the speech of children as young as 4 years of age (Perry *et al.*, 2001; Sachs *et al.*, 1973). Although the perceptual distinction of gender may primarily be determined by formant characteristics of vowel segments (Bennett, 1981; Whiteside, 2001), it is unclear if additional acoustic cues, such as those included in this study, might also contribute to such distinctions.

Results from the current study indicate that sex-specific differences in the spectral speech patterns of young children may be associated with learned or behavioral factors that affect articulatory development, beginning at approximately 5 years of age. This conclusion is similar to previous research examining children's obstruent productions (Nissen and Fox, 2005;Whiteside and Marshall, 2001), vowel formant frequencies (Bennett, 1981; Whiteside, 2001), and the perceptual sex identification of children's voices (Perry *et al.*, 2001; Sachs *et al.*, 1973), which have also indicated that sex-specific acoustic and spectral differences in children's speech are not fully explained by anatomic differences alone, but likely the result of cultural or sociophonetic factors.

V. CONCLUSIONS

The findings of this study were based on acoustic measures collected from children in discrete age groups; however, it would be of interest, in future studies, to examine children's speech development in a longitudinal manner. Although more difficult to conduct, this type of study would provide a more comprehensive understanding of possible sex-specific differences in children's speech, while having greater control of inter-speaker variation. Future studies are needed to understand if such measures are perceptually salient, in isolation or in conjunction with other acoustic cues. It would also be of interest to conduct additional research examining the physiologic basis of these spectral characteristics acoustic dimension and thereby correlate spectral moment measures with articulatory movement or vocal fold shape. Furthermore, the acoustic measures utilized in the present study were limited to static acoustic and spectral cues. Future studies might examine how children develop dynamic speech cues, such as the change in spectral slope or mean over the duration of the stop burst. Despite these limitations, it is hoped the findings of this study will contribute to a greater understanding of speech development in prepubescent children and provide additional insight into possible sex-specific patterns of stop articulation, as well as the developmental stage such differences might typically emerge.

ACKNOWLEDGMENTS

This research was supported by a Graduate Student Alumni Research Association grant from The Ohio State University and a McKay School of Education Grant from Brigham Young University.

Adobe Systems Incorporated. (**2003**). ADOBE AUDITION (Version 1.3) (Computer software), San Jose, CA.

Arlt, P. B., and Goodban, M. T. (**1976**). "A comparative study of articulation acquisition as based on a study of 240 normals, aged three to six," Language, Speech, and Hearing Services in Schools **7**, 173–180.

Baer, T., Gore, J. C., Gracco, L. C., and Nye, P. W. (**1991**). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," J. Acoust. Soc. Am. **90**, 799–828.

Behrens, S., and Blumstein, S. E. (**1988a**). "Acoustic characteristics of Eng-

lish voiceless fricatives: A descriptive analysis," J. Phonetics **16**, 295–298.
Behrens, S., and Blumstein, S. E. (**1988b**). "On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants," J. Acoust. Soc. Am. **84**, 861–867.
Bennett, S. (**1981**). "Vowel formant frequency characteristics of preadolescent males and females," J. Acoust. Soc. Am. **69**, 231–238.
Blumstein, S. E., and Stevens, K. N. (**1980**). "Perceptual invariance and onset spectra for stop consonants in different vowel environments," J. Acoust. Soc. Am. **67**, 648–662.
Busby, P., and Plant, G. (**1995**). "Formant frequency values of vowels produced by preadolescent boys and girls," J. Acoust. Soc. Am. **97**, 2603–2606.
Byrd, D. (**1992**). "Preliminary results on speaker-dependent variation in the TIMIT database," J. Acoust. Soc. Am. **92**, 593–596.
Byrd, D. (**1994**). "Relations of sex and dialect to reduction," Speech Commun. **15**, 39–54.
Dang, J., Honda, K., and Suzuki, H. (**1994**). "Morphological and acoustic analysis of the nasal and the paranasal cavities," J. Acoust. Soc. Am. **96**, 2088–2100.
Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (**1977**). "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Percept. Psychophys. **22**, 109–122.
Fant, G. (**1966**). "A note on vocal tract size factors and nonuniform F-pattern scaling," *Speech Sounds and Features* (MIT, Cambridge, MA), pp. 84–93.
Fant, G., Kruckenberg, A., and Nord, L. (**1991**). "Prosodic and segmental speaker variation," Speech Commun. **10**, 521–531.
Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.
Fitzsimmons, M., Sheahan, N., and Staunton, H. (**2001**). "Gender and the integration of acoustic dimensions of prosody: Implications for clinical studies," Brain Lang **78**, 94–108.
Forrest, K., Weismer, G., Elbert, M., and Dinnsen, D. A. (**1994**). "Spectral analysis of target-appropriate /t/ and /k/ produced by phonologically disordered and normally articulating children," Clin. Linguist. Phonetics **8**, 267–281.
Forrest, K., Weismer, G., Hodge, M., and Dinnsen, D. A. (**1990**). "Statistical analysis of word-initial /k/ and /t/ produced by normal and phonologically disordered children," Clin. Linguist. Phonetics **4**, 327–340.
Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.
Fox, R. A., and Nissen, S. L. (**2005**). "Sex-related acoustic changes in voiceless English fricatives," J. Speech Lang. Hear. Res. **48**, 753–765.
Gibson, T., and Ohde, R. N. (**2007**). "F2 locus equations: Phonetic descriptors of coarticulation in 17–22 month old children," J. Speech Lang. Hear. Res. **50**, 97–108.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Goldman, R., and Fristoe, M. (**1986**). *Goldman-Fristoe Test of Articulation* (American Guidance Service, Circle Pines, MN).
Graddol, D. (**1986**). "Discourse specific pitch behavior," in *Intonation in Discourse*, edited by C. Johns-Lewis (Croom Helm, London), pp. 221–237.
Hedrick, M. S., and Jesteadt, W. (**1996**). "Effect of relative amplitude, presentation level, and vowel duration on perception of voiceless stop consonants by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **100**, 3398–3407.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.
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.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.
Mattingly, I. (**1966**). "Speaker variation and vocal-tract size," J. Acoust. Soc. Am. **39**, S1219A.
Mendoza, E., Valencia, N., Munoz, J., and Trujillo, H. (**1996**). "Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS)," J. Voice **10**, 59–66.
Miccio, A. W. (**1996**). "A spectral moments analysis of the acquisition of word-initial voiceless fricatives in children with normal and disordered phonologies," Ph.D. thesis, Indiana University, Bloomington, IN.
Moore, B. C. J. (**1997**). *An Introduction to the Psychology of Hearing* (Academic, New York).
Moore, C. A. (**1992**). "The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images," J. Speech Hear. Res. **35**, 1009–1023.
Nissen, S. L., and Fox, R. A. (**2005**). "Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective," J. Acoust. Soc. Am. **118**, 2570–2578.
Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.
Nittrouer, S. (**1995**). "Children learn separate aspects of speech production at different rates: Evidence from spectral moments," J. Acoust. Soc. Am. **97**, 520–530.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). "The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults," J. Speech Lang. Hear. Res. **32**, 120–132.
Ohde, R. N., and Stevens, K. N. (**1983**). "Effect of burst amplitude on the perception of stop consonant place of articulation," J. Acoust. Soc. Am. **74**, 706–714.
Perry, T. L., Ohde, R. N., and Ashmead, D. H. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**, 2988–2998.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 585–594.
Repp, B. H. (**1984**). "Closure duration and release burst amplitude cues to stop consonant manner and place of articulation," Lang Speech **27**, 245–254.
Ryalls, J. H., Zipprer, A., and Baldauff, P. (**1997**). "A preliminary investigation of the effects of gender and race on voice onset time," J. Speech Lang. Hear. Res. **40**, 642–645.
Sachs, J., Liberman, P., and Erickson, D. (**1973**). "Anatomical and cultural determinants of male and female speech," in *Language Attitudes*, edited by R. W. Shuy and R. W. Fasold (Georgetown University Press, Washington DC), pp. 74–84.
Smit, A. B., Hand, L., Feilinger, J. J., Bernthal, J. E., and Bird, A. (**1990**). "The Iowa articulation norms project and its Nebraska replication," J. Speech Hear Disord. **55**, 779–798.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.
Sulter, A. M., Miller, D. G., Wolf, R. F., Schutte, H. K., Wit, H. P., and Mooyaart, E. L. (**1992**). "On the relation between the dimensions and resonance characteristics of the vocal tract: A study with MRI," Magn. Reson. Imaging **10**, 365–373.
Sussman, H., Hoemeke, K., and McCaffrey, H. (**1992**). "Locus equation as an index of coarticulation for place of articulation distinctions in children," J. Speech Hear. Res. **35**, 769–781.
Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., and Smith, J. (**1996**). "Consonant-vowel interdependencies in babbling and early words: preliminary examination of a locus equation approach," J. Speech Hear. Res. **39**, 424–433.
Swartz, B. L. (**1992**). "Gender differences in voice onset time," Percept. Mot. Skills **75**, 983–992.
Sweeting, P. M., Baken, R. J. (**1982**). "Voice onset time in a normal-ages population," J. Speech Hear. Res. **25**, 129–134.
Titze, I. R. (**1989**). "Physiologic and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.
Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.
Vorperian, H. K., Wang, S., Chung, M. K., Schimek, E. M., Durtschi, R. B., Kent, R. D., Ziegert, A. J., and Gentry, L. R. (**2009**). "Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study," J. Acoust. Soc. Am. **125**, 1666–1678.

Whiteside, S. P. (**1996**). "Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences," J. Int. Phonetic Assoc. **26**, 23–40.
Whiteside, S. P. (**2001**). "Sex-specific fundamental and formant frequency patterns in a cross-sectional study," J. Acoust. Soc. Am. **110**, 464–478.
Whiteside, S. P., and Hodgsen, C. (**2000**). "Speech patterns of children and adults elicited via a picture-naming task: An acoustic study," Speech Commun. **32**, 267–285.
Whiteside, S. P., and Irving, C. J. (**1997**). "Speakers' sex differences in voice onset time: Some preliminary findings," Percept. Mot. Skills **85**, 459–463.
Whiteside, S. P., and Marshall, J. (**2001**). "Developmental trends in voice onset time: Some evidence or sex difference," Phonetica **58**, 196–210.

A cross-dialect acoustic description of vowels: Brazilian and European Portuguese

Paola Escudero[a] and Paul Boersma
Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210, 1012VT Amsterdam, The Netherlands

Andréia Schurt Rauber
Center for Studies in the Humanities, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

Ricardo A. H. Bion
Department of Psychology, Stanford University, Jordan Hall, Building 420, 450 Serra Mall, Stanford, California 94305

(Received 18 July 2008; revised 22 June 2009; accepted 24 June 2009)

This paper examines four acoustic correlates of vowel identity in Brazilian Portuguese (BP) and European Portuguese (EP): first formant (F1), second formant (F2), duration, and fundamental frequency (F0). Both varieties of Portuguese display some cross-linguistically common phenomena: vowel-intrinsic duration, vowel-intrinsic pitch, gender-dependent size of the vowel space, gender-dependent duration, and a skewed symmetry in F1 between front and back vowels. Also, the average difference between the vocal tract sizes associated with /i/ and /u/, as measured from formant analyses, is comparable to the average difference between male and female vocal tract sizes. A language-specific phenomenon is that in both varieties of Portuguese the vowel-intrinsic duration effect is larger than in many other languages. Differences between BP and EP are found in duration (BP has longer stressed vowels than EP), in F1 (the lower-mid front vowel approaches its higher-mid counterpart more closely in EP than in BP), and in the size of the intrinsic pitch effect (larger for BP than for EP). © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3180321]

PACS number(s): 43.70.Fq, 43.70.Kv, 43.72.Ar [AL] Pages: 1379–1393

I. INTRODUCTION

The aim of this article is to investigate the acoustic characteristics of the seven oral vowels that Brazilian Portuguese (BP) and European Portuguese (EP) have in common in stressed position, namely, the vowels /i, e, ɛ, a, ɔ, o, u/, and thereby to find out what aspects of the Portuguese vowel inventory are universal, Portuguese-specific, or dialect-specific.

Studies that described Portuguese vowels in phonological or impressionistic articulatory terms (e.g., Câmara, 1970; Mateus, 1990; Bisol, 1996; Mateus and d'Andrade, 1998, 2000; Barroso, 1999; Moraes, 1999; Cristófaro Silva, 2002; Barbosa and Albano, 2004; Mateus *et al.*, 2005) agree that the Portuguese vowel inventory has an internal symmetry: apart from the central low vowel /a/, there are three unrounded front vowels (i, e, ɛ) and three rounded back vowels (u, o, ɔ) between which we can identify three pairings, namely, two high vowels (i-u), two higher-mid vowels (e-o) and two lower-mid vowels (ɛ-ɔ).[1] Because of the general relation between vowel height and the first formant (F1), we expect that the members of each pair have almost identical F1 values, and one research question is whether this is true for Portuguese. In fact, languages with large symmetric vowel inventories have been reported to have slightly higher F1 values for each back vowel as compared to its corresponding front vowel: American English (Peterson and Barney, 1952; Clopper *et al.*, 2005; Strange *et al.*, 2007), Parisian French (Strange *et al.*, 2007), Northern German (Strange *et al.*, 2007), Dutch (Koopmans-van Beinum, 1980),[2] and BP (Moraes *et al.*, 1996, p. 35; Seara, 2000, pp. 80, 91, 102, 112, and 141); one research question is whether this holds for both varieties of Portuguese.

Portuguese has been reported to have no phonological length distinctions in vowels (Falé, 1998, p. 257; Mateus *et al.*, 2005, p. 140). For such languages, it has been reported that low vowels tend to have a longer duration than high vowels (e.g., for French: Rochet and Rochet, 1991, p. 57, Fig. 7b). The effect can even be seen in languages that do have phonological length, such as English (House and Fairbanks, 1953, p. 111). In fact, the effect is so widespread that Lehiste (1970, p. 18) calls it *intrinsic vowel duration*. As for the cause of the effect, a recent review on controlled and mechanical properties of speech (Solé, 2007, p. 303) follows Lindblom (1967) and Lehiste (1970, pp. 18 and 19) in regarding it as a universal physiological property of speech production: open vowels require more jaw lowering, hence more time, than closed vowels. Since speakers can in principle control duration and F1 independently, it is, however, an open question whether Portuguese follows this cross-linguistic tendency or not. If Portuguese does follow the tendency, it is relevant to know the extent to which Portuguese does this; if this extent is larger than in other languages, it would be evidence for an exaggeration of the use of duration as a cue to vowel height.

[a]Author to whom correspondence should be addressed. Electronic mail: paola.escudero@uva.nl

Portuguese has never been reported to have phonological tone. For such languages, it has been reported that low vowels tend to have a lower F0 than high vowels (for a long list of languages, see Whalen and Levitt, 1995). Lehiste and Peterson (1961) call the effect *intrinsic fundamental frequency*. Again, articulatory explanations have been proposed, mainly in terms of a pull of the tongue on the larynx (Ohala and Eukel, 1987), but speakers can also control F0 and F1 independently, so it is an open question whether Portuguese follows this universal tendency or not, and if so, whether it does so to a larger extent than other languages, i.e., whether it exaggerates F0 differences as a cue to vowel height.

Several Romance languages with a comparable symmetric seven-vowel inventory as Portuguese show signs that the lower-mid vowels are merging with the higher-mid vowels in some regional varieties: Italian (Maiden, 1997, p. 8), French (Landick, 1995), and Catalan (Recasens and Espinosa, 2009). One of our research questions is whether any signs of future merger can be observed in either of the two Portuguese varieties under scrutiny.

As for differences between female and male speakers, we expect Portuguese to exhibit the following near-universal effects. First, females have generally higher F0 and formants than males. Second, women tend to have a larger vowel space than men, even along logarithmic scales, i.e., in terms of a ratio of the F1 values of /a/ versus /i, u/; the cause of this effect has been sought in the physiology (Simpson, 2001) as well as in the idea that males reduce their F1 space size because their F1 values are easier to discriminate by listeners than female F1 values (Goldstein, 1980; Ryalls and Lieberman, 1982; Diehl *et al.*, 1996). Third, women have longer vowel durations than men (Simpson and Ericsdotter, 2003); the source of this effect has been sought in the physiology (Simpson, 2001, 2002) as well as in the idea that women put more effort in trying to speak clearly (Byrd, 1992; Whiteside, 1996). As for differences between BP and EP, Moraes *et al.* (1996) report, comparing their BP results with the EP results of Delgado-Martins (1973), that /i/ and /u/ have a higher F1 in BP than in EP; the question is whether this result will still hold when comparing BP and EP with identical measurement methods.

Answering these research questions on the basis of earlier acoustic descriptions of Portuguese vowels (Delgado-Martins, 1973, 2002, pp. 41–52; Callou *et al.*, 1996; Moraes *et al.*, 1996; Seara, 2000) is difficult, because none of these studies provided direct cross-dialectal comparisons, investigated a sufficient number of speakers, included female speakers, or reported all four acoustic characteristics of all vowels; also, the results of multiple studies can hardly be combined, as a result of differences in measurement methods. The methodology employed in the present study is designed to answer the research questions with more confidence: (1) it compares the acoustic properties of BP and EP vowels, and follows as closely as possible the methods of data collection reported in Adank *et al.* (2004) in order to allow future comparisons across experiments and languages; (2) 40 speakers, 20 BP and 20 EP, produced a total of 5600 vowel tokens; (3) half of the speakers in each dialect were male and half were female; and (4) acoustic analyses were made of vowel duration, fundamental frequency, and the first two vowel formants. This methodology allows us to address all of the research questions mentioned above, as well as to explore any unpredicted differences between females and males or between BP and EP.

Finally, the present paper aims at providing reliable values for duration by measuring vowels only between voiceless consonants, and at providing typical formant values by measuring vowels only between stops and fricatives. Elicitation of multiple tokens per speaker allows us to automatically define the formant ceiling of the LPC analysis on the basis of within-speaker and within-vowel variation, thus allowing more reliable automatic formant measurements. This methodology is explained in detail so that it can be used as a reference for future studies on vowel formant analyses.

II. METHOD

A. Participants

In order to obtain relatively homogeneous and comparable groups of BP and EP participants, all participants were chosen to be highly educated young adults from the largest metropolitan area in each country. They were selected from groups of volunteers that completed a background questionnaire: if they met three requirements, they could be enlisted as speakers for the present study. The requirements were that they had lived in either São Paulo or Lisbon throughout their lives, that they did not speak any foreign language with a proficiency of 3 or more on a scale from 0 ("I don't understand a word") to 7 ("I understand like a native speaker"), and that they were undergraduate students under 30 years of age. In this way, 20 BP speakers from São Paulo and 20 EP speakers from Lisbon were selected. For each "dialect" (more precisely: "age-, social-economic-status-, and region-dependent variety of the standard language") there were equal numbers of men and women, so that the gender-dependence of the vowels could be investigated as easily as the dialect-dependence. For BP, the females' mean age was 23.2 years (standard deviation 4.3 years) and the males' mean age was 22.5 years (s.d. 4.7); for EP speakers, the females' mean age was 19.8 years (s.d. 1.5), and the males' mean age was 18.7 years (s.d. 0.8).

B. Data collection procedure

All 40 recordings were made in a quiet room with a Sony MZ-NHF800 minidisk recorder and a Sony ECM-MS907 condenser microphone, with a sample rate of 22 kHz and 16-bit quantization. The 20 BP recordings were made at the Escola Superior de Propaganda e Marketing (ESPM) in São Paulo, and the 20 EP recordings were made at the Instituto de Engenharia de Sistemas e Computadores (INESC) and at the University of Lisbon, both in Lisbon.

The target vowels /i, e, ɛ, a, ɔ, o, u/ were orthographically presented to the speakers as *i*, *ê*, *é*, *a*, *ó*, *ô*, and *u*, respectively, embedded in a sentence written on a computer screen. Each vowel was produced as the first vowel in a disyllabic CVCV sequence (C=consonant, V=vowel), where the two consonants were two identical voiceless stops or fricatives; this yielded nonce words such as /pepo/ and

/saso/ (*pêpo* and *sasso*) where the underlined vowel is the target vowel. The consonants were always voiceless so as to allow easy measurement of duration; the analysis was restricted to the five consonants /p, t, k, f, s/, i.e., the voiceless consonants that Portuguese shares with Spanish, in order to allow future cross-language comparisons. The speakers always stressed the first syllable of the nonce word, helped by the orthographic conventions of Portuguese. In the final unstressed syllable, where Portuguese has only three vowels, the participants only read the vowels /e/ and /o/, which are usually pronounced as [ɪ] and [ʊ] in BP (Cristófaro Silva, 2002, p. 86) and (if audible at all) as [i̥] and [u̥] in EP (Mateus and d'Andrade, 2000, p. 18).

The disyllabic nonce words were read in two phrasal positions, namely, in isolation and embedded in an immediately following carrier sentence similar to the one used in Adank *et al.* (2004). The sentences were read twice in two blocks; in the first block the isolated word had a final /e/, and in the second block it had a final /o/. An example of an isolated word with sentence in block 1 was therefore "*Pêpe*. Em *pêpe* e *pêpo* temos *ê*," which means '*Pêpe*. In *pêpe* and *pêpo* we have *ê*.' The corresponding example from block 2 would be "*Pêpo*. Em *pêpe* e *pêpo* temos *ê*."

The words and sentences were presented on a computer screen. In case the participants misread a word or sentence, they were asked to repeat it before the next word or sentence was presented.

Each participant thus produced six tokens of each vowel embedded in each consonant context. From these six tokens, we chose the two isolated words (i.e. one with final *e*, and one with final *o*) and the two best exemplars of the tokens embedded in the carrier sentence (one with final *e*, and one with final *o*). Two native speakers of Portuguese chose these best exemplars on the basis of their recording quality, i.e., the tokens with no background noise or hesitation during the production of the whole sentence. The final isolated vowels were not considered in the analysis. Thus, 20 productions (2 phrasal positions × 2 word-final vowels × 5 consonantal contexts) were analyzed for each of the 7 vowels of each participant. This yielded a total of 2800 vowel tokens per dialect (20 productions × 7 vowels × 20 speakers).

C. Acoustic analysis: Duration

For duration measurements, the start and end points of each of the 5600 vowel tokens were labeled manually in the digitized sound wave. Because all flanking consonants were voiceless and unaspirated, the start and end points of the vowel could be determined relatively easily by finding the first and last periods that had considerable amplitude and whose shape resembled that of more central periods, with both points of the selection chosen to be at a zero crossing of the waveform.

D. Acoustic analysis: Fundamental frequency

In order to determine the F0 of each of the 5600 vowel tokens, the computer program PRAAT (Boersma and Weenink, 2008) was used to measure the F0 curves of all recordings by the cross-correlation method, which is especially suitable for measuring short vowels. The pitch range for the analysis was set to 60–400 Hz for men and 120–400 Hz for women. If the analysis failed on any of the speaker's vowel tokens, i.e., if PRAAT considered the entire vowel center voiceless, the analysis for that token was redone in a way depending on the speaker's gender: if the analysis failed for a woman (which happened for six of the 2800 tokens, which were creaky), the analysis was retried with a pitch floor of 75 Hz, and if it failed for a man (which happened for 1 of the 2800 tokens, which was noisy), the analysis was retried with a lower criterion for voicedness. In this way, all 5600 vowel tokens eventually yielded F0 values. To get a robust measure of the F0 of the vowel, the median F0 value was taken of values measured in steps of 1 ms in the central 40% of the vowel: ignoring the first and last 30% of the vowel reduces the effect of the flanking consonants, and taking the median rather than the mean reduces the effect of F0 measurement errors.

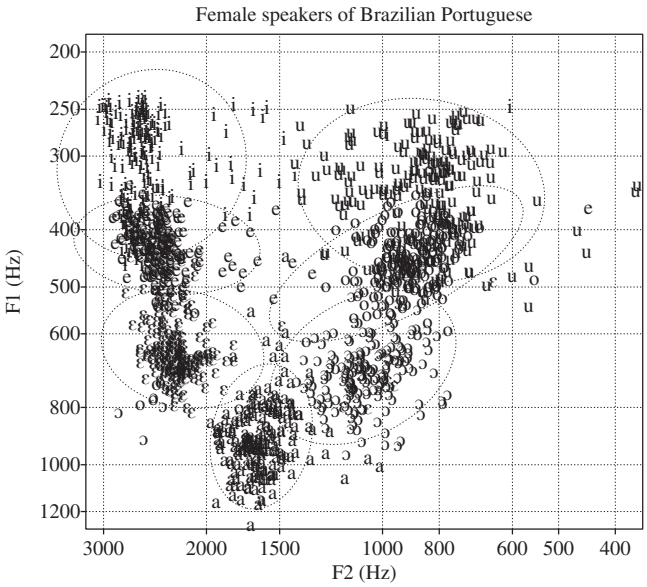


FIG. 1. The first and second formants of the 1400 vowel tokens of the Brazilian women, measured with a fixed (gender-specific) formant ceiling of 5500 Hz. The ellipses show two estimated standard deviations and have been designed to cover 86.5% of the data points (for normally distributed data).

E. Acoustic analysis: Optimized formant ceilings

For each of the 5600 vowel tokens, F1 and F2 were determined with the BURG algorithm (Anderson, 1978), as built into the PRAAT program. The analysis was done on a single window that consisted of the central 40% of the vowel.[3] As an initial approximation, PRAAT was made to search for five formants in the range from 50 Hz to 5500 Hz (for female speakers) or 5000 Hz (for male speakers). These gender-specific *formant ceilings* of 5000 and 5500 Hz reflect the different average vocal tract lengths of men versus women (since looking for five formants entails that the ceiling is meant to lie between F5 and F6, one can estimate the vocal tract length as $5c/(2\cdot\text{ceiling})$, where c is the speed of sound). The 1400 F1-F2 pairs thus measured for the Brazilian women are plotted in Fig. 1.

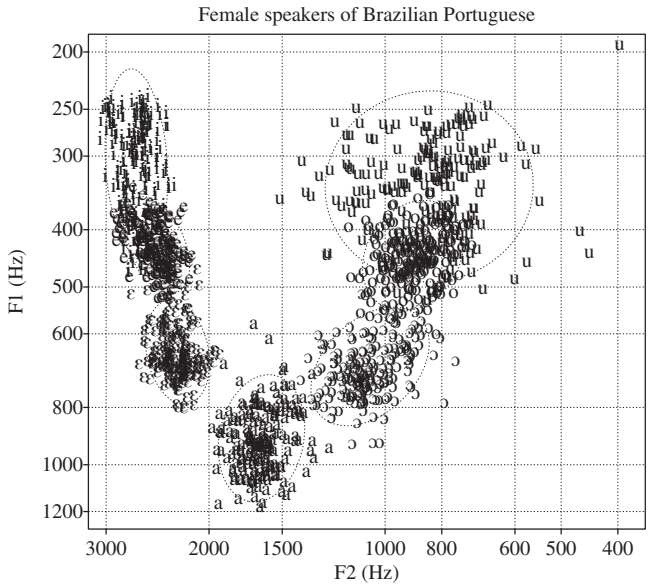


FIG. 2. The first and second formants of the 1400 vowel tokens of the Brazilian women, measured with optimized (speaker- and vowel-specific) formant ceilings.

Figure 1 shows several unlikely values for some formants: for several back vowels the F2 has been analyzed as nearly identical to F1; there are /ɔ/ and /o/ tokens in the lower left whose F2 has been incorrectly analysed as an F1, and the (weak) second tracheal resonance of /i/, between 1500 and 2000 Hz (Stevens, 1998, p. 300), has often been incorrectly analyzed as an F2. Figure 1 shows the large overlapping 2σ ellipses that these outliers cause. Such shifts in the numbering of formants indicate that the fixed gender-specific formant ceilings of 5000 and 5500 Hz could be problematic (too high for /ɔ/ and /o/, too low for /i/).

Although the manner of visualization in Fig. 1 overrepresents the outliers, a method was designed to adapt the formant ceilings to the speaker and the vowel at hand. This could be done by some general method that optimizes a formant track by a number of criteria (e.g., Nearey *et al.*, 2002: smallest bandwidths, continuity in time, correlation between original and LPC-generated spectrogram; also described by Adank, 2003, and used by Adank *et al.*, 2004), but the present paper instead takes advantage of the fortunate circumstance that each vowel was produced 20 times by each speaker.

The procedure to optimize the formant ceiling for a certain vowel of a certain speaker runs as follows. For all 20 tokens the first two formants are determined 201 times, namely, for all ceilings between 4500 and 6500 Hz in steps of 10 Hz (for women) or for all ceilings between 4000 and 6000 Hz in steps of 10 Hz (for men). From the 201 ceilings, the "optimal ceiling" is chosen as the one that yields the lowest variation in the 20 measured F1-F2 pairs. This variation is computed along the same logarithmic scales as seen in Fig. 1, namely, as the variance of the 20 log(F1) values plus the variance of the 20 log(F2) values. Thus, the procedure ends up with 280 optimal ceilings, one for each vowel of each speaker. With the 70 speaker-vowel-dependent ceilings for Brazilian women, Fig. 1 turns into Fig. 2.

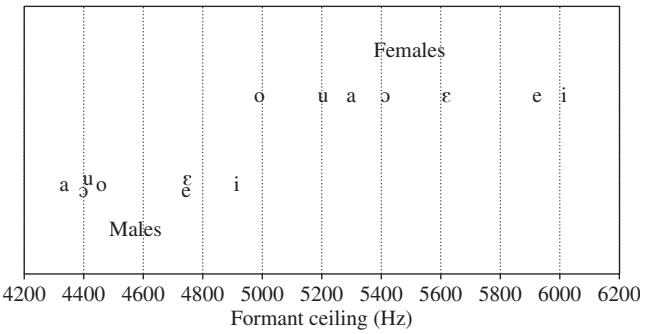


FIG. 3. Median optimal ceilings for each gender-vowel combination.

Figure 2 shows that the variation between the vowel tokens has decreased appreciably: almost all outliers have gone, and although only the variation in the formant values of a vowel *within* a speaker (not that *between* speakers) has been explicitly minimized, the 2σ ellipses have shrunk, especially in the F2 direction.

To illustrate that the ceiling optimization method does something sensible, Fig. 3 shows the effects of gender and vowel category on the optimal formant ceiling. Each vowel symbol in that figure represents the median of 20 optimal ceilings (because there are 20 speakers of each gender and the two dialects are pooled).

Figure 3 shows that both gender and vowel category have strong effects on what the optimal ceiling is. The median of the 140 optimal ceilings for the women is 5450 Hz, and the median of the 140 optimal ceilings for the men is 4595 Hz, which is a factor of 1.186 lower. This difference must reflect the difference in vocal tract lengths between men and women; it constitutes a justification for the use of different formant ceilings for men and women in computer analyses for formant frequencies. Interestingly, however, the effect of vowel category is of comparable size as the effect of gender: the median of the 40 optimal ceilings for /u/ is 4600 Hz, and the median of the 40 optimal ceilings for /i/ is 5625 Hz, which is a factor of 1.223 higher. This difference must reflect a difference in the length of the channel between upper and lower lip (rounded and protruded for /u/, and spread and retracted for /i/) and probably a difference in the height of the larynx (lowered for /u/: Ewan and Krones, 1974; Riordan, 1977). Generally, the three spread vowels /i/, /e/, and /ɛ/ come with shorter vocal tracts than the three rounded vowels /u/, /o/, and /ɔ/, and this must be reflected in the values of the higher formants (Kent and Read, 2002, p. 32); as the formant ceiling lies between F5 and F6, the formant ceiling will on average be higher for the spread than for the rounded vowels. Since a correct formant ceiling influences the reliability of the measurements of *all* formants, including F1 and F2, this result suggests that automated formant measurement methods should take into account vowel-related vocal tract lengths to a larger extent than they usually do.

III. SUMMARY OF RESULTS

Sections IV–VI present the detailed results of the acoustic measurements and statistical analyses aimed at answering

TABLE I. Geometric averages of vowel duration, F0, F1, F2, F3, and formant ceilings for female (F) and male (M) speakers of BP and EP. Between parentheses: the standard deviations, converted back to ratios of ms and Hz. Every cell represents ten speakers.

			/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
BP	Duration (ms)	F	99 (1.210)	122 (1.195)	141 (1.192)	144 (1.173)	139 (1.145)	123 (1.151)	100 (1.201)
		M	95 (1.216)	109 (1.200)	123 (1.232)	127 (1.186)	123 (1.209)	110 (1.189)	100 (1.205)
	F0 (Hz)	F	242 (1.096)	219 (1.098)	210 (1.092)	209 (1.088)	211 (1.093)	225 (1.098)	252 (1.087)
		M	137 (1.199)	131 (1.186)	124 (1.183)	122 (1.199)	122 (1.178)	132 (1.194)	140 (1.223)
	F1 (Hz)	F	307 (1.198)	425 (1.082)	646 (1.076)	910 (1.078)	681 (1.087)	442 (1.094)	337 (1.192)
		M	285 (1.077)	357 (1.077)	518 (1.089)	683 (1.095)	532 (1.160)	372 (1.100)	310 (1.070)
	F2 (Hz)	F	2676 (1.056)	2468 (1.061)	2271 (1.051)	1627 (1.062)	1054 (1.099)	893 (1.054)	812 (1.054)
		M	2198 (1.078)	2028 (1.076)	1831 (1.072)	1329 (1.088)	927 (1.108)	804 (1.092)	761 (1.100)
	F3 (Hz)	F	3296 (1.073)	3074 (1.048)	2897 (1.077)	2625 (1.119)	2653 (1.114)	2627 (1.158)	2691 (1.123)
		M	2952 (1.066)	2719 (1.077)	2572 (1.050)	2324 (1.084)	2335 (1.069)	2380 (1.060)	2309 (1.078)
	Ceiling (Hz)	F	6001 (1.086)	5933 (1.094)	5463 (1.166)	5577 (1.076)	5260 (1.137)	4938 (1.113)	5090 (1.095)
		M	5230 (1.155)	5063 (1.181)	5010 (1.137)	4463 (1.105)	4436 (1.077)	4522 (1.068)	4458 (1.064)
EP	Duration (ms)	F	92 (1.154)	106 (1.151)	115 (1.137)	122 (1.144)	118 (1.141)	110 (1.158)	94 (1.208)
		M	84 (1.142)	97 (1.147)	106 (1.162)	108 (1.183)	104 (1.149)	99 (1.144)	83 (1.151)
	F0 (Hz)	F	216 (1.084)	211 (1.082)	204 (1.075)	201 (1.086)	204 (1.076)	211 (1.084)	222 (1.092)
		M	126 (1.177)	122 (1.165)	117 (1.156)	115 (1.151)	117 (1.151)	123 (1.171)	127 (1.187)
	F1 (Hz)	F	313 (1.243)	402 (1.125)	511 (1.154)	781 (1.186)	592 (1.270)	422 (1.150)	335 (1.230)
		M	284 (1.085)	355 (1.090)	455 (1.131)	661 (1.075)	491 (1.111)	363 (1.107)	303 (1.085)
	F2 (Hz)	F	2760 (1.033)	2508 (1.040)	2360 (1.031)	1662 (1.078)	1118 (1.091)	921 (1.184)	862 (1.144)
		M	2161 (1.048)	1987 (1.058)	1836 (1.068)	1365 (1.060)	934 (1.078)	843 (1.090)	814 (1.127)
	F3 (Hz)	F	3283 (1.054)	3007 (1.043)	2943 (1.042)	2535 (1.170)	2729 (1.086)	2636 (1.188)	2458 (1.204)
		M	2774 (1.057)	2559 (1.057)	2475 (1.049)	2333 (1.041)	2414 (1.077)	2429 (1.072)	2315 (1.041)
	Ceiling (Hz)	F	5875 (1.090)	5734 (1.087)	5662 (1.096)	5278 (1.085)	5259 (1.132)	5165 (1.123)	5066 (1.119)
		M	4570 (1.153)	4733 (1.148)	4792 (1.098)	4523 (1.120)	4537 (1.137)	4512 (1.108)	4366 (1.065)

the specific research questions mentioned in the Introduction and finding differences between the two dialects and between the two genders. These sections report the effects of vowel category, gender and dialect on formants, duration, and fundamental frequency. Table I summarizes the average values for all these quantities (also shown in Figs. 6–8); each number in the table is a geometric average over ten speaker values, each of which is a median over 20 tokens (2 phrasal

positions × 2 word-final vowels × 5 consonant environments, see Sec. II B; using the median minimizes the influence of occasional measurement errors). Following much existing cross-dialectal work (Hagiwara, 1997; Adank *et al.*, 2004; Clopper *et al.*, 2005), the table has been split not only for dialect but also for gender, because males may speak differently as a group from females, and sound change (which is a likely source of any difference between BP and EP) may proceed with a different speed for males than for females (Labov, 1994, p. 156).

Since duration, F0, and formants are by definition positive quantities, they are expected to be normally distributed along logarithmic scales, and all statistical investigations in this and the following sections are therefore performed on log-transformed values; this decision is also inspired by the fact that duration is perceived and represented logarithmically (Gibbon, 1977; Allan and Gibbon, 1991), that F0 ranges are comparable for men and women only along a logarithmic scale (Henton, 1989; Tielen, 1992), and that the influence of a specific articulation on the height of formants (in hertz) must be expressed as a *ratio* (rather than as a difference) that is independent of the vocal tract size (if the vocal tract shape is constant). For readability, all averages of logarithmic values are transformed back to milliseconds or hertz, so that the reported averages are in effect geometric averages over the original values in milliseconds or hertz, as in Table I. Also, observed differences between groups in the log domain are reported as ratios between groups, and an observed reliable difference between groups in the log domain is reported as a (duration, F0, F1, or F2) ratio between groups that is reliably different from 1. Another consequence is that all figures use logarithmic axes. In Table I, the standard deviations in the log domain are expressed as ratios in the milliseconds or hertz domains; for example, if a certain average is 400 Hz and the corresponding standard deviation is 1.100, then one standard deviation up from the average is 440 Hz, two standard deviations up is 484 Hz, and one standard deviation down is 363.636 Hz.

Table I does not express what kind of variation the seven standard deviations in a row are due to; do the standard deviations of F0, for instance, reflect the fact that every speaker comes with a different small pitch range, or do they reflect the fact that every speaker randomly determines which vowel has what F0? To thus separate main speaker effects from speaker-vowel interaction effects, and to evaluate the differences between the dialects and between the genders, each of the statistical investigations into duration, F0, F1, and F2 (Secs. IV B, IV F, V, and VI) starts out with an exploratory repeated-measures analysis of variance (conducted with SPSS) on 280 logarithmic values (40 speakers × 7 vowels), which are the median values of the 20 tokens of each of the 7 vowels produced by the 40 speakers. In every repeated-measures analysis, dialect and gender act as between-subjects factors and vowel category acts as a within-subjects factor. For all four acoustical dimensions, Mauchly's sphericity test suggests that the numbers of degrees of freedom for the vowel effects have to be reduced. Accordingly, we decided to use Huynh–Feldt's correction, which multiplies the number of degrees of freedom (6 for the numerator, 216 for the denominator in the *F*-test) by a factor ε, which tends to be around 0.5. After each exploratory analysis we perform tests that directly address a specific research question raised in the Introduction, by investigating the behavior of a within-speaker measure specifically designed for the purpose.

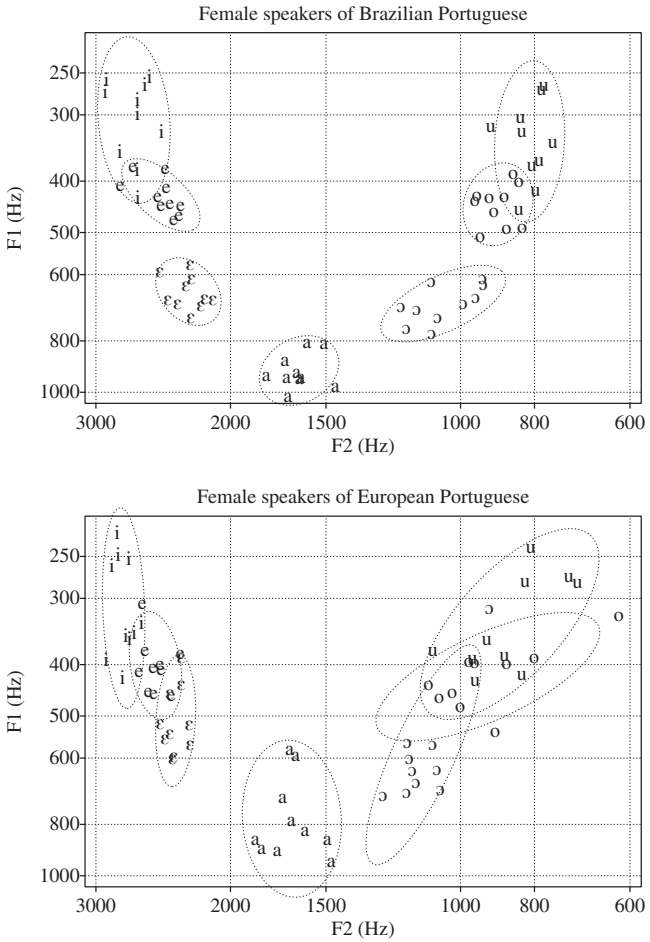


FIG. 4. First and second formants of ten BP and ten EP women.

IV. RESULTS FOR FORMANTS

A. The speakers' median formants

Figures 4 and 5 show the median F1 and F2 values for the ten female and ten male speakers of each dialect. In each of the four figures, each vowel occurs ten times because there were ten speakers of that gender and dialect. Each vowel symbol's vertical position represents the median of the speaker's 20 F1 values, and its horizontal position represents the median of the speaker's 20 F2 values. The 20 F1-F2 pairs that lie behind each vowel symbol were all measured with the same formant ceiling, namely, the formant ceiling that minimizes the variation among the 20 F1 and F2 values (Sec. II E).

Figure 6 shows the mean F1 and F2 values for the seven vowels for the four groups. Each symbol represents a geometric mean of ten speakers' median F1 and F2 values. The following sections consider F1 and F2 separately.

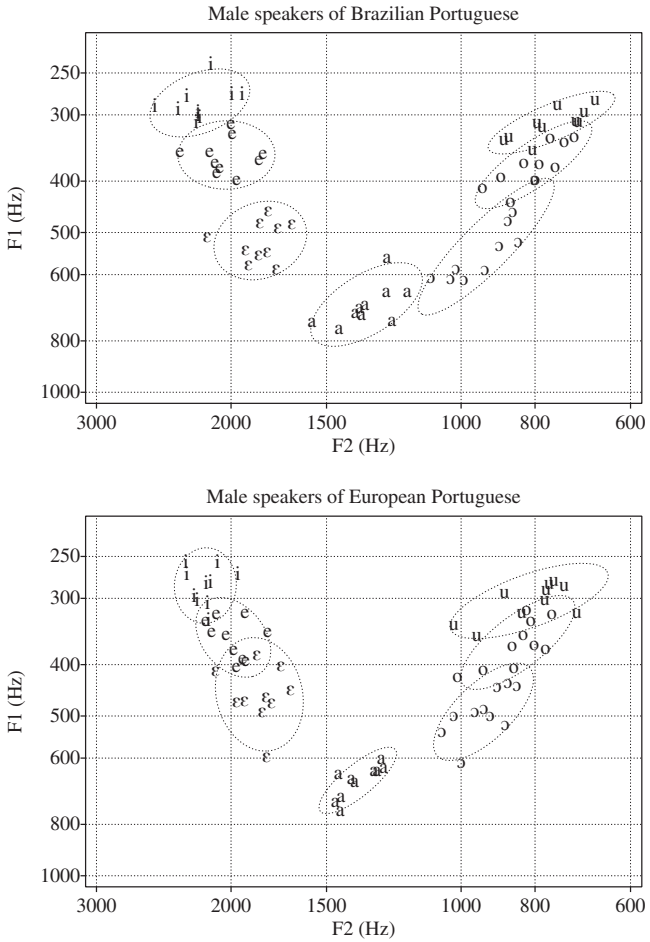


FIG. 5. First and second formants of ten BP and ten EP men.

B. Exploratory analysis of F1

The exploratory repeated-measures analysis of variance reveals a large main effect of vowel category on F1 (η_p^2 =0.950; $F[6\varepsilon,216\varepsilon,\varepsilon=0.609]=684.926$; $p=9\times10^{-85}$). As expected from the Introduction, and clearly visible in Fig. 6, the main determiner of F1 is the phonological vowel height: coarsely speaking, the low vowel /a/ has the highest F1, followed by the lower-mid vowels /ɛ/ and /ɔ/, then the higher-mid vowels /e/ and /o/, and finally the high vowels /i/ and /u/ which have the lowest F1. A subtler effect (of vowel place) is investigated in Sec. IV C.

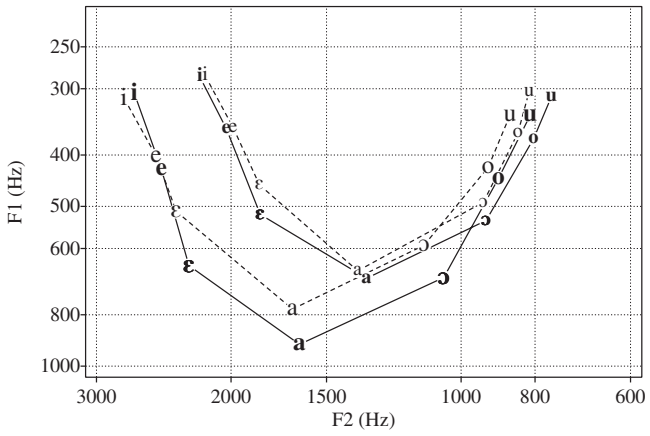


FIG. 6. The vowel spaces of the four groups. Solid lines and bold symbols=BP; dashed lines=EP. Large font: women; small font: men.

As expected, the analysis also reveals a large main effect of gender on F1 (η_p^2=0.394; $F[1,36]=23.430$; $p=2.4\times10^{-5}$): Portuguese-speaking women tend to have higher F1 values (geometric average: 478 Hz; the 95% confidence interval runs from 456 to 501 Hz) than Portuguese-speaking men (409 Hz; c.i.=390–429 Hz). The gender effect on F1 is therefore a ratio of 1.170 (c.i.=1.095–1.249), which compares well (as it should) with the female-male ratio of 1.186 found for the optimal formant ceilings in Sec. II E.

It is possible that the gender effect on F1 may have to be viewed in relation to interaction effects. Since the interaction of gender and dialect is not reliably different from zero ($F[1,36]=0.492$; $p=0.488$), and neither is the triple interaction of gender, dialect, and vowel ($F[6\varepsilon,216\varepsilon,\varepsilon=0.609]$ =1.219; $p=0.306$), it remains to consider the interaction of gender and vowel, which is indeed reliable (η_p^2=0.113; $F[6\varepsilon,216\varepsilon,\varepsilon=0.609]=4.604$; $p=0.0023$). Figure 6 suggests that this is because women take up a greater part of the F1 continuum than men. This is investigated in detail in Sec. IV D.

Finally, the analysis reveals a nearly significant main effect of dialect on F1 ($F[1,36]=4.052$; $p=0.052$), but the cause of this is probably the reliable interaction effect of dialect and vowel on F1 (η_p^2=0.158; $F[6\varepsilon,216\varepsilon,\varepsilon=0.609]$ =6.777; $p=9.5\times10^{-5}$). Apparently, some vowels have different heights in (São Paulo) BP than in (Lisbon) EP. This is investigated in detail in Sec. IV E.

C. The effect of vowel place on F1

One of the research questions in the Introduction is whether Portuguese follows the cross-linguistic trend that (rounded) back vowels tend to have higher F1 values than the corresponding (unrounded) front vowels. Figure 6 does show that for all four groups of speakers each back vowel has a higher average F1 than its front counterpart, but the figure does not show that this can be generalized to the Portuguese-speaking population. The exploratory analysis of Sec. IV A does yield an answer by reporting within-subjects comparisons. That is, a speaker's F1 of /u/ is higher than that of his or her /i/ by a factor of 1.082, the F1 of /o/ is higher than that of /e/ by a factor of 1.039, and the F1 of /ɔ/ is higher than that of /ɛ/ by a factor of 1.078. All three factors are reliably greater than 1 (uncorrected two-tailed $p=9.1\times10^{-12}$, 5.6×10^{-5}, and 7.1×10^{-5}, respectively): their 98.30% confidence intervals (i.e., Šidák-corrected for three planned comparisons) are 1.060–1.103, 1.017–1.061, and 1.034–1.125, respectively. The conclusion is that in the Portuguese-speaking population, each back vowel has a higher mean F1 than its corresponding front vowel. A multivariate analysis of variance with dialect and gender as factors and the three front-back differences as dependents reveals no influence of dialect, gender, or dialect × gender on the front-back differences.

Simple sign counting reveals that this correlation between F1 and backness holds for a majority of individual

speakers: for 38 of the 40 speakers, the F1 of /u/ is higher than the F1 of the same speaker's /i/. Likewise, the /o/-/e/ difference is positive for 32 of the 40 speakers, and the /ɔ/-/ɛ/ difference for 35 of the 40 speakers (the 15 exceptions happen to be maximally evenly distributed over the four groups, and maximally randomly distributed over the speakers). By not labeling the vowel symbols for speaker, Figs. 4 and 5 obscure this consistent effect (for instance, the four EP speakers with the conspicuously low F1 values for /i/ in Fig. 4 are the same as those with the conspicuously low F1 values for /u/). Sign counting therefore confirms again that there is a consistent correlation between F1 and phonological backness.

D. The effect of gender and dialect on the size of the F1 space

One of the research questions in the Introduction is whether Portuguese-speaking females have larger vowel spaces (along logarithmic axes) than Portuguese-speaking males. To answer this, we define a speaker's *F1 space size* as the ratio of the F1 of his or her low vowel /a/ and the (geometric) average F1 of his or her high vowels /i/ and /u/. We thus compute 40 F1 space sizes and subject these to a two-way analysis of variance with dialect and gender as factors. Since an interaction between gender and dialect was not found ($F[1,36]=2.395$, $p=0.130$), we report here only the two main effects.

The average F1 space size of the 20 women turns out to be 2.613, and that of the 20 men only 2.276. The female F1 space is therefore $2.613/2.276=1.148$ times (0.199 octaves) larger than the male F1 space (c.i.=1.046–1.260; the ratio is reliably different from 1 with $F[1,36]=9.052$, $p=0.0048$). As suggested at the end of Sec. IV B, therefore, Portuguese-speaking women indeed take up a larger part of the F1 space than men. For a comparison with other languages see Sec. VII A.

The F1 space size may also depend on the dialect. The average F1 space size of the 20 Brazilians is 2.552, and that of the Europeans 2.331. For the combined population of men and women, the Brazilian F1 space is therefore 1.095 times larger than the European F1 space (c.i.=0.998–1.201). This is not very reliably different from 1 ($F[1,36]=3.895$, $p=0.056$).

E. Vowel height differences between the two dialects

One of the research questions in the Introduction is which vowels are different in the two dialects. We first investigate this by a multivariate analysis of variance on the seven F1 values, with dialect and gender as factors. Since the dialect-gender interaction is not significant (Wilks' $\Lambda[7,30]=0.837$, $p=0.566$), we focus on the main effect of dialect. The vowel /ɛ/ turns out to be very reliably lower (higher F1) in BP than in EP ($F[1,36]=27.468$, $p=7.1\times10^{-6}$). A difference in the same direction is found for its back counterpart /ɔ/ ($F[1,36]=4.973$, $p=0.032$) and for the vowel /a/ ($F[1,36]=7.162$, $p=0.011$), although these differences are not very reliable (regarding the multiple comparisons). The hypothesis by Moraes *et al.* (1996) mentioned in the Introduction is not confirmed: for the 40 speakers, /u/ has indeed a higher F1 in BP than in EP (ratio 1.013), but /i/ has a lower F1 in BP than in EP (ratio 0.992); neither of these ratios generalize reliably to the populations (they are different from 1 with $p=0.779$ and 0.866); in fact, the upper bounds of the confidence intervals (0.923–1.112 and 0.900–1.093) show that the extent of any lowering of the high vowels cannot be greater than 11.2%.

From the mere fact that we found that /ɛ/ is lower in BP than in EP whereas we found no difference for /e/, we cannot yet conclude that in BP /ɛ/ is lowered more than /e/ (from differences in p values no inferences can be made about the relative sizes of an effect), and we cannot therefore answer yet our research question about the difference between the /ɛ/-/e/ distances in BP and EP. Both of these problems are addressed in the remainder of this section.

In order to establish any dialectal difference in /ɛ/-/e/ distance, one can take advantage of the fact that all seven vowels have been spoken by the same 40 speakers, i.e., we have information about the internal structure of each speaker's vowel space. Thus, the log(F1) differences between every speaker's /ɛ/ and /e/ were computed, as well as those between every speaker's /ɔ/ and /o/. A multivariate analysis of variance with dialect and gender as factors was performed on the two sets of 40 values. The only significant effect is that of dialect ($\Lambda[2,35]=0.451$, $p=8.8\times10^{-7}$), and it turns out that the F1 ratio of /ɛ/ and /e/ is very reliably greater in BP (observed average 1.485; uncorrected 95% c.i. =1.437–1.535) than in EP (1.276; c.i.=1.235–1.319): the ratio of these ratios is $1.485/1.276=1.164$ (c.i. =1.111–1.219), which is reliably different from 1 ($F[1,36]=43.391$, $p=1.1\times10^{-7}$). Likewise, the F1 ratio of /ɔ/ and /o/ is greater for the 20 Brazilians (1.482; c.i.=1.409–1.559) than for the 20 Europeans (1.377; c.i.=1.309–1.449); the ratio of these ratios is 1.076 (c.i.=1.002–1.156), which is reliably different from 1 at the $\alpha=0.05$ level ($F[1,36]=4.326$, $p=0.045$). We conclude that the acoustic distance between lower-mid and higher-mid vowels is indeed larger in BP than in EP.

We subsequently address the other question, namely, what is behind these observed differences in the acoustic mid-vowel distances: are these differences due to /ɛ/ and /ɔ/ being lower in BP than in EP or due to /e/ and /o/ being higher in BP than in EP? Table I and Fig. 6 indicate that the latter possibility is unlikely: for both women and men, the mean BP /e/ and /o/ are *lower* than the mean EP /e/ and /o/. The next hypothesis to consider is that the relative openness of the lower-mid vowels in BP is due to the larger F1 space that BP speakers may be using (Sec. IV D). In that case, the lowness of /ɛ/ and /ɔ/ should disappear if the F1 values are normalized for the F1 space size. To assess whether this is the case, we compute the *relative heights* of the four mid vowels for each speaker. For instance, the relative height of /ɛ/ within the front vowel space can be defined as $(\log F1(a)-\log F1(\varepsilon))/(\log F1(a)-\log F1(i))$, and the relative height of /o/ within the back vowel space can be defined as $(\log F1(a)-\log F1(o))/(\log F1(a)-\log F1(u))$.

A multivariate (four vowels) two-way (dialect, gender) analysis of variance reveals no effect of gender on relative

height ($\Lambda[4,33]=0.883$, $p=0.376$) and no interaction of dialect and gender ($\Lambda[4,33]=0.961$, $p=0.855$). We therefore only report on the main effect of dialect ($\Lambda[4,33]=0.423$, $p=1.0\times10^{-5}$). If all vowels were equally spaced along the log(F1) dimension, the lower-mid vowels would have a relative height of 0.333. The average Brazilian /ɛ/ indeed has a relative height of 0.315 (c.i.=0.275–0.355), but the average EP /ɛ/ has a relative height of 0.455 (c.i.=0.415–0.496), i.e., it lies close to the center of the F1 dimension; the difference between the dialects is highly reliable ($F[1,36]=25.022$; $p=3.0\times10^{-5}$). For /ɔ/, the difference between BP and EP is in the same direction (0.303 versus 0.353), but is not significant ($F[1,36]=1.250$; $p=0.271$). The higher-mid vowels seem to have very similar relative heights in the two dialects: /e/ has 0.730 for BP and 0.737 for EP, and /o/ has 0.752 for BP and 0.748 for EP. We conclude that the lower BP /ɛ/ remains even after normalizing for BP's larger F1 space.

The results of the previous paragraph suggest that the cause of the smaller /ɛ/-/e/ distance in EP could lie in a lower F1 for /ɛ/, but to be absolutely statistically certain (again, different degrees of statistical significance do not entail different effect sizes) one has to investigate whether the dialectal difference in the relative height of /ɛ/ is greater than that of /e/. This can be determined by subjecting the 40 *average mid vowel heights*, namely, $(\log F1(a)-(\log F1(\varepsilon)+\log F1(e))/2)/(\log F1(a)-\log F1(i))$, to a two-way analysis of variance. The effect of dialect on this measure is indeed significant ($F[1,36]=6.450$; $p=0.016$). We conclude that the smaller /ɛ/-/e/ distance in EP as compared to BP is due more to a raised /ɛ/ than to a lowered /e/ (within a normalized F1 space). For a discussion of the implications see Sec. VII A.

F. Effects on F2

As expected, the repeated-measures analysis of the variance of F2 reveals a large main effect of gender ($F[1,36]=120.857$; $p=4.7\times10^{-13}$): women's F2 values are higher than those of men by an average factor of 1.183, which compares well with the values found for the formant ceiling in Sec. II E and for F1 in Sec. IV B. The EP speakers turn out to have higher F2 values than the BP speakers, but this difference cannot be reliably generalized to their populations ($F[1,36]=3.009$; $p=0.091$). An interaction of dialect and gender is not found ($F[1,36]<1$).

As for the within-subject effects, the analysis reveals the expected main effect of vowel category on F2 ($F[6\varepsilon,216\varepsilon,\varepsilon=0.423]=1826.704$; $p=1.6\times10^{-78}$), as well as a reliable interaction between vowel and gender ($F[6\varepsilon,216\varepsilon,\varepsilon=0.423]=9.339$; $p=5.5\times10^{-5}$). From Fig. 6, the cause of the latter appears to be that the size of the F2 space (the /u/-/i/ distance) is larger for females than for males; this is investigated in detail below. The analysis reveals no interaction between vowel and dialect ($F<1$) and no triple interaction between vowel, dialect, and gender ($F<1$).

A multivariate analysis of variance on the F2 values of the seven vowels reveals neither a main effect of dialect[4] nor an effect of the interaction of dialect and gender; the main

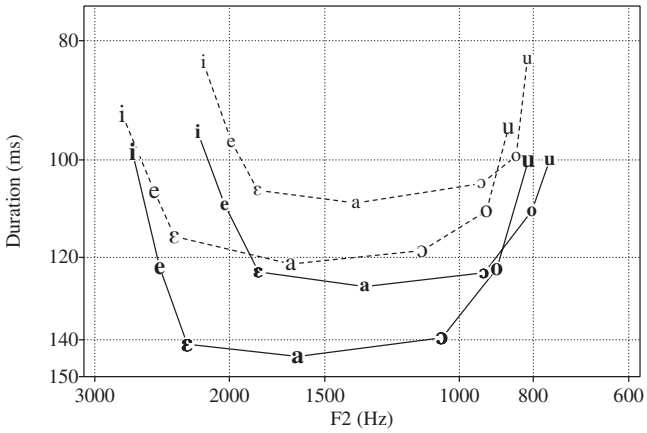


FIG. 7. Mean duration as a function of vowel category. The purpose of the inclusion of the F2 axis and the reversal of the vertical axis is to provide vowel space shapes that are similar in orientation and extent as the more usual ones in Fig. 6. Solid lines and bold symbols=BP; dashed lines=EP. Large font: women; small font: men.

effect of gender ($\Lambda[7,30]=0.143$, $p=5.0\times10^{-11}$) is that /a,ɛ,e,i,ɔ/ have a very reliably higher F2 for women than for men ($F[1,36]\geqslant28.953$, $p\leqslant4.7\times10^{-6}$); for /u/ ($F[1,36]=3.329$; $p=0.076$) and /o/ ($F[1,36]=8.125$; $p=0.0072$), the observed average effect is in the same direction but in itself less reliably generalizable to the population (given the multiplicity of the tests). The hypothesis that all vowels simultaneously have a higher F2 for women than for men is nevertheless confirmed at the $\alpha=0.10$ level (in the case of such an inclusive hypothesis, the multiplicity of tests also raises the chance of a type II error, so that one is allowed to use a higher α than usual: Winer, 1962, p. 13).

Analogously to the F1 space size of Sec. IV C, we define a speaker's *F2 space size* as the ratio of the F2 of his or her /i/ and the F2 of his or her /u/. When we subject the 40 sizes to a two-way analysis of variance, we find no effect of dialect ($F[1,36]=2.076$, $p=0.158$) or of dialect×gender ($F<1$), and the main effect of gender ($F[1,36]=16.504$, $p=2.5\times10^{-4}$) is that for the 20 men, the average ratio is 2.768 (c.i.=2.616–2.929), and for the 20 women it is 3.249 (c.i.=3.070–3.437); the ratio of these ratios is 1.174 (c.i.=1.083–1.271). We conclude that the size of the F2 space is greater for Portuguese-speaking women than for men, i.e., that the gender difference in F2 is larger for /i/ than for /u/.

V. RESULTS FOR DURATION

The fact, mentioned in the Introduction, that the Portuguese vowel system does not use vowel length as a phonological feature does not preclude that different vowels may have quite different phonetic durations, and that vowel durations may differ between dialects and between genders. Figure 7 shows the dependence of duration on vowel, dialect, and gender. Each symbol represents a value of duration (and F2) averaged over the median duration (and F2) values of ten speakers.

A. Exploratory analyses

The repeated-measures analysis of the variance of duration reveals that the main effect of vowel category is very

reliable ($F[6\varepsilon,216\varepsilon,\varepsilon=0.811]=243.358$, $p=5\times10^{-76}$); this issue is investigated in detail in Sec. V B. The duration of the vowels is influenced by dialect ($\eta_p^2=0.180$; $F[1,36]=7.915$, $p=0.008$): vowels are longer in BP than in EP by a factor of 1.148 (c.i.=1.039–1.269); this is investigated further in Sec. V C. The expected main effect of gender (see Introduction) is barely significant ($\eta_p^2=0.103$; $F[1,36]=4.125$, $p=0.050$): women's vowels are longer than men's vowels by a ratio of 1.105 (c.i.=1.0001–1.221); this is discussed in Sec. V C as well. The analysis does not reveal an interaction between gender and dialect ($F<1$), i.e., the difference between the two solid curves in Fig. 7 is not reliably different from the difference between the two dashed curves. The two-way interactions between gender and vowel and between dialect and vowel, and the three-way interaction between gender, dialect, and vowel are reliable, at least under the somewhat forgiving Huynh–Feldt correction ($F[6\varepsilon,216\varepsilon,\varepsilon=0.811]=2.426$, 3.829, 3.671; $p=0.039$, 0.0028, 0.0038); Fig. 7 suggests, for instance, that specifically /u/ is shortened specifically by EP men.

A multivariate analysis of variance on all vowel durations shows that at the $\alpha=0.10$ level, all seven vowels are longer in BP than in EP (/a,ɛ,ɔ/: $F[1,36]\geqslant10.770$, $p\leqslant0.0023$; /e/: $F=6.480$, $p=0.015$; /u/: $F=5.020$, $p=0.031$; /o/: $F=4.981$, $p=0.032$; /i/: $F=3.648$, $p=0.064$).

B. Vowel-intrinsic duration

From the Introduction, one can expect an effect of vowel height on duration, and Fig. 7 confirms this expectation. In fact, for 39 of the 40 speakers, the median of his or her 20 measured /i/ tokens is shorter than the median of his or her 20 measured /e/ tokens. Within the analysis of Sec. V A, pairwise comparisons between the seven vowels yield the following results for vowels of adjacent phonological heights: /i,u/ are shorter than /e,o/ (all four uncorrected two-tailed $p<3\times10^{-13}$), /e,o/ shorter than /ɛ,ɔ/ (all four $p<2\times10^{-10}$), /ɛ/ shorter than /a/ ($p=0.0072$), and /o/ shorter than /a/ ($p=0.000\,34$). We conclude with confidence that lower vowels are longer than higher vowels in Portuguese.

Given the structure of the phonological vowel space, a second potential effect may be worth investigating, namely, whether duration depends on the front-back distinction. The result of the three relevant pairwise comparisons is that /i/ is shorter than /u/ ($p=0.036$) and /e/ is shorter than /o/ ($p=0.029$); the difference between /ɛ/ and /ɔ/ is not significant ($p=0.940$). This subject is not pursued further here (a possible explanation is given in Sec. VII C), and the focus below is solely on the traditional vowel-intrinsic duration effect, which is the relation between duration and height.

To investigate the size (rather than just the existence) of the vowel-intrinsic duration effect (for cross-linguistic comparison), we define for each speaker the *vowel-intrinsic duration ratio* as the ratio between the duration of his or her /a/ and the average duration of his or her /i/ and /u/. We subject the 40 values thus obtained to a two-way analysis of variance. The average vowel-intrinsic duration ratio of the 40 speakers is 1.339 (c.i.=1.304–1.374). The ratio is comparably slightly influenced by dialect ($\eta_p^2=0.100$; $F[1,36]=3.988$, $p=0.053$), gender ($\eta_p^2=0.118$; $F[1,36]=4.794$, $p=0.035$), and an interaction of dialect and gender ($\eta_p^2=0.110$; $F[1,36]=4.454$, $p=0.042$); a one-way analysis of variance with the four speaker groups as the levels of the single factor confirms that the BP females have a larger vowel-intrinsic duration ratio than any of the other three groups (Tukey's "honestly significant difference" *post hoc* test: all three $p\leqslant0.030$), which do not differ significantly among themselves (all three $p\geqslant0.999$). Comparisons with other languages, and their implications, are discussed in Sec. VII C.

C. Dialect and gender differences in duration: Results of speaking rate?

The observed differences in vowel duration between the groups might potentially arise from between-group differences in speaking rate. To investigate whether such differences exist, we perform three between-group analyses of speaking rate.

For the first analysis we measured the durations of the utterance parts "em susse e susso," "em sasse e sasso," and so on, for all seven vowels but only for the consonant /s/; averaging over the seven vowels yields one typical sentence duration per speaker. When we subject the 40 values to a two-way analysis of variance, we find no reliable effect of dialect, gender, or dialect × gender (all three $p\geqslant0.142$). Hence, no difference in speaking rate is detected here.

For the second analysis we measured the durations of the /s/ *before* the target vowel, i.e., the initial consonants "s" of the words "susse," "sasse," and so on, for all seven vowels; averaging over the seven vowels yields one typical initial /s/ duration per speaker. A two-way analysis of variance again finds no reliable effect of dialect, gender, or dialect × gender (all three $p\geqslant0.219$). So again no difference is found between the dialects.

For the third analysis we measured the durations of the /s/ *after* the target vowel, i.e., the medial consonants "ss" of the words susse, sasse, and so on, for all seven vowels; averaging over the seven vowels yields one typical medial /s/ duration per speaker. A two-way analysis of variance reveals an effect of dialect alone ($p=0.012$; the other two $p\geqslant0.205$): the postvocalic /s/ is *shorter* in BP than in EP, opposite to the difference in vowel durations. Hence, it looks as if the Brazilians compensate for their longer stressed vowels by shortening the following consonant. This suggests that the duration difference in the stressed vowels is not caused by a difference in speech rate between the dialects.

VI. RESULTS FOR FUNDAMENTAL FREQUENCY

The fact, mentioned in the Introduction, that the Portuguese vowel system does not use tone as a phonological feature does not preclude that different vowels may have quite different fundamental frequencies, and that fundamental frequencies may differ between dialects (as they are expected to do between genders). Figure 8 shows the dependence of F0 on vowel, dialect, and gender. Each symbol represents a value of F0 (and F2) averaged over the median F0 (and F2) values of ten speakers.

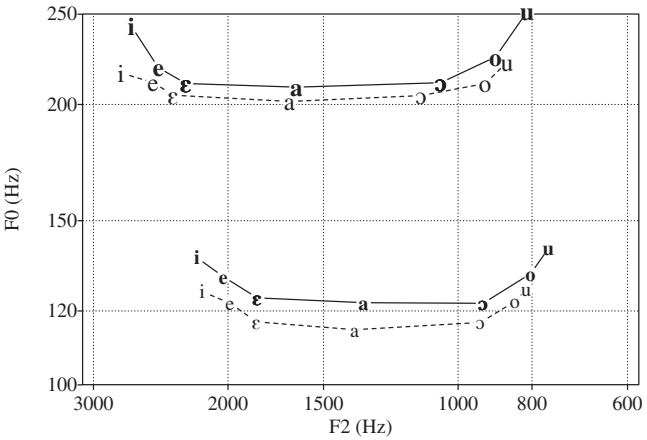


FIG. 8. Mean F0 as a function of vowel category. Solid lines and bold symbols=BP; dashed lines=EP. Top: women; bottom: men.

A. Exploratory analysis

The exploratory analysis of variance of F0 finds the expected large main effect of gender (η_p^2=0.833; $F[1,36]$ =179.793, $p=1.4\times10^{-15}$): the 20 women have a (geometric) average F0 of 216.60 Hz, the 20 men one of 125.07 Hz; the F0 of Portuguese-speaking women is therefore a factor of 1.732 higher than that of Portuguese-speaking men (c.i. =1.567–1.913). We find no reliable main effect of dialect ($F[1,36]$=0.007, p=0.932). Within speakers we find a main effect of vowel category ($F[6\varepsilon,216\varepsilon,\varepsilon$=0.492]=136.121, $p=5.3\times10^{-36}$) and an interaction of vowel and dialect (F =11.224, $p=2.1\times10^{-6}$), both of which can be observed in Fig. 8 and are discussed in Sec. VI B. We find no reliable interaction of vowel and gender (F=2.499; p=0.064) or triple interaction of vowel, gender, and dialect (F=2.276; p =0.085).

B. Vowel-intrinsic F0

From the Introduction, one can expect an effect of vowel height on F0, and Fig. 8 confirms this expectation. In fact, for all 40 speakers, both /i/ and /u/ have a higher F0 than /a/. Within the analysis of Sec. VI A, pairwise comparisons between the seven vowels yield the following results for vowels of adjacent phonological heights: /i,u/ have a higher F0 than /e,o/ (all four $p<2\times10^{-9}$), /e,o/ higher than /ɛ,ɔ/ (all four $p<4\times10^{-11}$), and /ɛ,ɔ/ higher than /a/ (p=0.000 55 and 0.0040). We conclude with confidence that lower vowels have a lower F0 than higher vowels in Portuguese. The fundamental frequency also seems to depend on place: /u/ has a higher F0 than /i/ (p=0.000 22) and /o/ than /e/ (p=0.049); the difference between /ɔ/ and /ɛ/ is less than one standard error (and in the wrong direction; p=0.334).

To investigate the size of the vowel-intrinsic F0 effect, we define for each speaker the *vowel-intrinsic F0 ratio* as the ratio between the average F0 of the high vowels /i/ and /u/ and the F0 of the low vowel /a/. When we subject the 40 values thus obtained to a two-way analysis of variance, we find a reliable main effect of dialect ($F[1,36]$=12.301, p =0.0012): the average ratios are 1.158 for the 20 Brazilians and 1.095 for the 20 Europeans. The ratio is therefore greater for BP than for EP by a factor of 1.057 (c.i.=1.024–1.092; p=0.000 62). Neither a main effect of gender ($F[1,36]$ =0.987, p=0.327) nor an interaction between gender and dialect ($F[1,36]$=4.454, p=0.079) is reliably detected.

VII. DISCUSSION

This section compares the results of Secs. IV–VI to earlier findings in the literature and tries to find explanations for the phenomena observed. Universal aspects, Portuguese-specific aspects, and dialect-specific aspects are identified.

A. First formant: Universal, Portuguese-specific, dialect-specific

Section IV B has found that the four-way phonological vowel height contrast of Portuguese is a strong determiner of F1. That is, the seven vowels divide up into four F1 regions, where each back vowel has an F1 similar to its corresponding front vowel. This is an unsurprising observation given the phonological discussions in the Introduction and given the fact that most languages with large vowel inventories exhibit this kind of symmetry. Section IV B has also found that women tend to have higher F1 values than men. This is an unsurprising observation reported abundantly in the previous literature (e.g., Peterson and Barney, 1952), and well understood in terms of the differences in vocal tract length between women and men. The gender effect on F1 is a ratio of 1.170. Section IV C finds that back vowels consistently have slightly higher F1 values than their front counterparts. We speculate that a universal principle might be involved, because this effect has been found for several languages with large vowel inventories (mentioned in the Introduction), and even for five-vowel inventories the relation still seems to apply to the /i/-/u/ contrast: Iberian Spanish (the control subjects of Cervera *et al.*, 2001), Japanese (Nishi *et al.*, 2008), Czech (Chládková *et al.*, 2009), and Hebrew (Most *et al.*, 2000).

According to Sec. IV D, the BP F1 space size is 1.201 times larger for females than for males, and for the EP speakers this *female-to-male F1 space size ratio* is 1.097. In order to assess the universality of these gender differences, one can compare these ratios to those of other languages. It is difficult to compare F1 values between studies because of the different data collection methods (speaking rate, speaking style) and different formant analysis methods (formant ceilings, number of formants measured, pre-emphasis). One can hope, however, that most of these issues have little influence on the female-male F1 ratio that one can extract from any specific study. For the American English speakers of Peterson and Barney (1952), then, the ratio is 0.978. For the American English speakers of Hillenbrand *et al.* (1995), the ratio is also 0.978. This suggests that American English women have a vowel space that may be shifted with respect to that of American English men, but is not larger (along a logarithmic scale). For the Northern Standard Dutch speakers of Adank *et al.* (2004), the ratio is 1.260, and for the Southern Standard Dutch speakers in that study the ratio is 1.032. Apparently, there can be large differences between languages and even closely related varieties in this respect.

Both Portuguese values happen to fall in between the two Dutch ones.

The combined evidence of Sec. IV E leads to the conclusion that /ɛ/ is higher (less open, having a lower absolute and relative F1) in EP from Lisbon than in BP from São Paulo. None of the studies on Portuguese vowels mentioned in the Introduction reported this dialectal difference. Regarding the ideas in the Introduction, and the location of /ɛ/ near the center of the F1 continuum, we might well be watching an impending merger (in EP) of /ɛ/ into /e/, as is also happening in Italian, French, and Catalan (see Introduction).

B. Second formant: Universal, Portuguese-specific, dialect-specific

Section IV F makes four observations. First, phonological front- and backness is a strong determiner of F2 in Portuguese. This is an unsurprising observation given that Portuguese, as most languages, uses vowel place to distinguish between vowel categories. Second, women have higher F2 values than men. As with F1, the well-understood explanation lies in the differences between the vocal tract sizes (the gender effect on F2 is a ratio of 1.183, which is comparable to the effect on F1). Third, /u/ might be more fronted in EP than in BP.[4] This could have been seen by comparing earlier publications on BP (Callou *et al.*, 1996) and EP (Delgado-Martins, 1973).

Fourth, Portuguese-speaking women not only have larger F1 space sizes than men, they also have larger F2 space sizes. The average Portuguese *female-to-male F2 space size ratio* is 1.174. For the American English speakers of Peterson and Barney (1952), the ratio is 1.116; for those of Hillenbrand *et al.* (1995), it is 1.089. For the Northern Dutch speakers of Adank *et al.* (2004), the ratio is 1.002, for the Southerners it is 1.166 (when compared with the F1 case, it is now the opposite group that exhibits large gender differences). The Portuguese ratio seems to be larger than that of English and Dutch. However, the large confidence interval reported in Sec. IV F, together with the presumably equally large uncertainties in the values reported for other languages, do not allow firm conclusions to be drawn.

C. Duration: Universal, Portuguese-specific, dialect-specific

Section V identifies four influences on duration in Portuguese. First, vowels are longer for women than for men (Sec. V A). This influence of gender on duration is not specific to Portuguese. Simpson and Ericsdotter (2003) report on many studies which find that female speakers produce longer vowels than male speakers in many Indo-European languages, such as English, German, Jamaican Creoles, French, and Swedish, but also in non-Indo-European languages, such as Creek. This gender effect may have a socio-phonetic origin (Byrd, 1992; Whiteside, 1996), e.g., women tend to speak more clearly than men, or a physiological one, e.g., men tend to have stiffer articulators than women (as speculated by Simpson, 2001, 2002, but not confirmed by Simpson 2003).[5]

Second, vowels are longer in BP than in EP (Sec. V A). A comparable difference has been found in the Spanish-speaking neighbors: Morrison and Escudero (2007) found that Peruvian Spanish vowels (from Lima) were 34% longer than European Spanish vowels (from Madrid). Causation by dialectal differences in speaking rate can probably be ruled out (Sec. V C).

Third, lower vowels are longer than higher vowels (Sec. V B). In Portuguese, this vowel-intrinsic duration effect turns out to be strong: the duration ratio of low and high vowels is 1.339. The effect is stronger than in most other languages without a phonological length contrast, such as Iberian Spanish (the control subjects of Cervera *et al.*, 2001: a ratio of 1.14; Morrison and Escudero, 2007: 1.04), Peruvian Spanish (Morrison and Escudero, 2007: 0.94), or European French (Rochet and Rochet, 1991: a ratio of 1.13; Strange *et al.*, 2007: 1.11). This language-dependence suggests that in Portuguese the effect is not solely of an automatic articulatory nature: it seems that Portuguese has turned duration into a language-specific (minor) cue for phonological vowel identity, analogously to how, e.g., English vowel duration has become a cue for the phonological voicing of a following obstruent, both in production (Heffner, 1937; House and Fairbanks, 1953; Luce and Charles-Luce, 1985) and in perception (Denes, 1955; Raphael, 1972).

Fourth, back vowels might be longer than their front counterparts (Sec. V B). For the high vowels, this was also found by Seara (2000). This effect may be epiphenomenal: back vowels have higher F1's than front vowels (Sec. VII A), and since F1 covaries with duration (see previous paragraph), back vowels are expected to have longer durations than front vowels.

D. Fundamental frequency: Universal, Portuguese-specific, dialect-specific

Section VI identifies three influences on F0. First, the ratio by which Portuguese-speaking women have a higher average F0 than men is 1.732 (Sec. VI A). It can be compared to the ratios of 1.687 and 1.690 found for American English by Peterson and Barney (1952) and Hillenbrand *et al.* (1995), respectively. The data of Adank *et al.* (2004) reveal ratios of 1.497 for Northern Dutch and 1.730 for Southern Dutch; Most *et al.* (2005) report a ratio of 1.518 for Hebrew. All these ratios are much smaller than the ratio found for Japanese (Yamazawa and Hollien, 1992), where the gender difference in F0 is apparently culturally influenced. Since Portuguese joins in with the majority of languages, it can be concluded that the cultural influence of gender on F0 in Portuguese is the same as that in this majority of languages, and might therefore well be zero, so that the effect could just be physiologically determined. However, comparing the gender-dependence of F0 across studies may be less than reliable, because the F0 difference between men and women tends to be largest at the age of our subjects (young adults) and tends to fall at later ages (Baken, 2005).

Second, high vowels have a higher F0 than low vowels, with a ratio of 1.158 for the Brazilians and a reliably smaller ratio of 1.095 for the Europeans (Sec. VI B). This vowel-intrinsic F0 effect is comparable to those reported for Ameri-

can English (House and Fairbanks, 1953: a ratio of 1.092) and Dutch (Koopmans-van Beinum, 1980: 1.098; Adank *et al.*, 2004: 1.222). In Portuguese, the dialect-dependence suggests that the intrinsic F0 is not an automatic consequence of articulation. However, this dependence might be caused by the dialect-dependence of duration, but the literature has never identified a universal negative correlation between F0 and duration (for vowels with a constant F1), so such a cause does not seem likely.

Third, back vowels seem to have a higher F0 than front vowels in Portuguese (Sec. VI B). This was also reliably found for English in a meta-analysis by Whalen and Levitt (1995). No causes for the effect seem to be known.

VIII. CONCLUSION

The present study finds several general properties of Portuguese vowels that they have in common with vowels in many other languages: they exhibit intrinsic F0 (Secs. VI B and VII D) and intrinsic duration (Secs. V B and VII C), the sizes of the F1 and F2 spaces are larger for women than for men (Secs. IV D, IV F, VII A, and VII B), F0 and formant values are higher for females than for males (Secs. IV A, IV F, VI A, VII A, VII B, and VII D), females' vowels are longer than those of males (Secs. V A and VII C), and the structure of the vowel inventory is basically symmetric (Secs. IV B and VII A) although back vowels have slightly higher F1 values than their front counterparts (Secs. IV C and VII A).

A Portuguese-specific finding is that Portuguese speakers seem to have turned vowel duration into a cue for vowel identity, to an extent that goes beyond the automatic lengthening of open vowels (Secs. V B and VII C); just as happened with the voicing-dependent vowel lengthening in English, one can predict that Portuguese *listeners* use this cue to a greater extent than listeners of other languages. Future research will have to verify this prediction.

There are three reliably established dialect-specific findings. One is that BP vowels are longer than EP vowels (Secs. V A, V C, and VII C). Another is that the vowel-intrinsic F0 effect is greater in BP than in EP (Secs. VI B and VII D). The third is that the lower-mid vowel /ɛ/ is higher in EP than in BP, and that it is closer to /e/ in EP than in BP (Secs. V B and VII C), a situation which might signal a future merger. To establish whether we are really witnessing a sound change in progress, a larger investigation with more age groups, social-economic strata, and regional varieties is called for. Such a more comprehensive study could also address some other questions that we had to leave open, such as the possible lowering of high vowels and the degree of articulatory automaticity of the intrinsic duration and intrinsic F0 effects.

At the methodological level, the proposed formant ceiling optimization method found that the average difference of the vocal tract lengths associated with /i/ and /u/ is comparable to the average difference of the female and male vocal tract lengths. Future investigations involving automatic formant measurements could benefit from this observation.

ACKNOWLEDGMENTS

This research was supported by NWO (Netherlands Organization for Scientific Research) Grant No. 016.024.018 to P.B. and by a CAPES (Committee for Postgraduate Courses in Higher Education, Brazilian Ministry of Education) grant to A.S.R. We would like to acknowledge the contribution of Denize Nobre Oliveira on the testing of participants and manual vowel segmentation, and of Ton Wempe for technical support and preliminary analyses.

[1]Some of the authors (Mateus *et al.*, 2005, p. 79) group /ɛ/ and /ɔ/ with /a/ by calling them "low vowels;" there seems to be no reason for this move other than minimizing the number of phonological features.

[2]Adank *et al.* (2004) do not confirm this result for either of the two regional standard varieties of Dutch that they investigate.

[3]A technical detail: the Gaussian-like shape of the window requires tails that capture another 20% of the vowel duration on each side of the central 40%.

[4]One could look specifically into the degree of fronting of /u/, knowing that /u/ was historically fronted (auditorily) in several European languages (dates approximate): 1st-century BC Greek (Sihler, 1995, p. 37), 5th-century Slavic (Stieber, 1979, p. 23), Old Dutch (Schönfeld, 1932, p. 82), 9th-century French (Meyer-Lübke, 1908, p. 53), 15th-century Swedish (Kock, 1911, p. 191), 20th-century southern British English (Harrington *et al.*, 2008). The European speakers indeed have a higher F2 than the Brazilians, but this cannot at this point be reliably generalized to the populations ($F[1,36]=3.676$; $p=0.063$).

[5]If vowel duration is related to speaking rate, identical utterances should be longer when spoken by women than when spoken by men. Whiteside (1996) did find this, but Simpson (2001) did not. Our Portuguese data can neither confirm nor disconfirm such gender differences in speaking rate (Sec. V C).

Adank, P. (**2003**). "Vowel normalization: A perceptual-acoustic study of Dutch vowels," Ph.D. thesis, University of Nijmegen.

Adank, P., Van Hout, R., and Smits, R. (**2004**). "An acoustic description of the vowels of Northern and Southern standard Dutch," J. Acoust. Soc. Am. **116**, 1729–1738.

Allan, L. G., and Gibbon, J. (**1991**). "Human bisection at the geometric mean," Learn Motiv **22**, 39–58.

Anderson, N. (**1978**). "On the calculation of filter coefficients for maximum entropy spectral analysis," in *Modern Spectral Analysis* (IEEE, New York).

Baken, R. J. (**2005**). "The aged voice: A new hypothesis," J. Voice **19**, 317–325.

Barbosa, P. A., and Albano, E. C. (**2004**). "Brazilian Portuguese: Illustrations of the IPA," J. Int. Phonetic Assoc. **34**, 227–232.

Barroso, H. (**1999**). *Forma e substância de expressão da língua portuguesa (Form and substance of the Portuguese language expression)* (Almedina, Coimbra).

Bisol, L. (**1996**). *Introdução a estudos de fonologia do português brasileiro (Introduction to studies on the phonology of Brazilian Portuguese)* (Editora Universitária da Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre).

Boersma, P., and Weenink, D. (**2008**). "Praat: doing phonetics by computer (Version 5.0.43)" [Computer program], retrieved 9 December 2008 from http://www.praat.org/.

Byrd, D. (**1992**). "Preliminary results on speaker-dependent variation in the TIMIT database," J. Acoust. Soc. Am. **92**, 593–596.

Callou, D., Moraes, J., and Leite, Y. (**1996**). "O vocalismo do português do Brasil (The vocalism of the Portuguese of Brazil)," Letras de Hoje (Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre) **31**(2), 27–40.

Câmara, J. M., Jr. (**1970**). *Estrutura da língua portuguesa (Structure of the Portuguese Language)* (Vozes, Petrópolis).

Cervera, T., Miralles, J. L., and González-Álvarez, J. (**2001**). "Acoustical analysis of Spanish vowels produced by laryngectomized subjects," J. Speech Lang. Hear. Res. **44**, 988–996.

Chládková, K., Boersma, P., and Podlipský, V. J. (**2009**). "On-line formant shifting as a function of F0," in Proceedings of Interspeech 2009.

Clopper, C. G., Pisoni, D. B., and De Jong, K. (**2005**). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118**, 1661–1676.
Cristófaro Silva, T. (**2002**). *Fonética e fonologia do português (The Phonetics and Phonology of Portuguese)* (Contexto, São Paulo).
Delgado-Martins, M. R. (**1973**). "Análise acústica das vogais orais tônicas em português (Acoustic analysis of the stressed oral vowels in Portuguese)," Boletim de Filologia (University of Lisbon) **22**, 303–314.
Delgado-Martins, M. R. (**2002**). *Fonética do português: trinta anos de investigação (The Phonetics of Portuguese: Thirty Years of Research)* (Caminho, Lisbon).
Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.
Diehl, R. L., Lindblom, B., Hoemeke, K. A., and Fahey, R. P. (**1996**). "On explaining certain male-female differences in the phonetic realization of vowel categories," J. Phonetics **24**, 187–208.
Ewan, W., and Krones, R. (**1974**). "Measuring larynx movement using the thyroumbrometer," J. Phonetics **2**, 327–335.
Falé, I. (**1998**). "Duração das vogais tónicas e fronteiras prosódicas: uma análise em estruturas coordenadas (Duration of stressed vowels and prosodic boundaries: An analysis on coordinated structures)," Actas do XIII Encontro Nacional da Associação Portuguesa de Linguística (Colibri, Lisbon), pp. 255–269.
Gibbon, J. (**1977**). "Scalar expectancy theory and Weber's Law in animal timing," Psychol. Rev. **84**, 279–325.
Goldstein, U. (**1980**). "An articulatory model for the vocal tracts of growing children," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.
Hagiwara, R. (**1997**). "Dialect variation and formant frequency: The American English vowels revisited," J. Acoust. Soc. Am. **102**, 655–658.
Harrington, J., Kleber, F., and Reubold, U. (**2008**). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," J. Acoust. Soc. Am. **123**, 2825–2835.
Heffner, R.-M. (**1937**). "Notes on the length of vowels," Am. Speech **12**, 128–134.
Henton, C. G. (**1989**). "Fact and fiction in the description of female and male pitch," Language & Communication **9**, 299–311.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
House, A. S., and Fairbanks, G. (**1953**). "The influence of consonant environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.
Kent, R. D., and Read, C. (**2002**). *The Acoustic Analysis of Speech*, 2nd ed. (Singular, San Diego).
Kock, A. (**1911**). *Svensk ljudhistoria (Swedish Sound History)* (Gleerup, Lund), Vol. **2**.
Koopmans-van Beinum, F. J. (**1980**). "Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions," Ph.D. thesis, University of Amsterdam.
Labov, W. (**1994**). *Principles of Linguistic Change. Volume I: Internal Factors* (Blackwell, Oxford).
Landick, M. (**1995**). "The mid-vowels in figures: hard facts," The French Review **69**, 88–102.
Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA).
Lehiste, I., and Peterson, G. E. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.
Lindblom, B. (**1967**). "Vowel duration and a model of lip-mandible coordination," Speech Transm. Lab. Q. Prog. Status Rep. **4**, 1–29.
Luce, P. A., and Charles-Luce, J. (**1985**). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," J. Acoust. Soc. Am. **78**, 1949–1957.
Maiden, M. (**1997**). "Vowel systems," in *The Dialects of Italy*, edited by M. Maiden and M. Parry (Routledge, London), pp. 7–14.
Mateus, M. H. M. (**1990**). *Fonética, fonologia e morfologia do português (The Phonetics, Phonology, and Morphology of Portuguese)* (Universidade Aberta, Lisbon).
Mateus, M. H. M., and d'Andrade, E. (**1998**). "The syllable structure in European Portuguese," DELTA [Documentação de Estudos em Linguística Teórica e Aplicada] (Pontifícia Universidade Católica de São Paulo, São Paulo) **14**, 13–32.
Mateus, M. H. M., and d'Andrade, E. (**2000**). *The Phonology of Portuguese* (Oxford University Press, Oxford).
Mateus, M. H. M., Falé, I., and Freitas, M. (**2005**). *Fonética e fonologia do português (Portuguese Phonetics and Phonology)* (Universidade Aberta, Lisbon).
Meyer-Lübke, W. (**1908**). *Historische Grammatik der französischen Sprache. 1. Laut- und Flexionslehre (Historical Grammar of the French Language. 1. Phonology and Inflectional Morphology)* (Carl Winter, Heidelberg).
Moraes, J. A. (**1999**). "Um algoritmo para a correção/simulação da duração dos segmentos vocálicos em português (An algorithm to correct/simulate duration in Portuguese vocalic segments)," in *Estudos da prosódia (Prosody Studies)*, edited by E. Scarpa (Editora da Unicamp, Campinas), pp. 69–84.
Moraes, J. A., Callou, D., and Leite, Y. (**1996**). "O sistema vocálico do português do Brasil: caracterização acústica (The vocalic system of the Portuguese of Brazil: Acoustic characterization)," in *Gramática do português falado (The Grammar of Spoken Portuguese)*, edited by M. Kato (Editora da Unicamp, Campinas), pp. 33–53.
Morrison, G. S., and Escudero, P. (**2007**). "A cross-dialect comparison of Peninsular- and Peruvian-Spanish vowels," in Proceedings of the 16th Congress of Phonetic Sciences, Saarbrücken, pp. 1505–1508.
Most, T., Amir, O., and Tobin, Y. (**2000**). "The Hebrew vowel system: Raw and normalized acoustic data," Lang Speech **43**, 295–308.
Nearey, T. M., Assmann, P. F., and Hillenbrand, J. M. (**2002**). "Evaluation of a strategy for automatic formant tracking," J. Acoust. Soc. Am. **112**, 2323.
Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., and Trent-Brown, S. (**2008**). "Acoustic and perceptual similarity of Japanese and American English vowels," J. Acoust. Soc. Am. **124**, 576–588.
Ohala, J. J., and Eukel, B. (**1987**). "Explaining the intrinsic pitch of vowels," in *In Honor of Ilse Lehiste*, edited by R. Channon and L. Shockey (Foris, Dordrecht), pp. 207–215.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.
Recasens, D., and Espinosa, A. (**2009**). "Dispersion and variability in Catalan five and six peripheral vowel systems," Speech Commun. **51**, 240–258.
Riordan, C. J. (**1977**). "Control of vocal-tract length in speech," J. Acoust. Soc. Am. **62**, 998–1002.
Rochet, A. P., and Rochet, B. L. (**1991**). "The effect of vowel height on patterns of assimilation nasality in French and English," in Proceedings of the 12th International Congress of Phonetic Sciences, Aix, Vol. **3**, pp. 54–57.
Ryalls, J. H., and Lieberman, P. (**1982**). "Fundamental frequency and vowel perception," J. Acoust. Soc. Am. **72**, 1631–1634.
Schönfeld, M. (**1932**). *Historiese grammatika van het Nederlands (Historical Grammar of Dutch)* (Thieme, Zutphen).
Seara, I. C. (**2000**). "Estudo acústico-perceptual da nasalidade das vogais do português brasileiro (Acoustical-perceptual study on the nasality of the vowels of Brazilian Portuguese)," Ph.D. thesis, Universidade Federal de Santa Catarina, Florianópolis.
Sihler, A. L. (**1995**). *New Comparative Grammar of Greek and Latin* (Oxford University Press, New York).
Simpson, A. P. (**2001**). "Dynamic consequences of differences in male and female vocal tract dimensions," J. Acoust. Soc. Am. **109**, 2153–2164.
Simpson, A. P. (**2002**). "Gender-specific articulatory-acoustic relations in vowel sequences," J. Phonetics **30**, 417–435.
Simpson, A. P. (**2003**). "Possible articulatory reasons for sex-specific differences in vowel duration," in Proceedings of the sixth International Seminar on Speech Production, Sydney, pp. 261–266.
Simpson, A. P., and Ericsdotter, C. (**2003**). "Sex-specific durational differences in English and Swedish," in Proceedings of the 15th Congress of Phonetic Sciences, Barcelona, pp. 1113–1116.
Solé, M. J. (**2007**). "Controlled and mechanical properties in speech: a review of the literature," in *Experimental Approaches to Phonology*, edited by M. J. Solé, P. Beddor and M. Ohala (Oxford University Press, Oxford), pp. 302–321.
Stieber, Z. (**1979**). *Zarys gramatyki prorównawczej języków slowiańskich (An Outline of the Comparative Grammar of the Slavic Languages)* (Państwowe Wydawnictwo Naukowe, Warsaw).
Stevens, K. (**1998**). *Acoustic Phonetics* (MIT, Cambridge, MA).
Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (**2007**). "Acoustic variability within and across German, French, and

American English vowels: Phonetic context effects," J. Acoust. Soc. Am. **122**, 1111–1129.
Tielen, M. T. J. (**1992**). "Male and female speech: An experimental study of sex-related voice and pronunciation characteristics," Ph.D. thesis, University of Amsterdam.
Whalen, D. H., and Levitt, A. G. (**1995**). "The universality of intrinsic F_0 of vowels," J. Phonetics **23**, 349–366.
Whiteside, S. P. (**1996**). "Temporal-based acoustic-phonetic patterns in read speech: Some evidence for speaker sex differences," J. Int. Phonetic Assoc. **26**, 23–40.
Winer, B. J. (**1962**). *Statistical Principles in Experimental Design* (McGraw-Hill, New York).
Yamazawa, H., and Hollien, H. (**1992**). "Speaking fundamental frequency patterns of Japanese women," Phonetica **49**, 128–140.

Acoustic markers of sarcasm in Cantonese and English

Henry S. Cheang[a] and Marc D. Pell
School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada

(Received 19 June 2008; revised 16 June 2009; accepted 18 June 2009)

The goal of this study was to identify acoustic parameters associated with the expression of sarcasm by Cantonese speakers, and to compare the observed features to similar data on English [Cheang, H. S. and Pell, M. D. (2008). Speech Commun. **50**, 366–381]. Six native Cantonese speakers produced utterances to express sarcasm, humorous irony, sincerity, and neutrality. Each utterance was analyzed to determine the mean fundamental frequency (F0), F0-range, mean amplitude, amplitude-range, speech rate, and harmonics-to-noise ratio (HNR) (to probe voice quality changes). Results showed that sarcastic utterances in Cantonese were produced with an elevated mean F0, and reductions in amplitude- and F0-range, which differentiated them most from sincere utterances. Sarcasm was also spoken with a slower speech rate and a higher HNR (i.e., less vocal noise) than the other attitudes in certain linguistic contexts. Direct Cantonese-English comparisons revealed one major distinction in the acoustic pattern for communicating sarcasm across the two languages: Cantonese speakers *raised* mean F0 to mark sarcasm, whereas English speakers *lowered* mean F0 in this context. These findings emphasize that prosody is instrumental for marking non-literal intentions in speech such as sarcasm in Cantonese as well as in other languages. However, the specific acoustic conventions for communicating sarcasm seem to vary among languages.
 [DOI: 10.1121/1.3177275]

PACS number(s): 43.70.Kv, 43.70.Fq, 43.70.Mn [AL] Pages: 1394–1405

I. INTRODUCTION

Speech prosody (i.e., intonation and stress patterns) conveys many types of information to listeners, including whether or not a speaker intends to be ironic or sarcastic (the latter being a subtype of verbal irony). Verbal irony occurs when the intended meaning of statements is opposite to, or different from, the literal sense of the words used (see Gibbs, 2000 for a detailed description). The unique feature of sarcasm as a form of verbal irony is that it is chiefly used to express *negative* critical attitudes. Several research sources have highlighted the importance of prosody as a cue for detecting sarcasm; for example, adult listeners have been found to identify sarcastic intent in content-filtered utterances (Bryant and Fox Tree, 2005; Rockwell, 2000a). It has also been shown that young children can recognize the intonational markers of sarcasm, and this ability is developmentally distinct from the ability to recognize sarcasm through semantic or contextual cues of speech (Ackerman, 1983, 1986; Capelli *et al.*, 1990; Laval and Bert-Erboul, 2005; Winner and Leekman, 1991).

Based on acoustic studies, it is likely that sarcasm is encoded in speech through various global manipulations in acoustic parameters such as fundamental frequency (F0) and F0 variability, amplitude and amplitude variability, speech rate, voice quality, and resonance (Attardo *et al.*, 2003; Bryant and Fox Tree, 2005; Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005; Schaffer, 1982). However, owing to methodological differences across studies, the available data are marked by uncertainty, and the relative importance of particular acoustic parameters for signaling sarcasm and their directionality cannot be fully determined. One limitation in literature on the expression of various attitudinal states is that some studies did not control for the possibility that speakers employ semantic cues (i.e., words or phrases; henceforth, "keyphrases") independently of, or in conjunction with, prosodic cues to express such information (see Scherer *et al.*, 1984). With particular respect to sarcasm, there is evidence that some keyphrases are used so frequently by speakers to convey sarcasm that these keyphrases become semantic markers of sarcasm regardless of overlaid prosodic features (i.e., "enantiosemantic," Haiman, 1998, p. 39). Another possible shortcoming in literature is that early acoustic descriptions of sarcastic prosody may be conflated with the acoustic markers of other forms of verbal irony; for example, Anolli *et al.* (2002) found significant acoustic differences between sentences, which were meant to convey positive verbal irony (i.e., a playful, humorous attitude) and negative verbal irony (i.e., sarcasm). However, this distinction is often not controlled in literature, promoting some uncertainty about the true prosodic markers of sarcasm.

Recently, Cheang and Pell (2008) reported a detailed acoustic investigation of sarcasm expressed in English. In that study, six native English speakers produced a common set of utterances to convey sarcasm, positive humorous irony, sincerity, or neutrality. Some of the sentences included keyphrases that previous authors claimed to be enantiosemantic of sarcasm in English (e.g., "I suppose."), whereas other sen-

[a]Author to whom correspondence should be addressed. Present address: Laboratoire de Recherche en Neurosciences et Électrophysiologie Cognitive, Service de Recherche, Hôpital Rivière-des-Prairies, 7070 Boulevard Perras, Montréal, Québec H1E 1A4, Canada. Electronic mail: henry.cheang@mail.mcgill.ca

tences could only be understood as sarcastic from the speaker's prosody. Following acoustic analyses of the recordings, it was found that sarcastic utterances were associated with a significantly lower mean F0 than utterances, which conveyed humorous irony or sincerity. Moreover, sarcasm was characterized by reductions in F0 variation (standard deviation) and in the harmonics-to-noise ratio (HNR) (a measure of voice quality) when compared to sincerity. Since these acoustic changes were observed for sentences with and without enantiosemantic keyphrases, they appear to be central vocal cues for encoding sarcasm in English, which are text-independent (Cheang and Pell, 2008). Speech rate and resonance changes also distinguished sarcasm from humor and sincerity but only in the context of specific sentence types. Cheang and Pell (2008) concluded that certain prosodic cues (e.g., reduced mean F0) may be central for expressing sarcasm in English, whereas speakers employ other acoustic cues to signal sarcasm in particular linguistic contexts.

However, these findings are restricted to how sarcasm is expressed in English and do not address potential cross-language differences in how prosody is used to convey intentions such as sarcasm. There is evidence that culture norms promote differences in how particular affective or attitudinal states are communicated across languages (Grabe *et al.*, 2003; McCluskey *et al.*, 1975). In fact, careful inspection of literature on sarcasm implies that there are certain cross-linguistic differences in the acoustic cues to sarcasm. French sarcastic utterances have been associated with higher F0, restricted F0-range, and a slower speech rate (Laval and Bert-Erboul, 2005). In contrast, sarcasm in Italian has been characterized as having a higher F0, increased F0-range, and greater amplitude (Anolli *et al.*, 2002). Despite differences in how acoustic cues appear to be used across languages (especially F0), it is possible that some of the similarities noted may reflect common physiological tendencies associated with sarcastic speech. While sarcasm cannot be considered an affective state, this intention is inherently marked by its negative valence; certain vocal and facial gestures associated with the expression of negative affect (such as a disgusted sneer) sometimes accompany the expression of sarcasm in speech (Cutler, 1974; Haiman, 1998). One can assume that any effects of these negative physiological responses on prosody would be relatively comparable across languages. For example, the disgusted sneer is associated with heightened tension in the orofacial region, which contributes to predictable changes in resonance and voice quality (Scherer, 1986).

Further studies of how sarcasm is conveyed in different languages, and especially in a language, which is highly distinct from English, would help to reconcile the observed differences in literature, and identify possible similarities in the acoustic expression of sarcasm across languages. Cantonese is a good language to study for this purpose because Cantonese and English have no common linguistic roots, and there are enormous cultural distance and communication style differences between speakers of these languages (Bond, 1991; Ho, 1986; Huang and Kok, 1999; Snow, 2004). Little research has been done to quantify the acoustic and linguistic markers of sarcasm in Cantonese; however, it appears that some idioms or syntactic forms predispose listeners to recognize sarcastic intentions in this language (Chan, 2001; Killingley, 1986; Matthews and Yip, 1994). There are hints that Cantonese speakers also use acoustic cues to signal sarcasm [Kwok and Luke, 1986, as cited by Bauer and Benedict (1997)], but no empirical evidence is available and it is difficult to dissociate potential prosodic markers from cues derived from semantics, connotation, or communication rules. A prosodic evaluation of Cantonese sarcasm should shed light on its acoustic specifications, which can also be usefully compared to known acoustic cues of sarcasm in other languages.

Following Cheang and Pell (2008), the major goal of this study was to evaluate whether Cantonese speakers utilize specific acoustic cues to express sarcasm when compared to other attitudes, and whether these acoustic features interact with specific phrase types (i.e., keyphrases) associated with sarcasm. A second aim was to compare these new acoustic findings for Cantonese with the data on English sarcastic speech (Cheang and Pell, 2008). To allow for these comparisons, a procedure highly similar to that of Cheang and Pell's (2008) investigation was adopted: Native (Cantonese) speakers were recruited to convey sarcasm, humor, sincerity, and neutrality in simple sentences using only prosodic cues. Following a perceptual validation procedure, involving a separate group of native Cantonese-speaking listeners, acoustic analyses of the sentences were conducted to determine whether the four attitudes were characterized by specific patterns of prosodic cues. The Cantonese sentences were then acoustically compared with the English sentences analyzed previously (Cheang and Pell, 2008). In a broad context, this research will contribute to a new understanding of how attitudes are conveyed extra-linguistically and across languages, and could represent an important step toward bridging potential cross-cultural misunderstandings in verbal communication.

As the context and approach of this study were relatively novel, firm predictions about the Cantonese speakers could not always be made, and cross-language comparisons of sarcastic speech should be considered exploratory in nature. Nonetheless, based on the literature cited above, it was hypothesized that sarcastic utterances in Cantonese would differ significantly from utterances conveying positive, humorous irony, as well as sincerity on measures of F0 [Kwok and Luke, 1986, as cited by Bauer and Benedict (1997)]. Also, since speaker manipulations of amplitude, speech rate, and voice quality are frequently cited as markers of sarcasm (Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005), it was anticipated that some of these features would also distinguish sarcasm from the other attitudes in Cantonese. Given the considerable linguistic and cultural distinctions between Cantonese and English, no strong predictions about the nature of Cantonese-English differences in sarcastic prosody could be made, although it was speculated that F0 would be used in some way to communicate sarcasm in both languages.

TABLE I. Text through which speakers articulated the four target attitudes (sarcasm, humor, sincerity, and neutrality). "A" items denote combined sentences, "B" items denote single sentences, and "C" items denote keyphrases. The text of the exemplars across the Cantonese and English samples were highly comparable. Biasing sentences for the Cantonese tokens are not presented, as they were close translations of the English materials. As there are presently no truly universal transcription schemes for Cantonese, the transcriptions of the text are based on those summarized in Matthews and Yip (1994), where words are transcribed with the phonemic tone identified following vowel identification. Note that there are six Cantonese vowel tones: high level (55), mid level (33), low level (22), high rising (35), low rising (23), and low falling (21).

Item	Attitude	Biasing sentence	Target utterance (Cantonese)	Target utterance (English)
1	Sarcasm	Don't you just love how your stupid mother-in-law always smirks and snorts loudly when you misspeak?	A: 係^{口}卦，呢個係個好客氣^{口}既表示。 hai22 gwa33, lei55 go33 hai22 go33 hou35 haak33 hei33 ge5 biu35 si22.	A: I suppose; it's a respectful gesture.
	Humor	Not everyday that you see a priest give the finger, is it?	B: 呢個係個好客氣^{口}既表示。 lei55 go33 hai22 go33 hou35 haak33 hei33 ge5 biu35 si22.	B: It's a respectful gesture.
	Sincerity	It was nice of your supervisor to send flowers.	C: 係^{口}卦。 hai22 gwa33.	C: I suppose.
2	Sarcasm	That horrid woman smokes a pack a day.	A: 係咩; 佢係個好健康^{口}既女人。 hai22 me55, keui5 hai22 go33 hou35 gin22 hong55 ge23 neui23 yan35.	A: Is that so; she is a healthy lady.
	Humor	Your friend Shelley can't even do a single pushup.	B: 佢係個好健康^{口}既女人。 Keui23 hai22 go33 hou35 gin22 hong55 ge23 neui23 yan35.	B: She is a healthy lady.
	Sincerity	She runs 10 miles everyday.	C: 係咩。 hai22 me55	C: Is that so?
3	Sarcasm	Our moronic boss gave us all food poisoning.	A: 嘩哎；佢係個好鬼叻^{口}既廚師。 wa55 aai55, keui23 hai22 go33 hou35 gwai35 lek55 ge23 cheui21 si55.	A: Oh boy; he is a superior chef.
	Humor	Your brother singed his eyebrows while making toast.	B: 佢係個好鬼叻^{口}既廚師。 Keui23 hai22 go33 hou35 gwai35 lek55 ge23 cheui21 si55.	B: He is a superior chef.
	Sincerity	Butch just won another cooking contest; this is his twentieth win in the last three years.	C: 嘩哎。 wa55 aai55.	C: Oh boy.
4	Sarcasm	The arrogant front-runner finished dead last.	A: 係囉； 呢個係個犀利^{口}既結果。 hai22 lo55, lei55 go33 hai22 go33 sai55 lei22 ge5 git33 gwo35.	A: Yeah, right; what a spectacular result.
	Humor	Fascinating how she lost the eating contest to someone half her size huh?	B: 呢個係個犀利^{口}既結果。 lei55 go33 hai22 go33 sai55 lei22 ge23 git33 gwo35.	B: What a spectacular result.
	Sincerity	He broke three records in that race!	C: 係囉。 hai22 lo55.	C: Yeah, right.

II. METHOD

A. Stimulus production: Encoders and materials

The "encoders" were six native Cantonese speakers who were living in Montreal, Canada (3 males, 3 females; mean age: 22.7 years, SD: 3.2 years; mean education: 17.0 years, SD: 2.0 years). All encoders were born, raised, and educated in Hong Kong or Guangzhou (i.e., cities where Cantonese is primarily spoken) and moved to Canada as young adults. They were late learners of additional languages (English and/or French) and used Cantonese exclusively when at home. As with the six English encoders studied by Cheang and Pell (2008), the present encoders had no formal training in acting.

A set of 96 Cantonese utterances was recorded from each encoder—24 utterances representing each of the 4 attitudes investigated by Cheang and Pell (2008): sarcasm, positive humorous irony, sincerity, and neutrality. Henceforth, positive humorous irony will be referred to as "humor" for brevity; while it is acknowledged that humor does not have to be ironic in nature, here "humor" will always refer to a positive form of irony associated with playful, humorous intent (Anolli *et al.*, 2002). Sincerity typically refers to a speaker's attempt to reinforce the literal meaning of their utterance using prosody or other cues. Neutrality, which marks instances where the speaker does not wish to convey obvious emotions or intentions through prosody, is known to have a distinct prosodic form in different languages (e.g., Pell *et al.*, 2009) and was selected as a baseline category for data interpretation. Given the secondary objective of evaluating whether Cantonese speakers modulate acoustic cues in conjunction with particular semantic cues, a subset of the items included Cantonese idioms ("keyphrases") associated with sarcastic messages. Overall, the 96 utterances consisted of an equal number of keyphrases, sentences, and "combined sentences," which were composed of the keyphrases and the sentences. An example of a combined sentence is "係咩; 佢係個好健康口既女人" (English meaning: "Is that so; she is a healthy lady"), for which "係咩" ("Is that so") is the keyphrase and "佢係個好健康口既女人"("She is a healthy lady") is the sentence. The text of the items ranged from 2 syllables for keyphrases to 9–11 syllables for combined sentences. Four such sets of utterance forms were created (see Table I).

In all cases, materials were devised to be as semanti-

cally, syntactically, and syllabically comparable as possible to the English materials constructed in Cheang and Pell, 2008. All items could be produced by the encoders to express different attitudes using identical text. Common words and idioms were used in the construction of all stimuli.[1] Also, care was taken to evenly distribute the six Cantonese phonemic vowel tones across the text of the tokens, except for the low falling tone, which occurred less frequently than the other tones (see transcriptions in Table I). This limitation is unlikely to be critical, as earlier work on Cantonese suggests that cues used to mark semantic information through phonemic tone shape are relatively independent of global cues, which mark extra-linguistic information in the sentence (Vance, 1976).

Each target utterance was produced in response to a biasing sentence, except in the neutral condition where encoders simply produced the utterance in isolation. For sarcasm, humor, and sincerity, the encoders produced each target utterance as if they were engaging in a scripted dialog. Biasing sentences provided a context that would facilitate production of the associated attitude; in the case of sarcasm, these sentences included insulting, cruel, or unfairly critical cues (e.g., "That horrid woman smokes a pack a day" to bias a sarcastic rendition of "She is a healthy lady"). Sentences biasing humor production included overt declaration of a friendly relation or playful cues (e.g., "Your friend can't even do a single push-up"). Biasing sentences used to help elicit sincerity did not contain information that suggested positive, negative, or other possible associations (e.g., "She runs 10 miles every day"). Moreover, encoders were presented with detailed explanations regarding the attitudes they were to reproduce prior to the recording of each set of attitudes, although specific acoustic attributes were intentionally not highlighted to them. For each attitude, identical biasing sentences were used for each item of the three phrase types. Each utterance was recorded twice non-sequentially from each encoder (review Table I).

As part of stimulus development, the text of the recording materials was presented to four native Cantonese speakers in a pilot reading study to judge the suitability of the sentence pairs in portraying situational contexts appropriate to each target attitude. Raters had to classify the pairs as being "unnatural," "somewhat natural," or "natural." Target utterances and biasing sentences were refined based on these ratings and presented to different raters. Refinement continued in this fashion until there was a minimum of 75% agreement across raters that sentence pairs reflected natural interactions.

B. Recording procedure

Interactions between the experimenter and participants were conducted exclusively in Cantonese. The entire set of neutral utterances was always elicited first from each encoder. All encoders were instructed to read aloud target sentences printed on cards in a neutral voice, devoid of affect. After recording the full set of neutral utterances, three separate sets of utterances (each conveying only one of the remaining attitudes) were recorded one at a time in random sequence across encoders. The recording of each attitude was blocked to help encoders to successfully adopt each expressive mode. Within each attitude set, target sentences were presented in a fixed random order.

Prior to producing sentences representing each attitude, encoders were provided definitions of each attitude and short, standardized descriptions of situations under which these attitudes are expressed (e.g., "people use sarcastic utterances to respond to insulting comments directed at them;" "people use humorous statements to be playful with friends"). Encoders were given no indication of which acoustic cues to employ during the recording procedure; rather, they were instructed to use these descriptions and the biasing sentences to facilitate their enactments of the target attitude. Sentence pairs consisting of a biasing sentence and an associate phrase type (i.e., keyphrase, sentence, or combined sentence) were presented to the encoder on printed cards. Encoders read the biasing sentence silently and then produced the target sentence aloud to communicate the target attitude. Encoders were given practice trials for each attitude prior to recording the experimental items. Encoders were neither coached nor given feedback regarding their renditions of the target attitudes, although they were allowed to repeat their productions if desired (the final exemplar was always retained for analysis). Recordings were conducted in a sound-attenuated booth, captured by a AKG C-420 head-mounted professional microphone onto a Sony TCD-D100 digital audio tape recorder (sampling rate: 44.1 kHz, 16 bits, mono). The digital recordings were transferred directly to a computer (downsampled to 24 kHz but unfiltered) for editing and acoustic analyses using PRAAT software (Boersma and Weenink, 2006). A total of 576 utterances were recorded (4 attitudes $\times$ 4 items $\times$ 3 phrase types $\times$ 2 repetitions $\times$ 6 encoders).

C. Perceptual validation study

In this study, attitudes were "posed" to control for the linguistic-semantic structure of the items submitted to acoustic analysis. Given this approach, it was necessary to first establish the representativeness and validity of each token, to ensure that acoustic measures referred to perceptually identifiable exemplars of each attitude. Therefore, a perceptual validation study was conducted prior to the acoustic analyses, involving a separate group of 16 native Cantonese "decoders" (8 males, 8 females; mean age: 24.9 years, SD: 6.0 years; mean education: 17.9 years, SD: 3.5 years). The decoders were recruited from the same population as the encoders (i.e., were born, raised, and educated in Hong Kong or Guangzhou and were living in Montreal). Each decoder was required to identify the attitude expressed by each utterance, based on their knowledge of how prosodic cues (and when applicable, enantiosemantic terms) are used to convey sarcasm, humor, sincerity, and neutrality. This procedure was designed to eliminate utterances, which did not reliably convey the intended meanings due to difficulties at the stage of *simulating* attitudes, while establishing the perceptual validity of utterances identified by a majority of decoders as representing one of the target attitudes.

TABLE II. Number of utterances retained from the perceptual validation study (with percent agreement across decoders in parentheses) for each attitude and phrase type condition.

	Attitude				
Phrase type	Sarcasm	Sincerity	Humor	Neutrality	Total
Keyphrase	58 (65%)	8 (57%)	26 (69%)	23 (57%)	115
Sentence	5 (60%)	89 (73%)	22 (62%)	33 (74%)	149
Combined sentence	22 (59%)	44 (60%)	25 (55%)	30 (69%)	121
Total	85	141	73	86	385

Note: Items were retained and/or reclassified based on a minimum 50% agreement about the intended attitude when judged by the 16 decoders.

For the validation experiment, all 576 utterances were presented in 19 blocks (i.e., approximately 30 utterances per block), and each decoder identified the intended attitude of the speaker in a forced choice decision task (sarcasm, humor, sincerity, and neutrality). Following previous methods (Cheang and Pell, 2008), utterances were retained as "valid" exemplars for acoustic analysis if the final recognition consensus of the 16 decoders was at least two times chance performance levels (i.e., 50% or greater, where 25% consensus represented chance levels). When a speaker's intended attitude was strongly recognized as a different attitude, the item was reclassified in favor of the perceptual ratings; this procedure helped to control for possible mismatches between perceived and intended meaning in utterances and to retain the maximum number of items for acoustic analysis. Following the perceptual validation study, 385 (67%) of the initial utterances were retained as valid exemplars of the four attitudes, as summarized in Table II.

D. Acoustic analyses

The choice of acoustic parameters was guided by the authors' previous findings on English sarcasm (Cheang and Pell, 2008), trends in sarcasm research (e.g., Anolli *et al.*, 2002; Attardo *et al.*, 2003; Cutler, 1974, 1976; Haiman, 1998; Rockwell, 2000a), and empirical observations of acoustic patterns that characterize Cantonese (Bauer and Benedict, 1997; Fok, 1974; Vance, 1976). All acoustic analyses were performed using PRAAT; for F0/pitch, these analyses were conducted automatically using an autocorrelation method, the results were smoothed, and the output was visually inspected and manually corrected in the event of "halving" and "doubling" errors in the data. The specific acoustic measures derived for each utterance were given as follows:

(1) mean F0—measured in hertz for each utterance as a whole;
(2) F0-range—computed by subtracting the minimum F0 value from the maximum F0 value of the full utterance, to estimate the degree of F0 variation;[2]
(3) mean amplitude—measured in decibel for each utterance as a whole;
(4) amplitude-range—computed by subtracting the minimum amplitude value from the maximum amplitude value of each exemplar, to estimate amplitude variation;
(5) speech rate—calculated in syllables/s by dividing the number of syllables by the total utterance duration; and
(6) HNR—computed as the ratio of the averaged periodic component of a sound signal to the corresponding averaged noise component, expressed in decibel (Yumoto *et al.*, 1982); HNR measures were taken from 50-ms stable central portions of vowels segmented from stressed syllables (most vowels in unstressed syllables were shorter than 50 ms).

E. Statistical procedure

Prior to statistical analysis, all acoustic measures pertaining to tokens produced by a single encoder (irrespective of attitude) were converted into *z*-scores to allow comparisons across items and encoders. All individual values of F0, amplitude, and speech rate were standardized separately per encoder in reference to his or her entire set of productions by dividing the difference between the averaged value of all exemplars from an individual data point by the standard deviation of all exemplars [e.g., ([mean F0(one exemplar of sarcasm) − mean F0(all exemplars)])/F0 SD(all exemplars)]. For values of HNR only, all HNR values taken from the stressed vowels of a given exemplar were first averaged prior to standardization as described above.

Following normalization, *z*-scores of the acoustic data from the Cantonese exemplars were subjected to separate analyses of variance (ANOVAs) involving the factors of attitude (sarcasm, humor, sincerity, and neutrality) and phrase type (keyphrase, sentence, and combined sentence), independently of each normalized acoustic measure. After characterizing which acoustic cues contributed to the expression of different attitudes in Cantonese, a second set of ANOVAs directly compared acoustic values of Cantonese versus English; these analyses considered the fixed variable of language (Cantonese and English) with repeated measures on attitude (sarcasm, humor, sincerity, and neutrality). Phrase type was omitted in the cross-language comparison to concentrate analyses on how speakers of the two different languages vary in their acoustic expression of sarcasm. In all cases, significant effects and interactions were explored *post hoc* using Tukey's honestly significant difference (HSD) method ($\alpha = 0.05$). Significant main effects that were subsumed by higher-order interactions are reported but not described in the text.

TABLE III. Mean normalized acoustic measures (and standard deviation) of Cantonese utterances expressing each of the four attitudes, divided by phrase type.

Phrase type	Attitude	Mean F0 (Hz)	F0-range (Hz)[a]	Mean amplitude (dB)	Amplitude-range (dB)[a]	Speech rate (syllables/s)	HNR[b] (dB)
Keyphrase	Sarcasm	0.64(1.40)	−0.36(1.17)	0.66(0.90)	−0.89(0.73)	−0.96(0.69)	0.61(1.01)
	Sincerity	−0.29(0.74)	−0.88(0.46)	0.50(0.99)	−1.23(0.6)	−0.79(0.87)	0.66(1.22)
	Humor	0.25(1.26)	−0.16(1.00)	0.49(0.84)	−1.33(0.70)	−1.27(0.62)	−0.93(1.20)
	Neutrality	−0.50(0.89)	−0.68(1.28)	−0.27(0.98)	−0.64(0.91)	−0.87(0.78)	0.27(1.62)
Sentence	Sarcasm	−0.03(0.84)	−0.12(0.43)	0.25(0.39)	−0.14(0.49)	0.04(0.45)	0.04(0.90)
	Sincerity	−0.01(0.72)	0.15(0.82)	0.33(0.71)	−0.07(0.70)	1.03(0.47)	−0.17(0.71)
	Humor	0.03(1.07)	0.07(1.08)	0.12(0.89)	−0.01(0.61)	0.96(0.51)	−0.81(1.59)
	Neutrality	−0.74(0.39)	−0.33(0.55)	−0.38(0.84)	0.47(0.58)	0.05(0.67)	−0.04(0.85)
Combined sentence	Sarcasm	0.12(0.69)	0.47(0.70)	−0.34(0.91)	0.92(0.61)	−0.04(0.45)	−0.06(0.63)
	Sincerity	0.15(0.69)	0.54(0.85)	−0.42(0.89)	0.78(0.64)	0.22(0.59)	−0.02(0.73)
	Humor	0.34(1.02)	0.58(1.21)	−0.41(0.90)	0.65(0.60)	0.16(0.67)	0.24(0.97)
	Neutrality	−0.74(0.29)	−0.13(0.60)	−1.13(1.00)	1.06(0.52)	−0.41(0.69)	0.01(0.53)

[a]Range=maximum−minimum.
[b]HNR=harmonics-to-noise ratio.

III. RESULTS

Table III summarizes the (normalized) acoustic features of Cantonese expressions of sarcasm, humor, sincerity, and neutrality, separately by phrase type.

A. Fundamental frequency: Mean and range

A 4×3 (attitude×phrase type) repeated-measures ANOVA performed on mean F0 yielded a significant main effect for attitude, $F(3,373)=14.91$, $p<0.0001$. *Post hoc* comparisons indicated that utterances conveying sarcasm were produced with a significantly higher mean F0 than sincerity overall. In addition, neutrality was spoken with a significantly lower mean F0 than all other attitudes. No significant interactions were found. Patterns of mean F0 are illustrated in Fig. 1.

Analysis of F0-range yielded main effects for attitude, $F(3,373)=4.66$, $p=0.003$, and phrase type, $F(2,373)=20.12$, $p<0.0001$. The F0-range of sarcastic utterances was significantly narrower than that of sincere tokens, and neutral utterances were produced with a significantly smaller F0-range than humor and sincerity. For the phrase type main effect, *post hoc* comparisons showed that combined sentences were produced with the widest F0-range overall, which was greater than sentences, followed by keyphrases, which exhibited a relatively narrow F0-range. No significant interactions were found.

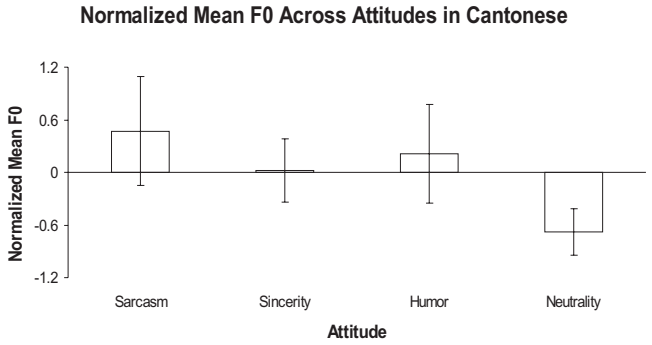


FIG. 1. Mean fundamental frequency (F0) of Cantonese utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

B. Amplitude: Mean and range

The 4×3 ANOVA on mean amplitude yielded a significant main effect of attitude, $F(3,373)=12.89$, $p<0.0001$. This effect was explained by the fact that neutrality was produced with a lower mean amplitude than the other three attitudes. There was also a significant main effect of phrase type, $F(2,373)=27.62$, $p<0.0001$, which revealed that keyphrases were spoken with the greatest amplitude, followed by sentences, which displayed greater amplitude than combined sentences. No significant interactions were found.

Analyses of amplitude-range yielded a significant main effect of attitude, $F(3,373)=9.61$, $p<0.0001$. Sarcasm and humor were produced with a significantly more restricted amplitude-range than sincerity, which in turn showed less amplitude variability than neutrality. A main effect of phrase type, $F(2,373)=171.62$, $p<0.0001$, could be explained by the fact that keyphrase exemplars displayed the most restricted amplitude-range and sentence exemplars showed less amplitude variation than combined sentence exemplars. No significant interactions were found.

C. Speech rate

Analysis of normalized speech rate yielded significant main effects for attitude, $F(3,373)=15.05$, $p<0.0001$, and phrase type, $F(2,373)=84.42$, $p<0.0001$, and a significant interaction of these factors, $F(6,373)=4.05$, $p=0.0006$. *Post hoc* tests on the interaction revealed that when producing sentences, sarcasm was expressed at a significantly slower speech rate than sincerity; also, neutrality was produced more slowly than humor and sincerity. For combined sentences, humor and sincerity were marked by a significantly faster speech rate than neutrality. There were no speech rate

TABLE IV. Mean normalized acoustic measures (and standard deviation) of Cantonese and English utterances expressing each of the four attitudes.

	Sarcasm		Sincerity		Humor		Neutrality	
Measure	Cantonese	English	Cantonese	English	Cantonese	English	Cantonese	English
Mean F0 (Hz)	0.47(1.25)	−0.45(0.72)	0.02(0.71)	0.42(1.05)	0.21(1.12)	0.34(1.17)	−0.68(0.55)	−0.49(0.48)
F0-range (Hz)	−0.13(1.09)	−0.05(1.02)	0.21(0.87)	0.25(0.91)	0.16(1.13)	0.09(0.96)	−0.36(0.84)	−0.44(0.99)
Mean amplitude (dB)	0.38(0.98)	0.14(1.24)	0.11(0.86)	0.22(0.72)	0.07(0.95)	0.33(0.72)	−0.61(1.00)	−0.65(0.89)
Amplitude-range (dB)	−0.38(1.05)	0.00(1.27)	0.13(0.84)	−0.10(0.90)	−0.25(1.06)	0.10(1.05)	0.38(0.94)	0.09(0.82)
Speech rate (syllables/s)	−0.67(0.76)	−0.80(0.99)	0.67(0.74)	0.46(0.84)	−0.11(1.11)	−0.03(1.00)	−0.36(0.79)	−0.02(0.72)
HNR (dB)	0.41(0.96)	−0.09(1.00)	−0.08(0.77)	−0.08(0.97)	−0.48(1.36)	0.18(0.96)	0.06(1.03)	0.10(1.06)

distinctions across the attitudes when speakers produced keyphrases.

D. HNR

The ANOVA on HNR produced a significant main effect of attitude, $F(3,365)=7.14$, $p=0.0001$, and a significant attitude by phrase type interaction, $F(6,365)=5.11$, $p=0.000\,05$.[3] For keyphrases only, humor exhibited significantly lower HNR values than utterances conveying all other attitudes. Both humorous keyphrases and humorous sentences also demonstrated significantly lower HNR values than humorous combined sentences.

E. Expressing sarcasm in Cantonese versus English

The second major goal was to highlight similarities and differences in the expression of sarcasm between languages by directly comparing the present acoustic data on Cantonese with published data on English (Cheang and Pell, 2008). This was possible as the Cantonese exemplars were devised, produced, validated, and measured acoustically in a manner identical to that of the prior study, although English exemplars needed to be renormalized in the same fashion as the Cantonese tokens to allow comparisons across data sets.[4] As noted earlier, phrase type was eliminated from these analyses by collapsing data along this factor to focus exclusively on the prosodic differences among attitudes; this manipulation is justified on the basis of the previous finding of global acoustic cues marking sarcasm in English (Cheang and Pell, 2008). The English and Cantonese data were entered into a series of ANOVAs involving the between-subjects factor of language (Cantonese and English) with repeated measures on attitude (sarcasm, sincerity, humor, and neutrality). Normalized data, which entered into the cross-language comparison, are furnished in Table IV (see Cheang and Pell, 2008 for the raw acoustic measures pertaining to English).

F. Fundamental frequency: Mean and range

Cross-language analysis of mean F0 data yielded a significant main effect of attitude, $F(3,866)=37.88$, $p<0.0001$, and an interaction of language by attitude, $F(3,866)=22.76$, $p<0.0001$. *Post hoc* elaboration of the interaction revealed that for Cantonese, sarcasm was produced with higher F0 levels than sincerity and neutrality (which was significantly lower than all attitudes). In contrast, English exemplars of sarcasm and neutrality exhibited significantly *lower* mean F0 than sincerity and humor. Across languages, sarcasm exhibited a significantly higher mean F0 in Cantonese than in English, and sincerity exhibited a significantly lower mean F0 in Cantonese than in English. There were no cross-linguistic differences in the mean F0 for neutral or humorous utterances (see Fig. 2).

The ANOVA of F0-range yielded a significant main effect of attitude, $F(3,866)=18.88$, $p<0.0001$. Irrespective of language, sarcasm was produced with a greater F0-range than neutral utterances, but with a narrower F0-range than sincere utterances. Neutrality exhibited the smallest F0-range, differentiating these utterances from all other attitudes. There was no interaction of attitude and language.

G. Amplitude: Mean and range

For mean amplitude, statistical analyses revealed a significant main effect of attitude, $F(3,865)=34.85$, $p<0.0001$, and a significant interaction of attitude and language, $F(3,865)=2.82$, $p=0.038$. The interaction was accounted for by the observation that neutral exemplars were produced with lower amplitude than all other attitudes in both languages. Also, there were differences between sarcasm and humor, which varied by language: In English, sarcasm was spoken with *lower* amplitude than humor, whereas in Cantonese sarcasm was spoken with *greater* amplitude than humor.

For amplitude-range, a significant main effect of attitude, $F(3,865)=6.48$, $p=0.0002$, and a significant language by attitude interaction, $F(3,865)=6.56$, $p=0.0002$, were found. In Cantonese, sarcasm was produced with a significantly restricted amplitude-range relative to neutral and sin-

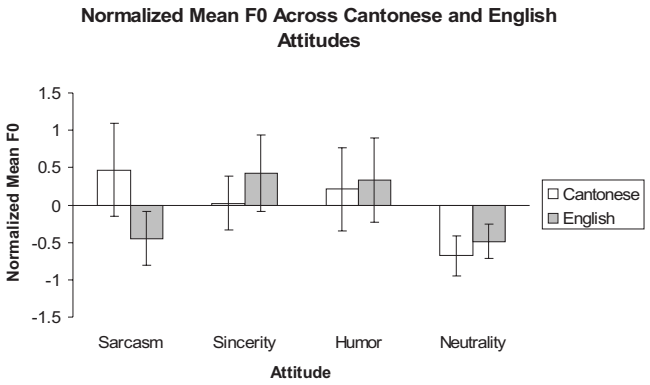


FIG. 2. Mean fundamental frequency (F0) of Cantonese versus English utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

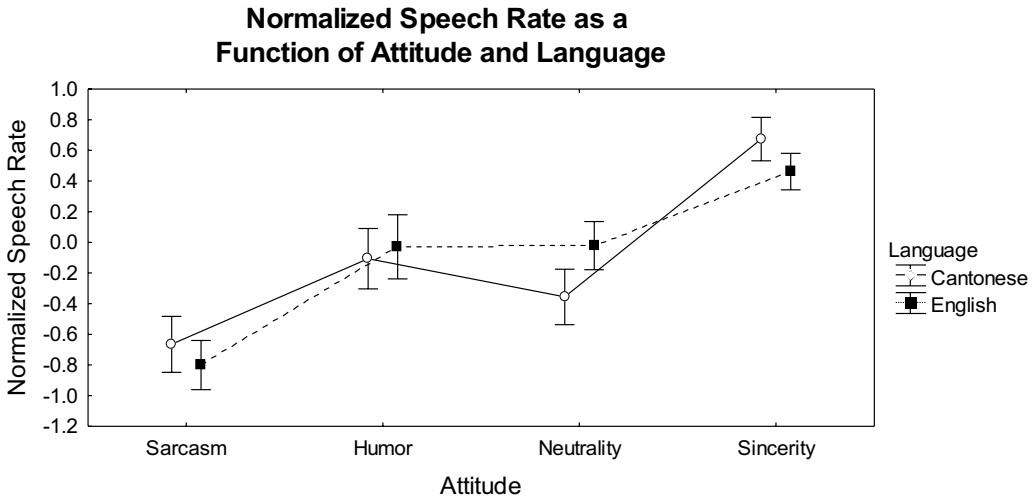


FIG. 3. Mean speech rate (syllables/s) of Cantonese versus English utterances conveying sarcasm, sincerity, humor, and neutrality. Error bars represent standard deviations.

cere utterances. Humor demonstrated a smaller amplitude-range than neutrality. English exemplars did not differ in amplitude-range as a function of attitude. Between languages, sarcastic utterances in Cantonese displayed a more restricted amplitude-range than in English.

H. Speech rate

Analyses of speech rate revealed a significant main effect of attitude, $F(3,866)=98.62$, $p<0.0001$, and an interaction between language and attitude, $F(3,866)=4.63$, $p=0.003$. In both Cantonese and English, sarcasm was produced with a significantly slower speech rate than all other attitudes (with the exception of neutrality in Cantonese). Sincerity was associated with the fastest articulation rate when compared to the other attitudes in both languages. In English, neutral utterances were spoken at a faster rate than humor, whereas in Cantonese neutral utterances were spoken more slowly than humor (see Fig. 3).

I. HNR

A significant main effect of attitude, $F(3,853)=3.71$, $p=0.01$, and a significant interaction of language and attitude, $F(3,853)=9.09$, $p<0.0001$, were found for measures of HNR. For Cantonese, the HNR of sarcastic utterances was significantly greater than that of sincere and humorous utterances, which in turn were greater than neutrality. There were no HNR differences as a function of attitude for English.[5] Between languages, Cantonese sarcasm displayed a higher HNR than English sarcasm, and Cantonese humor was produced with a lower HNR than in English.

IV. DISCUSSION

A. Overview

The current results provide novel data, which show that the expression of sarcasm in Cantonese speech is associated with a specific set of acoustic cues. Moreover, comparative analyses between languages show that the pattern in Cantonese is somewhat different from that used by English speakers to convey sarcasm (Cheang and Pell, 2008), although particular acoustic cues appear to share similar signaling functions between languages as discussed below. In Cantonese, the acoustic features of sarcastic utterances differentiated most clearly from sincere utterances: For this contrast, sarcasm displayed a higher mean F0, and narrower range in both F0 and amplitude, than utterances perceived as sincere in Cantonese speech. The data imply that Cantonese speakers encoded these particular features to convey sarcasm irrespective of the type/length of utterance produced (i.e., there was no influence of different phrase types on these acoustic variables). Other acoustic parameters were sometimes associated with sarcastic speech but only for specific phrase types; for example, sarcasm displayed higher HNR values (i.e., lower noise levels) than humor when keyphrase stimuli were produced.

Here, speaker mean F0 appeared to be the most important acoustic parameter for marking sarcasm. This finding is consistent with previous studies of sarcasm or verbal irony (Cutler, 1974, 1976; Haiman, 1998; Mueke, 1969, 1978; Myers-Roy, 1976; Rockwell, 2007, 2000a, 2005; Schaffer, 1982). This observation also fits the broader literature, which argues that mean F0 is a pivotal cue for signaling a variety of affective and attitudinal states across cultures and languages (see Banse and Scherer, 1996, for an overview). However, there were cross-language differences in how mean F0 was used to convey sarcasm; in Cantonese, mean F0 tended to be raised, whereas this parameter was lowered in the same context in English. Interestingly, Cantonese demonstrates similar patterns to Italian and French, which have both been associated with elevated mean F0 levels when speakers of these languages express sarcasm (Anolli *et al.*, 2002; Laval and Bert-Erboul, 2005). Collectively, the data suggest that the manner in which speakers exploit mean F0 to communicate sarcasm across languages is dictated to a considerable extent by social conventions. This differentiates how speakers use F0 to mark social intentions in spoken language, such as sarcasm, from how they use F0 to express basic emotions

such as fear, which demonstrate similar tendencies in speech irrespective of the speaker's language or culture (Pell *et al.*, 2009).

The finding that sarcasm is produced with less amplitude variation (range) in Cantonese, and frequently with a slower speech rate, coincides with data on English and other languages such as Japanese, Italian, and French (e.g., Adachi, 1996; Anolli *et al.*, 2002; Bryant and Fox Tree, 2005; Laval and Bert-Erboul, 2005; Rockwell, 2000a). It has been suggested that a reduced speech rate has the effect of drawing listener focus to a particular excerpt of discourse for a number of communicative purposes (Haiman, 1998; Kreuz and Roberts, 1995). The present results contribute to the argument that speakers of many languages use reduced speech rate to alert the listener to the intended, sarcastic meaning of utterances. It is possible that restrictions of amplitude variability serve a similar goal of the speaker as reduced speech rate to focus the listener to the fact that a non-literal message is intended. These suggestions could be usefully tested in future studies and in other contexts in which non-literal meanings are marked through prosody.

HNR was a significant cue, which differentiated expressions of sarcasm versus humor in certain contexts (when keyphrases were produced), indicating that speakers may modulate their voice quality at times to communicate sarcasm or irony. Attempts to convey a humorous attitude were associated with reliably greater amounts of noise (lower mean HNR values) when compared to analogous exemplars of sarcasm. Possibly, HNR fluctuations reflect changes in facial gestures, which occur during the expression of a playful intent; for example, smiling can be audibly detected by listeners (Auberge and Cathiard, 2003), although the mechanism through which this is accomplished is not yet fully specified (Tartter and Braun, 1994). Alternately, speakers may purposely deviate from their normal registers in attempting to convey humorous attitudes (Norrick, 2004), and this could introduce more noise into the speech signal. Interestingly, no other prosodic cues in these data clearly differentiated humorous utterances from the other attitudes overall; distinctions between humor and the other attitudes arose in unpredictable combinations of attitude and phrase type relative to sarcasm and sincerity. Nonetheless, data presented in Tables II–IV indicate that the two separate types of verbal irony targeted in this study show distinctions and should therefore not be treated as a single context with a unitary expression (Gibbs, 2000).

B. Distinctions between sarcasm and related attitudes

When expressions of sarcasm in Cantonese and English are compared to other attitudes, the clearest acoustic distinctions that emerged in both languages were between sarcasm and sincerity. This systematic distinction hints at a common communicative principle operating in both Cantonese and English: Sarcastic speech tends to be marked by a "play" cue, or a (meta-)message that signals to the listener that the speaker does not mean what they say, which is encoded by particular acoustic changes (Haiman, 1998). In contrast, sincere declarations are statements that are meant to reinforce the speaker's intended, literal meaning (even if implicature is required in some cases). Sincere utterances occur more frequently than sarcastic ones, and one can speculate that this context represents the "unmarked" mode of expression for most speakers. In the absence of linguistic cues to understand speaker intentions, it would be essential for sincere utterances to be highly distinct from (sarcastic) utterances that contain the play cue, which appears to be the case in both of the languages studied here.

When compared to sarcasm and sincerity, sarcasm and humor exhibited relatively few acoustic differences. This may be due to the fact that the sarcastic and humorous utterances evaluated here both represent instances of subtypes of verbal irony that share conceptual features (such as requiring an extra play cue to be recognized, Gibbs, 2000; Haiman, 1998). Rather than differences in acoustic features, contextual factors may be more important for recognizing humorous intent when compared to sarcasm. It has been argued that following the introduction of a play cue, a speaker must create a surprising incongruity in discourse and cohesively resolve it in order to successfully convey various forms of humor, ironic or otherwise (Berger, 1987; Berlyne, 1972; Brownell *et al.*, 1983; Cunningham and Derks, 2005; McGhee, 1976; Shultz and Horibe, 1974; Shultz, 1972; Suls, 1983; Wicker *et al.*, 1980, 1981). In the validation study, it may have been more difficult for decoders to appreciate the intended humorous intent of utterances because they did not have a context for interpreting humor. This likelihood is suggested by the relatively small number of humor exemplars retained for acoustic analysis (review Table II). Nonetheless, the present results do capture some of the possible acoustic correlates of humor, since many of these tokens were perceived correctly and subjected to acoustic analyses.

Finally, it was obvious that neutral utterances were acoustically distinct in nearly every respect from the other three attitudes. This point is important because many previous studies have simply divided utterances into the categories of "sarcastic (or ironic)" and "non-sarcastic (or non-ironic)," without particular attention to whether the comparison tokens were examples of neutrality or sincerity (e.g., Bryant, 2007; Bryant and Fox Tree, 2002; Rockwell, 2000a). The fact that previous studies may have inadvertently conflated neutral and sincere utterances as a baseline for characterizing sarcasm may underlie some of the disparities in the acoustic literature on sarcasm.

C. Issues of language

As reported previously for English (Cheang and Pell, 2008), the current data on Cantonese indicate that phrase type plays an important role in how prosody is used to convey speaker attitudes, especially for sarcasm and sincerity. A review of Table II shows that the presence of Cantonese keyphrases greatly increased the likelihood of decoders identifying sarcasm, whereas the absence of such phrases biased a sincere interpretation. This finding confirms that certain phrases or words can be largely associated with sarcastic meaning in both English and Cantonese (Haiman, 1998; Matthews and Yip, 1994). The present results may be a start-

ing point from which to detail the range of semantic terms that predispose sarcastic intentions in future studies, although they do not negate the fact that speakers use prosody in isolation to express sarcasm.

In the case of tone languages such as Cantonese, suggestions have been made that the set of acoustic cues available for expressing global meanings through prosody may be restricted in tonal languages due to the important role of pitch for signaling lexico-semantic information (Ross *et al.*, 1986). This notion was not borne out in the current study. Despite certain cross-language differences in the directionality of acoustic cues used to mark attitudes in Cantonese and English, speakers of both languages tended to exploit the *same* prosodic markers to convey attitudes (particularly mean F0). This finding suggests that local pitch phenomena in Cantonese do not hinder the use of F0 for other signaling functions over longer time domains (Vance, 1976).

D. Future directions

While great efforts were made to perceptually validate the present materials prior to acoustic analysis, the present findings are somewhat limited by the small number of items used and the small number of encoders who produced exemplars of the four attitudes. The generalizability of the present findings may be also somewhat restricted given that the present acoustic analyses were based on non-spontaneous speech. Since the encoders were intentionally given no explicit instructions as to how to acoustically pattern their speech productions, the encoders may have attempted to produce stylized "folk models" of sarcastic speech in addition to simulations of their personal style of articulating sarcasm in natural discourse. As such, the present results may suggest that speakers have less variable modulation of the individual acoustic components of sarcastic prosody than would be naturally observed in spontaneous conversation. Regardless, recognizably sarcastic prosody must have been adequately emulated by the encoders in the speech tokens, given the converging identification rates of a majority of naïve decoders during validation. Furthermore, speakers do use a diverse array of cues to signal sarcasm in natural speech, including highly stylized speech (Haiman, 1998). Therefore, although the current acoustic analyses are based on non-spontaneous tokens, the findings should reflect aspects of sarcastic prosody that occur in typical discourse.

Nonetheless, future studies would benefit from analyzing spontaneous excerpts of sarcasm elicited from more encoders; this should mitigate the relatively high number of recordings which were excluded from the present analyses and provide an even more comprehensive profile of sarcastic prosody. The nature of excluded tokens may also be of interest in future work: Since there appears to be a certain flexibility in how speakers communicate sarcasm (Haiman, 1998), analyzing tokens that are ambiguous or poorly recognized as sarcasm may be of value for revealing additional cues tied to this context. Future research should also consider an even larger set of acoustic parameters, which may be associated with sarcastic speech; some potential measures to examine include the number and length of pauses, different voice registers, and resonance changes (see Haiman, 1998 for an overview). The role of non-verbal cues such as facial expressions, and how these cues interact with prosody in the context of sarcasm, is also another avenue to explore (cf. Rockwell, 2000b, 2001, 2005).

Finally, given the differences and similarities noted here in how sarcasm is encoded in Cantonese versus English, these data raise the question of whether sarcasm can be detected from prosody in the speaker's non-native language. There is evidence that widely disparate cultures base their recognition of attitudes or emotions on culturally-defined factors such as overall communication style, at least in part (Elfenbein and Ambady, 2002, 2003; Kitayama and Ishii, 2002). Given the observed language-related differences in the direction of mean F0 change, which cues sarcasm in Cantonese versus English, one can speculate that listeners of one language would experience difficulties using cues in the other language correctly to infer when speakers intend to be sarcastic. This hypothesis is being tested by presenting both the Cantonese and English exemplars to monolingual speakers of each language (Cheang and Pell, in preparation).

ACKNOWLEDGMENTS

Support for this work was provided by doctoral training scholarships awarded to H.S.C. in the form of a CIHR-K.M. Hunter Doctoral Training Award from the Canadian Institutes of Health Research (CIHR) and a Bridge Funding Award from the Centre for Research on Language, Mind and Brain (CRLMB), an interdisciplinary unit in McGill University; and a Discovery Grant from the National Sciences and Engineering Research Council of Canada (NSERC) to M.D.P. The authors would also like to thank Elmira Chan and Marie Desmarteau for assistance in data collection.

[1]Cantonese speakers may add particles (that are constrained only by having to be in utterance-final position) that freely vary (i.e., can be paired with any context) to utterances in addition to (or instead of) modifying prosodic cues over the course of a sentence to help convey various attitudes or serve grammatical functions such as specifying utterance mode (Matthews and Yip, 1994). Inclusion of such particles in the present materials would limit comparability with earlier English materials devised by the current authors as there are no English equivalents. Perhaps, more importantly, acoustic variation may be (but not necessarily) concentrated disproportionately in particles for the purposes of attitude expression when such particles are present (Chan, 2001; Matthews and Yip, 1994). Such possible interactions between particles and prosody are not the present focus and, hence, particles were excluded from the ends of sentence exemplars. It should be noted that keyphrases could not be translated without the inclusion of particles, although particle inclusion here was justified as the keyphrases conveyed meaning as a unit rather than having the bulk of meaning expressed disproportionately by the particle.

[2]In Cheang and Pell, 2008, both F0 SD and F0-range were included as measures of F0 variation whereas the current study only included F0-range. F0 SD may be influenced by the very rapid lexical tonal shifts in Cantonese vowels (Bauer, 1998; Bauer and Benedict, 1997; Fok, 1974; Khouw and Ciocca, 2007; Vance, 1976, 1977) rather than from attitude expression. To avoid this confound F0 SD was dropped from consideration.

[3]The PRAAT program could not derive acoustic measurements for several tokens, hence the somewhat discrepant degrees of freedom reported across acoustic analyses.

[4]The current Cantonese utterances were subjected to *z*-score standardization. By contrast, all individual acoustic data points (from separate English utterances) were normalized per encoder in reference to the set of neutral exemplars spoken by that encoder in the authors' previous study {e.g.,

[mean F0(one keyphrase exemplar of humor) − mean F0(all keyphrase exemplars of neutrality)]/mean F0(all keyphrase exemplars of neutrality)} (Cheang and Pell, 2008). This change in strategy was adopted because naive Cantonese raters identified nearly none of the current utterances produced by one Cantonese encoder as conveying neutrality. However, the results (with respect to the acoustic cues of sarcasm for English) are very comparable across studies, suggesting that the prosodic features that emerged in the previous study were not spurious.

[5]The present lack of significant HNR differences as a function of attitude type in the English exemplars is at odds with the earlier finding of significantly reduced HNR values in English sarcasm (Cheang and Pell, 2008). The discrepancy stems from the fact that using the previous method of normalization, $p=0.046$ for the HNR attitude effect. Given the divergence, the renormalized English exemplars were analyzed in isolation using the same factors as in the ANOVAs of Cantonese tokens. Significant main effects of attitude for mean F0, significant effects of attitude and phrase type for F0 standard deviation, a significant effect of phrase type for mean amplitude and amplitude-range, and a significant attitude by phrase type interaction for speech rate were found. These effects were all attained in previous analyses of the English tokens normalized in reference to the neutral exemplars, and *post hoc* analyses yielded highly comparable results (Cheang and Pell, 2008). The great similarity in results for the analysis of the other acoustic parameters suggests that standardization was comparable across studies; the previous effect of HNR was genuine, albeit statistically weaker.

Ackerman, B. (**1983**). "Form and function in children's understanding of ironic utterances," J. Exp. Child Psychol. **35**, 487–508.

Ackerman, B. (**1986**). "Children's sensitivity to comprehension failure in interpreting a nonliteral use of an utterance," Child Dev. **57**, 485–497.

Adachi, T. (**1996**). "Sarcasm in Japanese," Studies in Language **19**, 1–36.

Anolli, L., Ciceri, R., and Infantino, M. G. (**2002**). "From 'blame by praise' to 'praise by blame': Analysis of vocal patterns in ironic communication," Int. J. Psychol. **37**, 266–276.

Attardo, S., Eisterhold, J., Hay, J., and Poggi, I. (**2003**). "Multimodal markers of irony and sarcasm," Humor **16**, 243–260.

Auberge, V., and Cathiard, M. (**2003**). "Can we hear the prosody of smile?," Speech Commun. **40**, 87–97.

Banse, R., and Scherer, K. R. (**1996**). "Acoustic profiles in vocal emotion expression," J. Pers. Soc. Psychol. **70**, 614–636.

Bauer, R. S. (**1998**). "Hong Kong Cantonese tone contours," in *Studies in Cantonese Linguistics*, edited by S. Matthews (Linguistic Society of Hong Kong, Hong Kong), pp. 1–34.

Bauer, R. S., and Benedict, P. K. (**1997**). *Modern Cantonese Phonology* (Mouton de Gruyter, Berlin), Vol. **102**.

Berger, A. A. (**1987**). "Humor: An introduction," Am. Behav. Sci. **30**, 6–15.

Berlyne, D. E. (**1972**). "Humor and its kin," in *The Psychology of Humor*, edited by P. E. McGhee and J. H. Goldstein (Springer-Verlag, New York), pp. 43–60.

Boersma, P., and Weenink, D. (**2006**). PRAAT: Doing phonetics by computer, version. 4.5.16, http://www.praat.org/ (Last viewed 6/15/2006).

Bond, M. H. (**1991**). *Beyond the Chinese Face* (Oxford University Press, Oxford).

Brownell, H. H., Michel, D., Powelson, J., and Gardner, H. (**1983**). "Surprise but not coherence: Sensitivity to verbal humor in right-hemisphere patients," Brain Lang **18**, 20–27.

Bryant, G. A. (**2009**). "Prosodic contrasts in ironic speech," Discourse Processes (in press).

Bryant, G. A., and Fox Tree, J. E. (**2002**). "Recognizing verbal irony in spontaneous speech," Metaphor and Symbolic Activity **17**, 99–117.

Bryant, G. A., and Fox Tree, J. E. (**2005**). "Is there an ironic tone of voice?," Lang Speech **48**, 257–277.

Capelli, C. A., Nakagawa, N., and Madden, C. M. (**1990**). "How children understand sarcasm: The role of context and intonation," Child Dev. **61**, 1824–1841.

Chan, M. (**2001**). "Gender-related use of sentence-final particles in Cantonese," in *Gender Across Languages: The Linguistic Representation of Women and Men*, edited by M. Hellinger and H. Bussmann (John Benjamins, Amsterdam), pp. 57–72.

Cheang, H. S., and Pell, M. D. (**2008**). "The sound of sarcasm," Speech Commun. **50**, 366–381.

Cunningham, W. A., and Derks, P. (**2005**). "Humor appreciation and latency of comprehension," Humor **18**, 389–403.

Cutler, A. (**1974**). "On saying what you mean without meaning what you say," in papers from the *Tenth Regional Meeting of the Chicago Linguistic Society*, edited by M. W. LaGaly, R. A. Fox, and A. Bruck (Chicago Linguistic Society, Chicago, IL), pp. 117–127.

Cutler, A. (**1976**). Beyond Parsing and Lexical Look-UP: An Enriched Description of Auditory Sentence Comprehension, in *New Approaches to Language Mechanisms*, edited by R. J. Wales and E. Walker (North-Holland, New York), pp. 133–149.

Elfenbein, H. A., and Ambady, N. (**2002**). "On the universality and cultural specificity of emotion recognition: A meta-analysis," Psychol. Bull. **128**, 203–235.

Elfenbein, H. A., and Ambady, N. (**2003**). "Cultural similarity's consequences: A distance perspective on cross-cultural differences in emotion recognition," J. Cross Cult. Psychol. **34**, 92–110.

Fok, C. Y.-Y. (**1974**). *A Perceptual Study of Tones in Cantonese* (Centre of Asian Studies, University of Hong Kong, Hong Kong), Vol. **18**.

Gibbs, R. W., Jr. (**2000**). "Irony in talk among friends," Metaphor Symb. **15**, 5–27.

Grabe, E., Rosner, B. S., Garcia-Albea, J. E., and Zhou, X. (**2003**). "Perception of English intonation by English, Spanish, and Chinese listeners," Lang Speech **46**, 375–401.

Haiman, J. (**1998**). *Talk is Cheap: Sarcasm, Alienation, and the Evolution of Language* (Oxford University Press, Oxford).

Ho, D. Y. F. (**1986**). "Chinese patterns of socialization: A critical review," in *The Psychology of the Chinese People*, edited by M. H. Bond (Oxford University Press, New York), pp. 1–37.

Huang, P. P.-F., and Kok, G. P. (**1999**). *Speak Cantonese*, 3rd ed. (Far Eastern, Detroit, MI), Vol. **1**.

Khouw, E., and Ciocca, V. (**2007**). "Perceptual correlates of Cantonese tones," J. Phonetics **35**, 104–117.

Killingley, S.-Y. (**1986**). "Normal and deviant classifier usage in Cantonese," Anthropological Linguistics **28**, 321–336.

Kitayama, S., and Ishii, K. (**2002**). "Word and voice: Spontaneous attention to emotional utterances in two languages," Cognit. Emotion **16**, 29–59.

Kreuz, R. J., and Roberts, R. M. (**1995**). "Two cues for verbal irony: Hyperbole and the ironic tone of voice," Metaphor and Symbolic Activity **10**, 21–31.

Kwok, H. H., and Luke, K. K. (**1986**). "Intonation in Cantonese: A preliminary study," paper presented at the *19th International Conference on Sino-Tibetan Languages and Linguistics* (Ohio State University, Columbus, OH).

Laval, V., and Bert-Erboul, A. (**2005**). "French-speaking children's understanding of sarcasm: The role of intonation and context," J. Speech Lang. Hear. Res. **48**, 610–620.

Matthews, S., and Yip, V. (**1994**). *Cantonese: A Comprehensive Grammar* (Routledge, London).

McCluskey, K. W., Albas, D. C., Niemi, R. R., Cuevas, C., and Ferrer, C. A. (**1975**). "Cross-cultural differences in the perception of the emotional content of speech: A study of the developmental of sensitivity in Canadian and Mexican children," Dev. Psychol. **11**, 551–555.

McGhee, P. E. (**1976**). "Children's appreciation of humor: A test of the cognitive congruency principle," Child Dev. **47**, 420–426.

Muecke, D. C. (**1969**). *The Compass of Irony* (Methuen, London).

Muecke, D. C. (**1978**). "Irony markers," Poetics Today **7**, 363–375.

Myers-Roy, A. (**1976**). "Towards a definition of irony," in *Studies in Language Variation*, edited by R. W. Fasold and R. Shuy (Georgetown University Press, Washington, DC), pp. 171–183.

Norrick, N. R. (**2004**). "Non-verbal humor and joke performance," Humor **17**, 401–409.

Pell, M. D., Monetta, L., Paulmann, S., and Kotz, S. A. (**2009**). "Recognizing emotions in a foreign language," J. Nonverbal Behav. **33**, 107–120.

Rockwell, P. (**2000a**). "Lower, slower, louder: Vocal cues of sarcasm," J. Psycholinguist. Res. **29**, 483–495.

Rockwell, P. (**2000b**). "Actors', partners', and observers' perceptions of sarcasm," Percept. Mot. Skills **91**, 665–668.

Rockwell, P. (**2001**). "Facial expression and sarcasm," Percept. Mot. Skills **93**, 47–50.

Rockwell, P. (**2005**). "Sarcasm on television talk shows: Determining speaker intent through verbal and nonverbal cues," in *Psychology of Moods*, edited by A. Clark (Nova Science, New York), pp. 109–140.

Rockwell, P. (**2007**). "Vocal features of conversational sarcasm: A comparison of methods," J. Psycholinguist. Res. **36**, 361–369.

Ross, E. D., Edmondson, J. A., and Seibert, G. B. (**1986**). "The effect of affect on various acoustic measures of prosody in tone and non-tone languages: A comparison based on computer analysis of voice," J. Phonetics

14, 283–302.
Schaffer, R. (**1982**). "Are there consistent vocal clues for irony?," in *Parasession on Language and Behavior*, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller (Chicago Linguistic Society, Chicago, IL), pp. 204–210.
Scherer, K. R. (**1986**). "Vocal affect expression," Psychol. Bull. **99**, 143–165.
Scherer, K. R., Ladd, D. R., and Silverman, K. E. (**1984**). "Vocal cues to speaker affect: Testing two models," J. Acoust. Soc. Am. **76**, 1346–1356.
Shultz, T. R. (**1972**). "The role of incongruity and resolution in children's appreciation of cartoon humor," J. Exp. Child Psychol. **13**, 456–477.
Schultz, T. R., and Horibe, F. (**1974**). "Development of the appreciation of verbal jokes," Dev. Psychol. **10**, 13–20.
Snow, D. (**2004**). *Cantonese as Written Language: The Growth of a Written Chinese Vernacular* (Hong Kong University Press, Hong Kong).
Suls, J. (**1983**). "Cognitive processes in humor appreciation," in *Handbook of Humor Research*, edited by P. E. McGhee and J. H. Goldstein (Springer-Verlag, New York), pp. 39–57.
Tartter, V. C., and Braun, D. (**1994**). "Hearing smiles and frowns in normal and whisper registers," J. Acoust. Soc. Am. **96**, 2101–2107.
Vance, T. J. (**1976**). "An experimental investigation of tone and intonation in Cantonese," Phonetica **33**, 368–392.
Vance, T. J. (**1977**). "Tonal distinctions in Cantonese," Phonetica **34**, 93–107.
Wicker, F. W., Barron, W. L. I., and Willis, A. C. (**1980**). "Disparagement humor: Dispositions and resolutions," J. Pers. Soc. Psychol. **39**, 701–709.
Wicker, F. W., Thorelli, I. M., Barron, W. L. I., and Ponder, M. R. (**1981**). "Relationships among affective and cognitive factors in humor," J. Res. Pers. **15**, 359–370.
Winner, E., and Leekman, S. (**1991**). "Distinguishing irony from deception: Understanding the speaker's second-order intention," Br. J. Dev. Psychol. **9**, 257–270.
Yumoto, E., Gould, W. J., and Baer, T. (**1982**). "Harmonics-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. **71**, 1544–1550.

Production and perception of French vowels by congenitally blind adults and sighted adults

Lucie Ménard[a] and Sophie Dupont
Département de Linguistique et de Didactique des Langues, Laboratoire de Phonétique, Center for Research on Language, Mind, and Brain, Université du Québec à Montréal, Montreal, Quebec H3C 3P8, Canada

Shari R. Baum
School of Communication Sciences and Disorders, Center for Research on Language, Mind, and Brain, McGill University, Montreal, Quebec H3G 1A8, Canada

Jérôme Aubin
Département de Linguistique et de Didactique des Langues, Laboratoire de Phonétique, Center for Research on Language, Mind, and Brain, Université du Québec à Montréal, Montreal, Quebec H3C 3P8, Canada

(Received 6 August 2008; revised 18 May 2009; accepted 30 May 2009)

The goal of this study is to investigate the production and perception of French vowels by blind and sighted speakers. 12 blind adults and 12 sighted adults served as subjects. The auditory-perceptual abilities of each subject were evaluated by discrimination tests (AXB). At the production level, ten repetitions of the ten French oral vowels were recorded. Formant values and fundamental frequency values were extracted from the acoustic signal. Measures of contrasts between vowel categories were computed and compared for each feature (height, place of articulation, roundedness) and group (blind, sighted). The results reveal a significant effect of group (blind vs sighted) on production, with sighted speakers producing vowels that are spaced further apart in the vowel space than those of blind speakers. A group effect emerged for a subset of the perceptual contrasts examined, with blind speakers having higher peak discrimination scores than sighted speakers. Results suggest an important role of visual input in determining speech goals. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3158930]

PACS number(s): 43.70.Mn, 43.71.Es [RSN] Pages: 1406–1414

I. INTRODUCTION

In the past decades, several studies have shown that visual cues provided by the lips and jaw are not simply redundant in the process of speech perception; in fact, they act as functional cues that supplement the auditory information transmitted by the speech signal (McGurk and MacDonald, 1976; Robert-Ribes *et al.*, 1998). In the audiovisual modality, speech intelligibility scores are higher than in the auditory modality alone or in the visual modality alone (Grant *et al.*, 1998; Reisberg *et al.*, 1987). The role played by visual cues in speech perception is crucial for perceivers without access to auditory input (Andersson *et al.*, 2001; Bernstein *et al.*, 2001). However, perceptual cues conveyed by the visual channel alone do not allow the listener to recover all the phonological contrasts of a language, as revealed by the fact that prelingually deaf speakers without hearing aids never fully gain the ability to perceive speech on the basis of speechreading alone (cf. Bernstein *et al.*, 2000).

Although the visual modality is crucial for deaf speakers, the fact that congenitally blind speakers learn to produce correct speech sounds suggests that visual cues are not mandatory in the control of speech movements. Nevertheless, several studies conducted with blind speakers revealed that their speech discrimination abilities differ from those of sighted speakers (Lucas, 1984; Hugdahl *et al.*, 2004; Gougoux *et al.*, 2004). This difference in auditory discrimination abilities between the groups may reflect differences at the level of production, given that the ability to discriminate speech sounds has been suggested to be related to individual differences in the amount of articulatory-acoustic contrast produced between the two sounds (Perkell *et al.*, 2004, for instance). The links between production and perception have been evidenced in hearing and sighted subjects. For example, in 11 subjects, speaker's judgments of synthetic vowel similarities were correlated with that speaker's produced formant values for corner vowels. Newman (2003) also found production-perception relationships in subjects' produced and perceived voicing onset times values. Bell-Berti *et al.* (1979) and Perkell *et al.* (2004) reported electromyographic (EMG) (for the former) and articulatory data (for the latter) significantly related to the subjects' perceptual abilities. Other studies have shown parallel changes in production and perception dimensions, such as Bradlow *et al.* (1997), Rvachew (1994), and Vick *et al.* (2001) for cochlear implant subjects.

Furthermore, apart from differences in discrimination abilities between congenitally blind speakers and sighted speakers, the lack of access to visual information might also

[a]Author to whom correspondence should be addressed. Electronic mail: menard.lucie@uqam.ca

induce differences in the use and/or control of the speech articulators (especially the visible ones). To the best of our knowledge, no study has addressed speech production abilities in adult speakers with visual impairments.

This paper aims to investigate auditory discrimination abilities and the production of vowel contrasts in 12 congenitally blind adults and 12 sighted adults, all native speakers of Canadian French. Vowels were chosen because they are perceptually salient in the speech stream and tend to yield relatively consistent percepts.

II. AUDITORY PERCEPTION IN BLIND SPEAKERS

Without visual information, blind speakers rely solely on the auditory signal to recover phonological information. In a review of studies conducted on blind and sighted speakers between 1960 and 1980, Miller (1992) showed that studies have produced somewhat contradictory results regarding auditory acuity in the two speaker groups. Stankov and Spilsbury (1978), for instance, studied rhythm perception and frequency discrimination in music and speech in clear and distorted (background noise or reduced tempo) conditions in 30 young speakers (between 10 and 15 years of age) belonging to three groups: totally blind, partially blind, and sighted. Blind speakers performed better than sighted speakers in frequency discrimination tasks, but no difference was found between the two groups in speech identification tasks in distorted conditions. Starlinger and Niemeyer (1981) and Niemeyer and Starlinger (1981) also conducted a series of perceptual experiments on blind and sighted adults. They found no difference in frequency discrimination thresholds, intensity discrimination thresholds, or duration discrimination thresholds. Despite the fact that those low-level tasks were performed equally well by both groups, in higher-level identification tasks, such as binaural integration of pure tones and noise, the blind speakers performed significantly better than the sighted speakers. Lucas (1984), Hugdahl *et al.* (2004), and Gougoux *et al.* (2004) also found superior non-speech auditory perceptual abilities in blind speakers. It should be mentioned, however, that the blind speaker groups in these studies were heterogeneous in many respects, for instance, concerning speaker age, age at blindness, degree of blindness, etc. Such variability is confounded with visual impairment and could have greatly influenced the results.

III. SPEECH PRODUCTION IN BLIND SPEAKERS

As reported by Kuhl and Meltzoff (1982) and Legerstee (1990), by the age of 4 months, sighted babies demonstrate strong capacities to associate sounds with the corresponding visual representation of the lips. Babies also imitate labial movements of sounds visually presented. It seems that at this language acquisition stage, babies recognize relationships between auditory parameters and visual events. Although most of the studies addressing auditory perceptual abilities in blind speakers have been conducted with adult subjects, speech production has been mainly described for blind children. As Elstner (1983) stated, visual impairment deprives the child of an important source of information that may have consequences for the strategies used to produce phonological targets. At the pre-babbling stage, Lewis (1975) reported less imitation of labial speech gestures by a blind baby compared to sighted babies. Elstner (1983) and Mills (1987) presented various studies showing phonological delays and phonetic/phonological disorders in older children. In a study of syllables produced by a small number of congenitally blind children (1–2 years of age), Mills (1987) reported a higher number of phonological confusions between groups of visually dissimilar consonants (labial /b/ vs velar /k/) for the blind children compared to sighted children. These data must, however, be interpreted with caution since they come from a very small sample. Furthermore, as reported by Elstner (1983), it is difficult to study homogeneous populations of blind speakers, and observed differences in speech production abilities between blind and sighted groups might just as well be related to the presence of uncontrolled variables, such as additional motor control disorders or language disorders, unrelated to the visual impairment.

In perhaps the most directly relevant study, Göllesz (1972) collected EMG data from 13-year-old and 14-year-old blind Hungarian male speakers uttering vowels. Sighted control subjects were also recorded. Despite reduced labial dynamics in blind speakers compared to sighted speakers, as measured by the degree of EMG activation, no significant differences were observed in the acoustic signal. These results suggest that visual impairment leads speakers to adopt different control strategies for the visible labial articulators. Some compensatory abilities of the other articulators are also likely involved to offset the limited movements of the lips to reach the acoustic target.

The objectives of the present study are the following. First, auditory discrimination abilities along the three phonological contrasts in French oral vowels (height, place of articulation, and roundedness) are investigated in 12 congenitally blind adults and 12 sighted adults. Second, the production of the French oral vowels by both groups of speakers in the acoustic space in terms of between-category contrast distances is studied. Third, production-perception relationships are analyzed through multiple regression analyses.

IV. METHOD

A. Participants

12 congenitally blind adults (6 males and 6 females) and 12 sighted adult control subjects (6 males and 6 females) participated in the study. All subjects were native speakers of Canadian French living in the Montreal area. (Although the majority had some exposure to English, all use French as their primary language.) The blind speakers had a congenital and complete visual impairment, classified as class 3, 4, or 5 in the International Disease Classification of the World Health Organization. They had never had any perception of light or movement. They ranged in age from 26 to 52 years (mean: 44). They did not demonstrate any language disorders or motor deficits by self-report. Table I presents pertinent characteristics of the blind speakers. Twelve sighted adult subjects were also recorded and formed the control group. They all had perfect vision (20/20) or impaired vision cor-

TABLE I. Characteristics of the 12 blind speakers.

Subject	Gender	Age	Etiology of blindness	Vision at birth	Current vision
DM	F	48	Retinitis pigmentosa	U[a]	R.E.[b] =3/210 L.E.[c] =0
FB	F	40	Congential cataract	U	R.E.=0 L.E.=6/1260
SS	F	26	U	U	U (total blindness)
CP	M	52	Optic atrophy	Total blindness	R.E.=0 L.E.=0
SN	M	40	Detachment of the retina	U	R.E.=2/180 L.E.=2/105
YL	M	42	Congential cataract et Congential glaucoma	U	U (total blindness)
MAR	M	36	Retinitis pigmentosa	Total blindness	R.E.=20/400 L.E.=20/400
AB	M	52	Congential cataract	Total blindness	R.E.=3/180 L.E.=2/180
IM	F	51	Retinitis pigmentosa	Total blindness	R.E.=2/400 L.E.=2/400
FM	F	45	Congential cataract	Total blindness	U (total blindness)
JL	F	52	Retinitis pigmentosa	U	U (total blindness)
MD	M	42	Retinitis pigmentosa	Total blindness	U (total blindness)

[a]Undetermind.
[b]Right eye.
[c]Left eye.

rected by lenses, resulting in near-perfect vision (according to self-report). The control subjects ranged in age from 22 to 39 years (mean: 33). Despite the mean age difference between the groups, it is unlikely to influence the results, as small age-related changes in perception and production that may exist tend to emerge at more advanced ages. Moreover, as will be seen, in this instance, the older (blind) group ultimately demonstrates more accurate auditory discrimination scores. All subjects passed a 20-dB HL pure-tone screening at 500, 1000, 2000, 4000, and 8000 Hz.

B. Experiment I: Perception

Five sets of vowels (including the first five formants) ranging from /i/ to /e/, /e/ to /ɛ/, /ɛ/ to /a/, /y/ to /u/, and /i/ to /y/ (all phonemically contrastive) were synthesized using the variable linear articulatory model (Boë and Maeda, 1997), which is based on Maeda's model (Maeda, 1979). Whereas one might suggest that the use of synthetic stimuli does not adequately reflect natural speech processing, it represents the most appropriate means of controlling the precise acoustic differences across the stimuli. The five continua corresponded to the three phonological features along which French oral vowels are produced: height (/i/ vs /e/, /e/ vs /ɛ/, and /ɛ/ vs /a/), place of articulation (/y/ vs /u/), and rounding (/i/ vs /y/). Formant values of the end-point stimuli for each of the three continua, listed in Table II, were those used in previous perceptual studies with similar synthesized stimuli (Ménard *et al.*, 2002; Ménard *et al.*, 2004). Formant bandwidths for the five formants were calculated based on an analog simulation (Badin and Fant, 1984). Several versions of the five continua were created based on different steps between adjacent stimuli. Those stimuli were submitted to two native Canadian-French-speaking judges in order to determine the version that would maximally avoid ceiling effects while yielding scores above chance level (good discrimination functions). For the /i/ vs /e/ continuum, five stimuli were created between the end-points at equally stepped F1 (0.22 bark/20.1 mel), F2 (0.08 bark/7.4 mel), and F3 (0.34 bark/30.7 mel) distances. For the /e/ vs /ɛ/ continuum, five stimuli were also synthesized; F1 values be-

TABLE II. Formant (F_i) and bandwidth (B_i) values, in hertz, of end-point stimuli /i/, /e/, /y/, and /u/ synthesized for the perceptual experiment.

Vowel	F1	F2	F3	F4	F5	B1	B2	B3	B4	B5
/i/	236	2062	3372	3466	5000	78	13	61	154	154
/e/	372	1918	2501	3466	5000	78	13	61	154	154
/ɛ/	492	1676	2445	3610	5000	48	40	148	67	67
/a/	711	1234	2311	3695	5000	37	57	71	98	98
/y/	236	1757	2062	3294	5000	88	40	19	19	19
/u/	236	705	2062	3294	5000	88	40	19	19	19

tween stimuli differed by 0.18 bark (16.5 mel), and F2 values varied by 0.15 bark (13.8 mel). Regarding the /ɛ/ vs /a/ contrast, eight stimuli were created by varying F1 and F2 in equal steps (0.19 bark/17.4 mel and 0.22 bark/20.1 mel). As a result, seven stimuli (including end-points) were created for the /i/-/e/ dimension, seven for the /e/-/ɛ/ continuum, and ten for the /ɛ/-/a/ continuum. The /y/-/u/ continuum was represented by 22 stimuli, spaced in F2 by 0.26 bark (23.6 mel). The rounding continuum, corresponding to the /i/-/y/ dimension, was represented by seven stimuli, equally stepped in F2 (0.18 bark/16.5 mel) and F3 (0.52 bark/46.3 mel). A cascade formant synthesizer was excited by a glottal waveform generated by the Liljencrants-Fant source model. The resulting signal was digitized at 22 kHz and was 600 ms long. A fall-rise amplitude contour was applied to the signal. The F0 values were 110 Hz.

Stimuli from the five continua were presented to each of the subjects in a discrimination task. A classic AXB design was used, with an interstimulus interval of 500 ms. Stimuli were grouped in triads where the first and the third were one step apart on the synthesized continuum, and the second was the same as either the first stimulus or the third one. After each triad was played, the subject had to decide whether the second stimulus was the same as the first or the third. Each triad was also presented in BXA form, where the order of the first and third stimuli was reversed. Each triad was repeated twice, in each order (AXB and BXA), yielding a total of four repetitions for a given pair of stimuli. All stimuli were randomized across listeners.

C. Experiment II: Production

1. Procedure

Each participant in the auditory discrimination test also served as a subject in a production task. Ten repetitions of the ten French oral vowels /i y u e ø o ɛ œ ɔ a/ were elicited from each speaker, in random order, in the following context: "V comme WORD" ("V as in WORD"), where V is one of the ten vowels mentioned above, and WORD is a French word with this vowel in initial position. Only the initial isolated, long, and sustained V was analyzed (not the V produced in the word context). All speakers repeated the sequence after hearing an adult speaker utter it. For the sighted group, no visual input was provided. The speech signals were recorded in a sound booth with a high-quality tabletop microphone (Shure SM-84) at a 15- to 20-cm distance from the subject's lips and digitized at 44 100 Hz by a digital audio tape recorder (DAT TASCAM DA-P1). Signals were then downsampled to 22 050 Hz after low-pass filtering (cutoff frequency of 10 000 Hz). For each of the ten vowel repetitions, the first three formant frequencies were then extracted for each vowel, using the Linear Predictive Coding algorithm integrated in the PRAAT speech analysis program (Boersma and Weenink, 2007). The number of poles varied from 10 to 14 in the range of parameters used by Lee *et al.* (1999) and Hillenbrand *et al.* (1995). A 14-ms Hamming window was used with a pre-emphasis factor of 0.98 (pre-emphasis from 50 Hz for a sampling frequency of 22 050 Hz). Formant measurement errors were detected by comparing, for each vowel, the automatically extracted formant values overlaid on a wide-band spectrogram with a spectral slice obtained by an Fast Fourier Transform analysis with a Hamming window. When major discrepancies were observed either (i) between the overlaid formant values and the spectrogram or (ii) between the overlaid formant values and the spectral slice, the prediction order of the automatic detection algorithm was readjusted and the analysis was performed again. Fundamental frequency values (F0) were extracted using an autocorrelation algorithm. The formant frequencies were then converted to the mel scale because this scale models the ear's integration of frequency according to the following formula: $F_{\mathrm{mel}}=550\ln(1+F_{\mathrm{Hz}}/550)$.

2. Data analysis

At the perceptual level, for each triad, the number of correct responses, referred to as the discrimination score, was computed. The highest discrimination score obtained for the triads of a given continuum will be referred to as the peak discrimination score. In such a perceptual task, this score reflects auditory discrimination abilities at the category boundary between two stimuli. Repeated-measures Analysis of Variance (ANOVAs) were then carried out with peak discrimination scores as the dependent variable, vowel continuum (/i/ vs /e/, /e/ vs /ɛ/, /ɛ/ vs /a/, /i/ vs /y/, or /y/vs /u/) as the within-subject factor, and subject group (blind or sighted) as the between-subject factor. Since the scores were relatively high, with data ranging from 62.5% to 100%, ceiling effects were obtained, resulting in right-skewed distributions. Scores were thus transformed into logarithmic-based scales.

At the production level, the stimuli produced were represented in the traditional F1 vs F2 vs F3 space in mels. This three-dimensional space was used rather than the two-dimensional F1 vs F2 space to account for possible shifts in formant-cavity affiliations across subjects, yielding greater contrast in the F2 vs F3 space between two vowel categories, than in the F1 vs F2 space. This is the case, for instance, for the /i/ vs /y/ rounding contrast in French (Schwartz *et al.*, 1993).

Following Lane *et al.* (2001) and Ménard *et al.* (2007), among others, measures of contrast distances between vowel categories were computed. Those parameters have already been used in studies of speech produced by cochlear implant users as measures of produced contrasts (Lane *et al.*, 2001; Ménard *et al.*, 2007). In such studies, for a given vowel contrast, it is assumed that greater contrast distance between vowels reflects greater control and precision in the ability to produce this vowel contrast.

For each participant, Euclidean distances were first calculated between the mean F1, F2, and F3 values in mels (for each of the vowels) for all possible pairs of vowels in the acoustic space. Euclidean distances are more appropriate for cross-speaker comparisons than raw F1, F2, and F3 data since the latter are closely related to vocal-tract morphology and are speaker dependent. The first dependent variable at the production level consisted of the Euclidean distances for the vowels corresponding to the stimuli used in the perceptual task (/i/ vs /e/, /e/ vs /ɛ/, /ɛ/ vs /a/, /i/ vs /y/, and /y/ vs

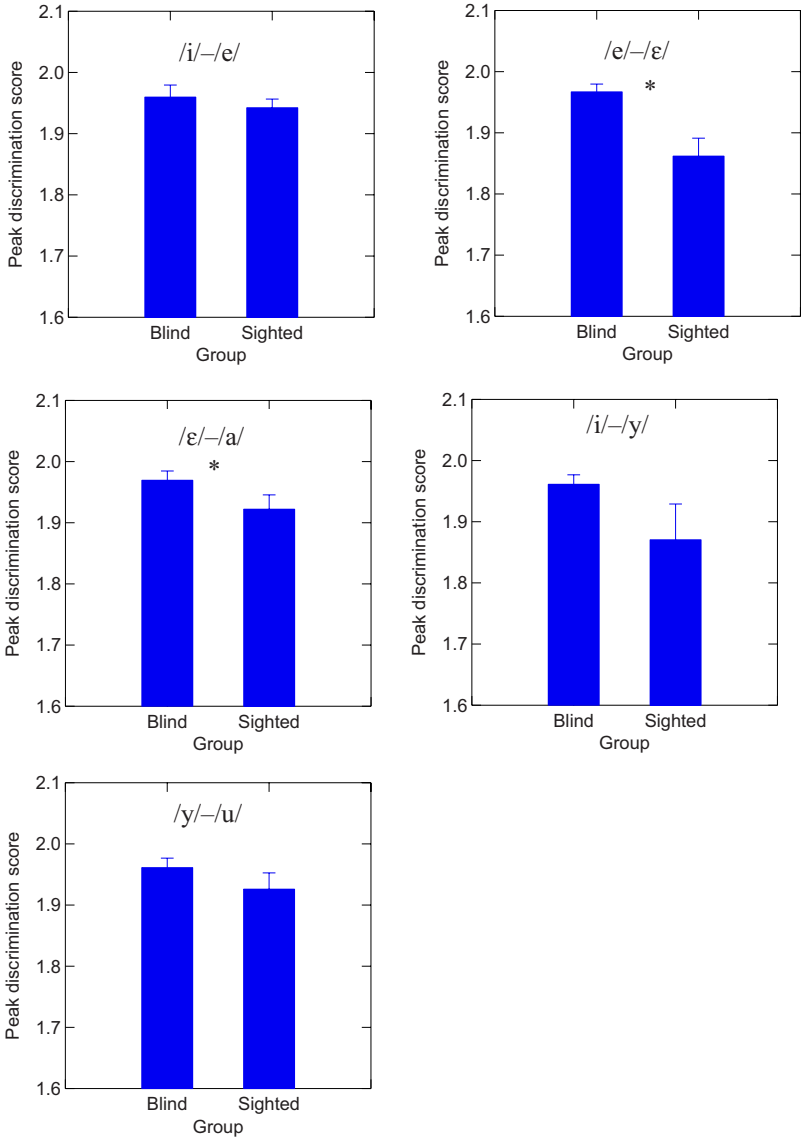


FIG. 1. (Color online) Average peak discrimination scores for both speaker groups for the five vowel contrasts: /i/-/e/ (upper left panel),/e/-/ɛ/ (upper right panel), /ɛ/-/a/ (middle left panel), /i/-/y/ (middle right panel), and /y/-/u/ (lower left panel). Error bars are one standard error of the mean.

/u/). The second dependent variable was the average vowel spacing (AVS) (Lane *et al.*, 2001) defined as the average of all Euclidean distances (including those between vowel pairs not used as stimuli at the perceptual level). Unlike produced Euclidean distances, AVS provides a global measure of produced vowel contrasts. Repeated-measures ANOVAs were then carried out on the data with subject group (blind or sighted) as the between-subject factor. Vowel continuum (/i/ vs /e/, /e/ vs /ɛ/, /ɛ/ vs /a/, /i/ vs /y/, and /y/ vs /u/) was the within-subject factor for the first dependent variable (Euclidean distances between vowels).

To further investigate the link between production and perception, multiple regression analyses were performed. For each of the five vowel continua, 12 data points (one for each speaker) were used. The dependent variable was produced Euclidean distance, in mels, and the independent variables were peak discrimination scores and speaker group.

V. RESULTS

A. Perception

Average peak discrimination scores for the three continua related to vowel height, rounding, and place of articulation for sighted and blind speakers are plotted in Fig. 1 (/i/-/e/: upper left panel, /e/-/ɛ/: upper right panel, /ɛ/-/a/: middle left panel, /i/-/y/: middle right panel, and /y/-/u/: lower left panel). As Fig. 1 shows, all participants had good discrimination acuity, as revealed by the rather high average values for the peak discrimination scores. A repeated-measures ANOVA with peak discrimination scores as the dependent variable, speaker group (sighted or blind) as the between-subject variable, and vowel contrast (/i/-/e/, /e/-/ɛ/, /ɛ/-/a/, /i/-/y/, /y/-/u/) as the within-subject variable did not reveal any significant main effects of speaker group or vowel contrast. However, a significant interaction of speaker group and vowel contrast was found $[F(4,88)=2.51; p<0.05]$. *Post hoc* tests (Tukey) showed that blind speakers had significantly higher peak discrimination scores than sighted speakers for the /e/-/ɛ/ contrast $[F(1,22)=15.60; p<0.05]$ as well as for the /ɛ/-/a/ contrast $[F(1,22)=5.12; p<0.05]$. The difference in peak discrimination scores for the /i/-/y/ continuum did not reach significance ($p<0.07$) but the observed pattern is similar to the significant one noted for the /e/-/ɛ/ and /ɛ/-/a/ contrasts, with blind speakers having higher scores than sighted speakers. A closer examination of the

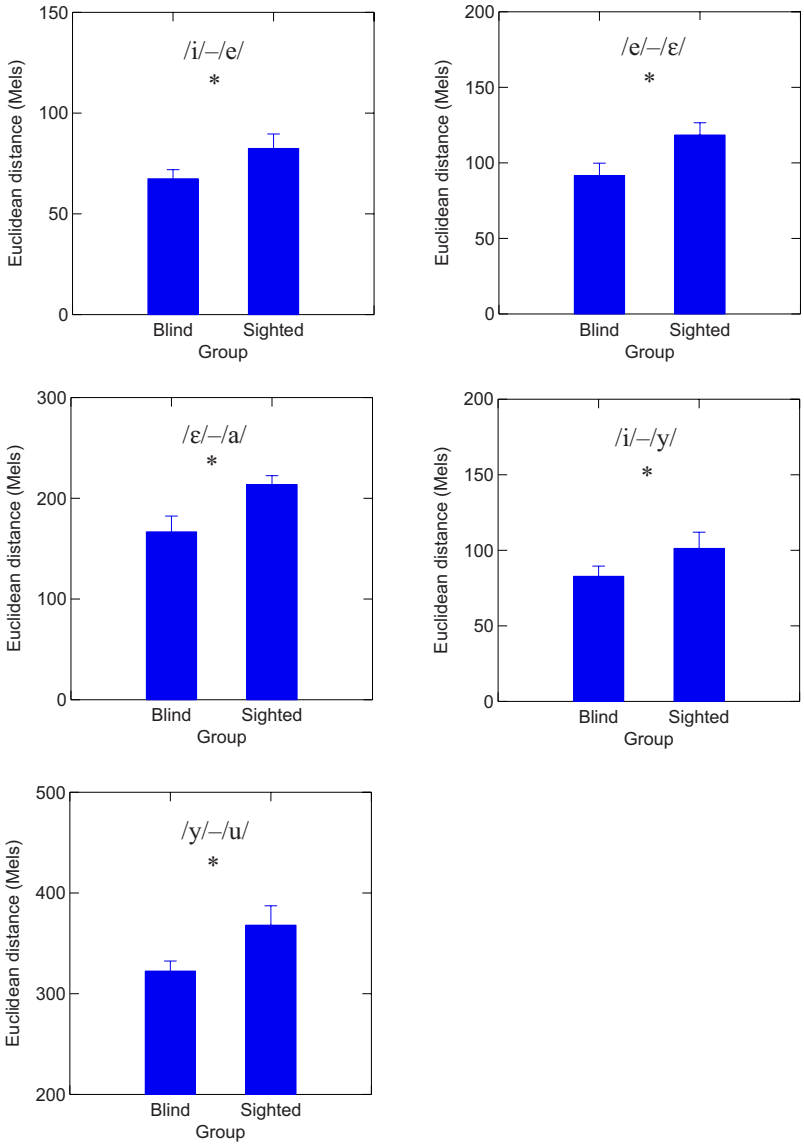


FIG. 2. (Color online) Average produced Euclidean distances for both speaker groups for the five vowel contrasts: /i/-/e/ (upper left panel),/e/-/ɛ/ (upper right panel), /ɛ/-/a/ (middle left panel), /i/-/y/ (middle right panel), and /y/-/u/ (lower left panel). Error bars are one standard error of the mean.

data revealed that standard deviation values (as suggested by the size of the error bars in Fig. 1) are higher for sighted subjects than for blind subjects. This pattern is mainly due to the perfect discrimination score (100%) of two sighted subjects, the remaining ten subjects having peak discrimination scores close to 80%. A ceiling effect is probably involved here, preventing the tendency toward higher discrimination scores for blind than sighted subjects from reaching significance.

B. Production

The average Euclidean distances measured for the /i/-/e/, /e/-/ɛ/, /ɛ/-/a/, /i/-/y/, and /y/-/u/ produced contrasts are plotted in Fig. 2. Error bars represent one standard error of the mean. A repeated-measures ANOVA with Euclidean distance as the dependent variable, speaker group (sighted or blind) as the between-subject variable, and vowel contrast (/i/-/e/, /e/-/ɛ/, /ɛ/-/a/, /i/-/y/, /y/-/u/) as the within-subject variable revealed a significant main effect of speaker group, with blind subjects having smaller contrast distances than sighted subjects $[F(1,22)=14.33;\ p<0.05]$. A significant effect of vowel contrast was also observed, with the distances between /y/ and /u/ being greater than the distances between the other vowel pairs $[F(4,88)=12.43;\ p<0.05]$. This result reflects the organization of the French vowel space, the rounded back vowel /u/ having no unrounded counterpart (in contrast to the front rounded /y/ vs the front unrounded /i/). The analysis did not reveal any significant effect of the interaction between the group variable and the vowel contrast variable.

Figure 3 plots the average AVS values for sighted speakers and blind speakers. A one-way ANOVA performed on this data set with speaker group (sighted or blind) as the between-subject factor revealed a significant effect of speaker group $[F(1,20)=6.20;p<0.05]$. Sighted speakers produced larger contrast distances between vowel categories (AVS) than blind speakers.

C. Relations between production and perception

For each of the five vowel continua, in Fig. 4, data are plotted in "production-perception" space (/i/-/e/: upper left panel, /e/-/ɛ/: upper right panel, /ɛ/-/a/: middle left panel,

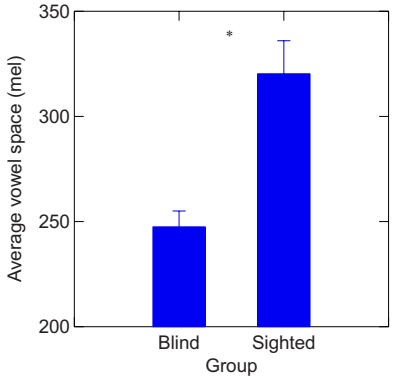


FIG. 3. (Color online) AVS in the F1 vs F2 vs F3 space, in mels, for both speaker groups.

/i/-/y/: middle right panel, and /y/-/u/: lower left panel). The production axis (x-axis) corresponds to the produced contrast distance (Euclidean distance), in mels, for each speaker. The perception dimension (y-axis) is the corresponding value of the peak discrimination score for each speaker. As a result, 12 data points are represented within each subject group (blind and sighted) and for each vowel continuum. Results of multiple regression analyses are presented in Table III. Beta weights, given for each of the independent variables, are interpretable in terms of magnitude of influence of a variable on the produced contrast distance. As shown in Table III, the

TABLE III. Values of beta weights (B) from multiple regression analyses for each vowel contrast. Dependent variables: produced Euclidean distances; independent variables: subject group and peak discrimination scores (* =significant at $p<0.05$).

Contrast	B group	B peak discrimination score
/i/-/e/	0.35	−0.06
/e/-/ɛ/	0.73*	0.46*
/ɛ/-/a/	0.57*	0.27
/i/-/y/	0.21	−0.27
/y/-/u/	0.44*	0.07

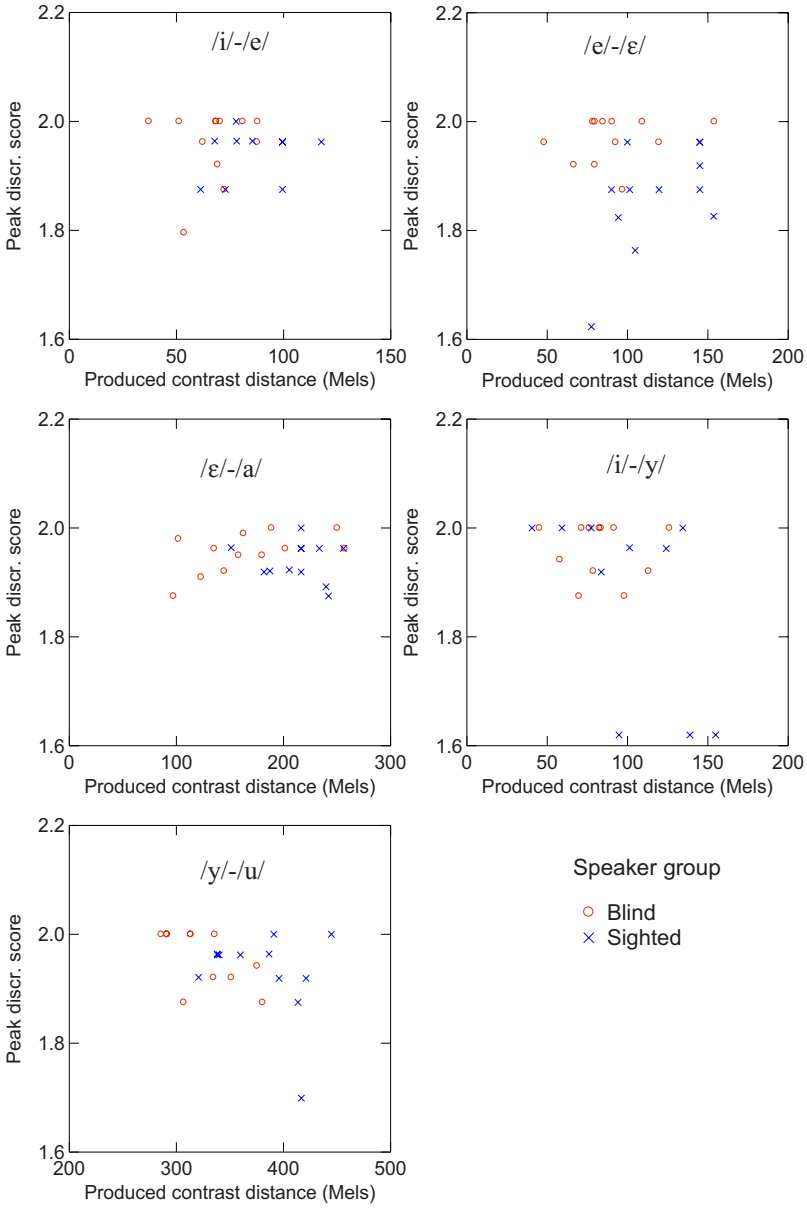


FIG. 4. (Color online) Average produced Euclidean distances in relation with peak discrimination scores for both speaker groups for the five vowel contrasts: /i/-/e/ (upper left panel), /e/-/ɛ/ (upper right panel), /ɛ/-/a/ (middle left panel), /i/-/y/ (middle right panel), and /y/-/u/ (lower left panel). Results of multiple regression analyses are shown in Table III.

effect of the group variable is significant for /e/-/ɛ/, /ɛ/-/a/, and /y/-/u/. The effect of peak discrimination score is significant only for the /e/-/ɛ/ contrast, and its beta weight is lower than the one calculated for the group variable. These results suggest that lack of visual input has a stronger influence on the production of the vowel contrasts than does perceptual acuity for those contrasts, although the two co-vary to some degree.

VI. DISCUSSION

In order to assess the role of visual deprivation on auditory perception and production of French vowels, discrimination tasks and acoustic recordings were conducted with congenitally blind subjects and sighted control subjects. Significant effects of speaker groups were found in both tasks (although for only a subset of vowel discrimination contrasts), confirming the importance of the role of visual experience in speech perception and production.

A. Auditory perception and produced contrast distances

At the perceptual level, the results of these experiments showed that congenitally blind adult speakers have more accurate auditory discrimination abilities than sighted adult speakers for some French oral vowels. Indeed, in AXB discrimination tasks performed on synthesized continua, blind speakers had significantly higher peak discrimination scores than sighted speakers for two continua (/e/-/ɛ/ and /ɛ/-/a/), and the same tendency almost reached significance for a third continuum (/i/-/y/). The fact that discrimination scores were higher only for two out of five continua is likely due to a ceiling effect. This result confirms those of earlier studies showing that blind speakers have better auditory acuity than sighted speakers (Lucas, 1984; Hugdahl *et al.*, 2004; Gougoux *et al.*, 2004; Doucet *et al.*, 2005). Those contrasts are related to both height and rounding features, two dimensions that are highly associated with visual correlates at the perceptual level in French. Perhaps blind listeners are more attuned to the acoustic properties of these contrasts because they cannot rely on the additional visual cues. Although it is possible that the blind listeners had more experience with synthetic speech in their lifetimes, given the high accuracy levels of both groups and the fact that our findings are consistent with previous investigations, it is unlikely that such experience (if present) contributed significantly to the results.

At the production level, contrast distances, measured by the value of AVS, were significantly higher for sighted speakers than for blind speakers. According to Perkell *et al.* (2004), speakers who are better able to discriminate auditorily between phoneme categories will tend to produce larger contrast distances between categories. Thus, blind speakers, who have more accurate auditory discrimination abilities, would be expected to produce larger AVS values. Interestingly, the opposite pattern was observed, suggesting that the effects of absence of visual feedback on vowel contrasts may be larger than the effects of auditory acuity. [Of course, additional unrelated factors, such as differences in language acquisition and socialization, educational environment, etc., may also play a role (e.g., Andersen *et al.*, 1993).] This result is somewhat similar to that reported by Perkell and colleagues (Perkell *et al.*, 2004; Ménard *et al.*, 2007) on postlingually deaf speakers of American English with cochlear implants. Although several differences exist between the two sets of studies, in the Perkell group's studies, absence of auditory feedback yielded reduced contrast distances between categories in acoustic space, as revealed by reduced AVS values. In the present study, absence of visual feedback from birth was found to lead to similar results.

The present results indicate that the absence of visual input may contribute to more accurate auditory discrimination scores, suggesting that the internal phonemic sensory goals may be more distinct (see, e.g., Guenther *et al.*, 2006). However, the contrast distances (AVS) measured in the production task were smaller for speakers with visual impairments compared to sighted speakers, suggesting that visual cues play an important part in shaping speech goals.

The findings are consistent with recent behavioral and neuroimaging investigations which have supported a close link between perception and production (e.g., Fadiga *et al.*, 2002; Watkins *et al.*, 2003; Wilson *et al.*, 2004; Sams *et al.*, 2005; Pulvermuller *et al.*, 2006; Gentilucci and Bernardis, 2007; Meister *et al.*, 2007; Skipper *et al.*, 2007; Tourville *et al.*, 2008). That is, despite the absence of a strong statistical relationship between the auditory-perceptual and production data, we interpret the results to suggest an important link between the perceptual representation (developed on the basis of both auditory and visual cues) and production patterns. Further studies conducted with a greater number of subjects will seek to investigate whether auditory acuity in blind speakers is specifically related to the amount of acoustic contrasts produced by those speakers.

ACKNOWLEDGMENTS

This work was supported by the Social Sciences and Humanities Research Council of Canada and the Natural Sciences and Engineering Research Council of Canada. We are grateful to Pascal Perrier for useful comments on earlier versions of this paper. Thanks to Zofia Laubitz for copy-editing the paper.

Andersen, E., Dunlea, A., and Kekelis, L. (**1993**). "The impact of input: Language acquisition in the visually impaired," First Lang. **13**, 23–50.

Andersson, U., Lyxell, B., Rönnberg, J., and Spens, K. E. (**2001**). "Cognitive correlates of visual speech understanding in hearing-impaired individuals," Journal of Deaf Studies and Deaf Education **6**, 103–116.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," Speech Transm. Lab. Q. Prog. Status Rep. **2–3**, 53–108.

Bell-Berti, F., Raphael, L. J., Pisoni, D. B., and Sawusch, J. R. (**1979**). "Some relationships between speech production and perception," Phonetica **36**, 373–383.

Bernstein, L. E., Auer, E. T., Jr., and Tucker, P. E. (**2001**). "Enhanced speechreading in deaf adults: Can short-term training/practice close the gap for hearing adults¿," J. Speech Lang. Hear. Res. **44**, 5–18.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62**, 233–252.

Boë, L.-J., and Maeda, S. (**1997**). "Modélisation de la croissance du conduit vocal. Espace vocalique des nouveaux-nés et des adultes. Conséquences pour l'ontogenèse et la phylogenèse," Journées d'Études Linguistiques: "La Voyelle dans Tous ces États," **1**, 98–105.

Boersma, P., and Weenink, D. (**2007**). PRAAT, Version 4.4.07, www.praat.org (Last viewed February, 2007).
Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.
Doucet, M.-E., Guillemot, J.-P., Lassonde, M., Gagné, J.-P., Leclerc, C., and Lepore, F. (**2005**). "Blind subjects process auditory spectral cues more efficiently than sighted individuals," Exp. Brain Res. **160**, 194–202.
Elstner, W. (**1983**). "Abnormalities in the verbal communication of visually-impaired children," in *Language Acquisition in the Blind Child*, edited by A. E. Mills (Croom Helm, London), pp. 18–41.
Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (**2002**). "Speech listening specifically modulates the excitability of tongue muscles: A TMS study," Eur. J. Neurosci. **15**, 399–402.
Gentilucci, M., and Bernardis, P. (**2007**). "Imitation during phoneme production," Neuropsychologia **45**, 608–615.
Göllesz, V. (**1972**). "Über die lippenartikulation der von geburt an blinden" ("About the lip articulation of the blind from birth"), in *Papers in Interdisciplinary Speech Research. Speech Symposium*, edited by S. Hirschberg, G. Y. Szépe, and E. Vass-Kovoics (Akadémiai Kiado, Budapest), pp. 85–91.
Gougoux, F., Lepore, F., Lassonde, M., Voss, P., Zatorre, R. J., and Belin, P. (**2004**). "Pitch discrimination in the early blind," Nature (London) **430**, 309.
Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.
Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang **96**, 280–301.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
Hugdahl, K., Ek, M., Rintee, T., Tuomainen, J., Haaral, C., and Hämäläinen, K. (**2004**). "Blind individuals show enhanced perceptual and attentional sensitivity for identification of speech sounds," Brain Res. Cognit. Brain Res. **19**, 28–32.
Kuhl, P. K., and Meltzoff, A. N. (**1982**). "The bimodal perception of speech in infancy," Science **218**, 1138–1141.
Lane, H., Matthies, M., Perkell, J., Vick, J., and Zandipour, M. (**2001**). "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation," J. Speech Lang. Hear. Res. **44**, 552–563.
Lee, S., Potamianos, A., and Narayanan, S. (**1999**). "Acoustics of children's speech. Developmental changes of temporal and spectral parameters," J. Acoust. Soc. Am. **105**, 1455–1468.
Legerstee, M. (**1990**). "Infants use multimodal information to imitate speech sounds," Infant Behav. Dev. **13**, 343–354.
Lewis, M. M. (**1975**). *Infant Speech: A Study of the Beginnings of Language* (Arno, New York).
Lucas, S. A. (**1984**). "Auditory discrimination and speech production in the blind child," Int. J. Rehabil. Res. **7**, 74–76.
Maeda, S. (**1979**). "An articulatory model of the tongue based on a statistical analysis," J. Acoust. Soc. Am. **65**, S22.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (**2007**). "The essential role of premotor cortex in speech perception," Curr. Biol. **17**, 1692–1696.
Ménard, L., Polak, M., Denny, M., Lane, H., Matthies, M. L., Perkell, J. S., Burton, E., Marrone, N., Tiede, M., and Vick, J. (**2007**). "Interactions of speaking condition and auditory feedback on vowel production in postlingually deaf adults with cochlear implants," J. Acoust. Soc. Am. **121**, 3790–3801.
Ménard, L., Schwartz, J.-L., and Boë, L.-J. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized french vowels from birth to adulthood," J. Speech Lang. Hear. Res. **47**, 1059–1080.
Ménard, L., Schwartz, J.-L., Boë, L.-J., Kandel, S., and Vallée, N. (**2002**). "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood," J. Acoust. Soc. Am. **111**, 1892–1905.
Miller, J. (**1992**). "Diderot reconsidered: Visual impairment and auditory compensation," J. Vis. Impair. Blind. **86**, 206–210.
Mills, A. E. (**1987**). "The development of phonology in the blind child," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates, London), pp. 145–163.
Newman, R. S. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.
Niemeyer, W., and Starlinger, I. (**1981**). "Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. II. Central functions," Audiology **20**, 510–515.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338–2344.
Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (**2006**). "Motor cortex maps articulatory features of speech sounds," Proc. Natl. Acad. Sci. U.S.A. **103**, 7865–7870.
Reisberg, D., McLean, J., and Goldfield, A. (**1987**). "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli," in *Hearing by Eye: The Psychology of Lip-Reading*, edited by B. Dodd and R. Campbell (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 97–113.
Robert-Ribes, J., Schwartz, J. L., Lallouache, T., and Escudier, P. (**1998**). "Complementarity and synergy in bimodal speech: Auditory, visual and audiovisual identification of French oral vowels in noise," J. Acoust. Soc. Am. **103**, 3677–3689.
Rvachew, S. (**1994**). "Speech perception training can facilitate sound production learning," J. Speech Hear. Res. **37**, 347–357.
Sams, M., Mottonen, R., and Sihvonen, T. (**2005**). "Seeing and hearing others and oneself talk," Brain Res. Cognit. Brain Res. **23**, 429–435.
Schwartz, J.-L., Beautemps, D., Abry, C., and Escudier, P. (**1993**). "Inter-individual and cross-linguistic strategies for the production of the [i] vs. [y] contrast," J. Phonetics **21**, 411–425.
Skipper, J. I., Van Wassenhove, V., Nusbaum, H. C., and Small, S. L. (**2007**). "Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception," Cereb. Cortex **17**, 2387–2399.
Stankov, L., and Spilsbury, G. (**1978**). "The measurement of auditory abilities of blind, partially sighted, and sighted children," Appl. Psychol. Measure. **2**, 491–503.
Starlinger, I., and Niemeyer, W. (**1981**). "Do the blind hear better? Investigations on auditory processing in congenital or early acquired blindness. I. Peripheral functions," Audiology **20**, 503–509.
Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429–1443.
Vick, J., Lane, H., Perkell, J. S., Matthies, M. L., Gould, J., and Zandipour, M. (**2001**). "Covariation of cochlear implant users' perception and production of vowel contrasts and their identification by listeners with normal hearing," J. Speech Lang. Hear. Res. **44**, 1257–1267.
Watkins, K. E., Strafella, A. P., and Paus, T. (**2003**). "Seeing and hearing speech excites the motor system involved in speech production," Neuropsychologia **41**, 989–994.
Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (**2004**). "Listening to speech activates motor areas involved in speech production," Nat. Neurosci. **7**, 701–702.

Role of mask pattern in intelligibility of ideal binary-masked noisy speech

Ulrik Kjems,[a] Jesper B. Boldt, and Michael S. Pedersen
Oticon A/S, Kongebakken 9, DK-2765 Smørum, Denmark

Thomas Lunner
Oticon Research Centre Eriksholm, Kongevejen 243, DK-3070 Snekkersten, Denmark and Department of Clinical and Experimental Medicine, and Technical Audiology, Linköping University, S-58183 Linköping, Sweden

DeLiang Wang
Department of Computer Science and Engineering and Center for Cognitive Science, The Ohio State University, Columbus, Ohio 43210

(Received 11 September 2008; revised 9 May 2009; accepted 22 June 2009)

Intelligibility of ideal binary masked noisy speech was measured on a group of normal hearing individuals across mixture signal to noise ratio (SNR) levels, masker types, and local criteria for forming the binary mask. The binary mask is computed from time-frequency decompositions of target and masker signals using two different schemes: an ideal binary mask computed by thresholding the local SNR within time-frequency units and a target binary mask computed by comparing the local target energy against the long-term average speech spectrum. By depicting intelligibility scores as a function of the difference between mixture SNR and local SNR threshold, alignment of the performance curves is obtained for a large range of mixture SNR levels. Large intelligibility benefits are obtained for both sparse and dense binary masks. When an ideal mask is dense with many ones, the effect of changing mixture SNR level while fixing the mask is significant, whereas for more sparse masks the effect is small or insignificant.
 [DOI: 10.1121/1.3179673]

PACS number(s): 43.71.An, 43.71.Gv, 43.66.Ba, 43.66.Dc [MSS] Pages: 1415–1426

I. INTRODUCTION

The human ability to understand speech in a variety of adverse conditions is remarkable, and the underlying processes are not well understood. According to Bregman's auditory scene analysis account, the auditory system processes the acoustic input in two stages: an analysis and segmentation stage where the sound is decomposed into distinct time-frequency (T-F) segments followed by a grouping stage (Bregman, 1990; Wang and Brown, 2006). The grouping stage is divided into primitive grouping and schema driven grouping that represent bottom-up and top-down processes, respectively. Hence, in order to recognize speech in background noise, the auditory system would employ a combination of bottom-up processing of available cues, and top-down application of schemas, which represent learned patterns.

In this paper these processes are studied using the technique of ideal T-F segregation (ITFS), which was proposed by Brungart *et al.* (2006) to induce idealized grouping when listening to a mixture of target speech and noise. ITFS is based on the use of ideal binary mask (IBM), which was originally proposed as a benchmark for measuring the segregation performance of computational auditory scene analysis systems (Wang, 2005). The ITFS technique applies an IBM to the mixture, and several recent studies have utilized the technique for revealing important factors for speech intelligibility in noise (Brungart *et al.*, 2006; Anzalone *et al.*, 2006; Li and Loizou, 2008; Wang *et al.*, 2009).

A binary mask is defined in the T-F domain as a matrix of binary numbers. We refer to the basic elements of the T-F representation of a signal as T-F units. A frequency decomposition similar to the human ear can be achieved using a bank of gammatone filters (Patterson *et al.*, 1988), and signal energies are computed in time frames (Wang and Brown, 2006). The IBM is defined by comparing the signal-to-noise ratio within each T-F unit against a local criterion (LC) or threshold measured in units of decibels. Only the T-F units with local signal to noise ratio (SNR) exceeding LC are assigned 1 in the binary mask. Let $T(t,f)$ and $M(t,f)$ denote target and masker signal power measured in decibels, at time t and frequency f, respectively, the IBM is defined as

$$\mathrm{IBM}(t,f) = \begin{cases} 1 & \text{if } T(t,f) - M(t,f) > \mathrm{LC}, \\ 0 & \text{otherwise}. \end{cases} \quad (1)$$

An IBM segregated signal can be synthesized from the mixture by deriving a gain from the binary mask, and applying it to the mixture before recombination in a synthesis filter bank. However, not all studies follow the same procedure—sometimes the short-time Fourier transform is used (for instance Li and Loizou, 2008) which typically yields lower frequency resolution at low frequencies, but much higher resolution at high frequencies.

In Brungart *et al.*, 2006, the IBM was used as a means to retain the effect of energetic masking, thereby separating the

[a]Author to whom correspondence should be addressed. Electronic mail: uk@oticon.dk

energetic masking and informational masking effects. They argued that since the IBM removes those T-F units dominated by the masker, ITFS can be said to retain the effect of energetic masking, while removing informational masking caused by the excluded units with relatively significant masker energy. Informational masking refers to the inability to correctly segregate audible target information from the mixture. Their study showed a plateau of nearly perfect intelligibility of ITFS processed mixtures when varying the value of LC from −12 to 0 dB. Meanwhile, the IBM with 0 dB LC is considered to be the theoretically optimal mask out of all possible binary masks in terms of SNR gain (Li and Wang, 2009). Brungart *et al.* (2006) noted that lowering the mixture SNR by 1 dB while fixing LC causes the exact same T-F units to be left out as increasing the LC by 1 dB while fixing the mixture SNR; in other words, the IBM remains the same in these two scenarios. They demonstrated remarkably similar performance curves by altering the test conditions in the two ways described, which they interpret as rough equivalence in the effect of energetic masking.

Anzalone *et al.* (2006) showed large intelligibility benefits of IBM segregation and reported positive results on hearing impaired subjects, although their IBM definition is different from the previously outlined ITFS procedure. They computed the IBM by comparing the target signal to a fixed threshold adjusted to retain a certain percentage of the total target energy. Furthermore they attenuated the T-F units designated as non-target by 14 dB, in contrast to the total elimination described above. Their results showed more than 7 dB improvement in speech reception threshold (SRT) for normal hearing and more than 9 dB improvement for hearing impaired subjects.

In a study comparing impaired and normal-hearing subjects, Wang *et al.* (2009) demonstrated large improvements in SRT for both normal-hearing and hearing impaired groups due to ITFS processing of speech mixtures. Their study of the normal-hearing group shows an 11 dB improvement in SRT with a cafeteria noise masker containing conversational speech and an improvement of 7 dB for speech-shaped noise (SSN). For the hearing impaired group, the SRT improvement was 16 dB in cafeteria noise and 9 dB in SSN. As a surprising result, the SRTs obtained from the normal-hearing and hearing impaired groups on the ITFS processed mixtures were comparable.

Li and Loizou (2008) used short time Fourier transforms to apply ideal binary masking to mixtures with a two-talker masker, as well as modulated and unmodulated SSN maskers. They found large intelligibility benefits similar to Brungart *et al.* (2006) when varying the LC parameter, although they reported wider plateaus of LC values with almost perfect intelligibility (−20 to +5 dB compared to −12 to 0 dB in Brungart *et al.*, 2006), which they attributed to differences in speech material and filterbank setup. They further suggested that it may be the pattern of the binary mask itself that matters for intelligibility, rather than the local SNR of each T-F unit.

Wang *et al.* (2008) demonstrated that applying a binary pattern of gains obtained from an IBM with a SSN masker to the masker signal alone produces high intelligibility scores, a type of process related to noise vocoding (Dudley, 1939; Shannon *et al.*, 1995). Using different numbers of filterbank bands, they showed that intelligibility is lost when the number of channels is 8 or smaller, a result which differs from that reported by Shannon *et al.* (1995) who used continuous, rather than binary, values for envelope manipulation. There, high intelligibility was reported using noise vocoded in just four channels.

A. Motivation

The large benefits in intelligibility outlined previously could make the IBM a candidate for applications such as hearing aids, provided that the IBM can be approximated sufficiently well. In this paper we will not consider how such estimation might be done. However, to devise such applications it is important to understand the mechanisms by which the IBM enhances intelligibility. In the above described literature, much attention has been given to explaining intelligibility of IBM segregated mixtures by considering audibility of the target signal. By focusing on absolute regions of LC (Brungart *et al.*, 2006), emphasis is put on the interpretation that the IBM reduces informational masking by directing listeners' attention to the T-F units containing target information (Li and Loizou, 2008). This view is basically related to models of intelligibility based on target audibility in additive noise, such as the speech intelligibility index (ANSI, 1997), where intelligibility is described as a function of the proportion of target signal that is audible in different frequency bands. Cooke (2006) and Srinivasan and Wang (2008) proposed related computational models that operate on mixture input directly and produce recognition results from automatic speech recognition that are compatible with human intelligibility performance.

However, some of the previous published results seem inconsistent with this view. In particular, the observation of Wang *et al.* (2008) that IBM-processed noise is intelligible suggests that the resulting temporal envelope of the processed mixture is important. The speech transmission index (Houtgast and Steeneken, 1971) considers how distortions to the envelope affect speech intelligibility. Recent extensions have been made to improve the model predictions of nonlinearly processed speech (Goldsworthy and Greenberg, 2004). While the speech intelligibility index model cannot explain the noise gating results of Wang *et al.* (2008), a model based on speech transmission index described by Goldsworthy and Greenberg (2004) may perform better. This means that the target modulation carried by the IBM may play a key role in intelligibility of processed mixtures.

Based on the observation that the IBM is insensitive to the covariation of LC and mixture SNR, we propose to focus on the *difference* between the LC and the mixture SNR levels when comparing performance across mixture SNR levels. We therefore introduce a *relative criterion* (RC), defined as RC=LC−SNR in units of decibels.

By focusing on RC and varying the mixture SNR, it is possible to vary the effects of the target component of the IBM processed mixture relative to that of the masker. For example, by taking the mixture SNR to a large negative

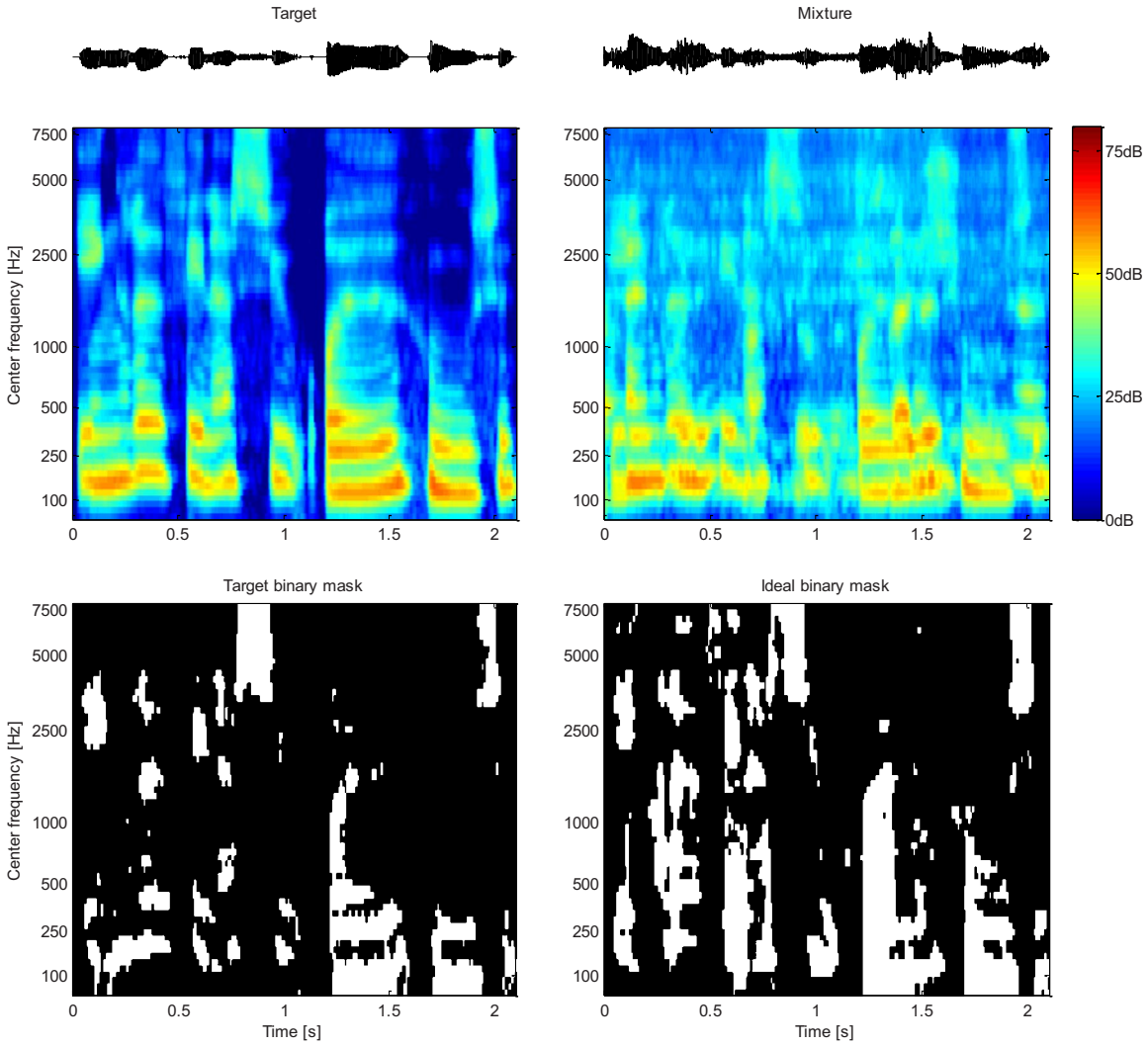


FIG. 1. (Color online) Illustration of IBM and TBM. Upper row shows waveform signal for a clean target sentence (left) and the sentence corrupted with cafeteria noise (right). Middle row shows the cochleagram representation of the two signals. Bottom left and right show the TBM and IBM, respectively, with white indicating the value of 1.

value, we can measure intelligibility of IBM-gated noise similar to Wang *et al.* (2008). On the other hand, by taking the mixture SNR to a level near the SRT we may measure how processing with the exact same binary mask affects intelligibility near the SRT.

B. Aims of the experiments

The change in focus from LC to RC brings up several research questions, which we will address in this paper. One aim is to investigate how the range of RC for optimal intelligibility depends on mixture SNR. Are there regions of RC where mixture SNR level has little or no effect on intelligibility? This question is directly addressed by the experiments in this paper. If under some circumstances mixture SNR level plays a minor role, the masker signal type may play a major role. So the second aim is to investigate the effects of masker type. We know that the plateau of optimal LC values is narrower for same-talker speech maskers (Brungart *et al.*, 2006) compared to a SSN masker. So far, intelligibility of IBM-processed noise has only been reported for stationary noise (Wang *et al.*, 2008).

Third, we wanted to compare the effects of alternative ideal mask definitions. The mask used by Anzalone *et al.* (2006) was computed based on the target signal alone; yet large intelligibility improvements were obtained. They define the target binary mask (TBM) as the one obtained by comparing, in each T-F unit, the target energy to that of a SSN reference signal matching the long-term spectrum of the target speaker. This comparison still uses the LC parameter as a SNR threshold. The binary mask that results from this process can then be applied to a mixture of the target and a *different* masker. Figure 1 illustrates an example TBM and IBM computed from a target sentence in cafeteria noise, and shows the differences between the resulting masks. The top row shows the time domain waveforms of the clean and noisy target sentences. The middle row shows cochleagrams of the clean and noisy target sentence using a filterbank of 1 ERB (equivalent rectangular bandwidth) wide gammatone filters with center frequencies from 55 Hz to 7.7 kHz. The bottom row shows the TBM (left) and IBM (right). The two masks are noticeably different. The TBM pattern resembles the target sentence and is unaffected by the specific masker.

On the other hand, the IBM pattern depends on the masker signal as well.

The TBM has several useful properties. The mask is, by this definition, identical to the IBM when SSN is the masker, so the TBM can be used as a measure of how general the IBM generated with the SSN masker is. Furthermore, relating to schema-based auditory scene analysis, the TBM could be interpreted as a simplified template of a learned pattern, indicating where in time and frequency to expect target energy. We therefore expect that the TBM leads to comparable benefits in intelligibility compared to the IBM. In some speech enhancement applications, it may be easier to estimate a TBM rather than an IBM, and it is therefore useful to know the extent to which the TBM results in intelligibility improvements.

The remainder of this paper is organized as follows. A listening experiment is described in Sec. II, and the results are reported and discussed in Sec. III. Section IV concludes the paper.

II. EXPERIMENTAL SETUP

A listening experiment was conducted to measure speech intelligibility of ITFS processed mixtures. The aim was to measure the influence of mixture SNR level, RC value, masker type, and to compare mask construction schemes: IBM and TBM.

A. Stimuli

The target phrases were from the Dantale II corpus (Wagener *et al.*, 2003) which is the Danish version of the Swedish Hagerman sentence (Hagerman, 1982) test and the German Oldenburg sentence test (Kollmeier and Wesselkamp, 1997). The corpus consists of 150 sentences designed to have low redundancy. The phrases were all spoken by the same female Danish speaker. The sentences were five words long following the same grammatical structure: name-verb-numeral-adjective-noun. An English translated example is "Michael had five new plants." Each word was randomly selected out of ten possibilities in each position of a sentence, taking coarticulation into account (Wagener *et al.*, 2003). Since long-term spectral characteristics are quite similar among different languages (Byrne *et al.*, 1994), the main observations of the present experiment could hold for English and other languages, though there are likely some language effects.

The target sentences were presented in nine second intervals, allowing the subjects time to repeat the words they recognize as well as guess. An operator recorded the number of correctly recognized words for each sentence.

Four masker signals were used: SSN, cafeteria noise, car interior noise, and noise from a bottling hall. We use the SSN included with the Dantale II corpus, which is produced by superimposing the speech material in the corpus. The cafeteria masker was a recording of an uninterrupted conversation between a male and a female Danish speaker in a cafeteria background (Vestergaard, 1998). The signal was equalized to match the long-term spectrum of the target sentences. This was done to isolate the effects of masker modulation and long-term average spectrum. The car interior noise was a recording during highway driving and was chosen to represent a quasi-stationary noise with strong low-frequency content. The fourth noise used was a recording of bottles rattling on a conveyor belt in a bottling hall (Vestergaard, 1998), and was chosen to represent a signal with strong high-frequency content. All stimuli were diotically presented through headphones.

TABLE I. Seven combinations of masker type and mask type. Note that TBM and IBM with SSN masker are identical.

	Speech shaped noise	Cafeteria noise	Car interior	Bottling noise
IBM	1	2	3	4
TBM		5	6	7

For each masker type, three mixture SNR levels were selected along with eight values of RC. Given that the IBM and the TBM are identical with the SSN masker, there were seven combinations of masker type and mask type, as shown in Table I.

Mixture SNR levels were set to match measured 20% and 50% SRTs for each masker type. The third SNR level was fixed at −60 dB to create IBM-gated noise similar to Wang *et al.* (2008).

B. Sessions

The experiment was divided into two sessions. In Session I, the slope and SRT of each subject's psychometric curve of the unprocessed mixtures and each of the four maskers were measured using the adaptive Dantale II procedure, and the mixture SNR levels for 20% and 80% correct word identification were derived (Brand and Kollmeier, 2002; Wagener *et al.*, 2003). In Session II, intelligibility was measured on a grid of three different mixture SNR levels and eight different RC values (including an "unprocessed" condition, see later) for each of the seven conditions in Table I. This generated a total of 3 SNR levels, 8 RC values, and 7 conditions of Table I, resulting in $3\times8\times7=168$ points, where intelligibility was measured. Each combination was tested on each subject using two sentences. Hence, each subject listened to a total of $2\times168=336$ sentences, which required reuse of sentences. To prevent memorization, order of the sentences was balanced as much as possible within and across subjects, and appeared random to the subjects.

From Session I measurements, logistic functions

$$P(\mathrm{SNR}) = (1+\exp(4s_{50}(L_{50}-\mathrm{SNR})))^{-1} \quad (2)$$

were fitted by means of the maximum likelihood method, assuming a binomial distribution of individual sentence scores (Brand and Kollmeier, 2002) yielding the 50% SRT (L_{50}) and slope (s_{50}) parameters for each subject and each masker type. The two initial sentences of each adaptation were discarded, and to reduce the effects of outliers, the data from the three best and three worst performing subjects were left out before averaging in order to derive the 20% and 50% SRT values. Pilot experiments revealed an effect of a princi-

TABLE II. SRT at 50% correct L_{50} and slope s_{50} parameters of the logistic function, Eq. (2), estimated from Session I measurements, using maximum likelihood with correction for gated noise (see text, Sec. II B). The next column shows the derived 20% SRT for average subject performance. The last two columns show the upper and lower RC values for the four masker types. Offline simulations were used to determine the RC values for obtaining IBM sparseness of 1.5% and 80% ones in the mask. The three TBM conditions 5–7 of Table I all used RC values corresponding to IBM/SSN with mixture SNR corresponding to masker type.

Masker type	50% SRT mixture SNR (L_{50}) (dB)	Slope at SRT (s_{50}) (%/dB)	20% SRT mixture SNR (dB)	RC for 1.5% ones in mask (dB)	RC for 80% ones in mask (dB)
Speech shaped noise	−7.3	15.1	−9.8	12.7	−30.3
Cafeteria	−8.8	7.5	−13.8	24.6	−27.4
Car interior	−20.3	12.7	−23.0	27.5	−25.2
Bottling noise	−12.2	5.7	−18.4	23.1	−34.9

pal difference between the continuous masker used in Session I, and the binary gated masker used with the ITFS signals in Session II. The effect caused a slightly decreased performance in the latter case. This effect has previously been described by Wagener (2003, Chap. 5) where a comparison of continuous versus gated noise indicated a 1.4 dB increase in SRT (L_{50}) and a decrease in slope (s_{50}) from 21%/dB to 18%/dB. Accordingly, we adjusted the measured SRT from Session I by adding 1.4 dB and slope by multiplying 18/21, resulting in the values listed in the first two columns of Table II. The third column shows 20% SRT derived from the adjusted parameters. The measured SRTs and slopes for speech-shaped and cafeteria noise all agree with previous results on the same material (Wagener, 2003; Wang *et al.*, 2009).

In order to determine the range of RC values to use, offline simulations were carried out to identify the RC values that yielded mask densities of 1.5% and 80% measured as percent ones in the mask within speech intervals (see Sec. II D for signal processing details). For each masker type seven RC values were then identified by equidistant sampling (in decibels) between these two points. For the three TBM conditions, the set of RC values equaled the set for IBM/SSN (condition 1 in Table I) since the binary masks are identical by definition. An eighth additional unprocessed condition was added, where the mask was set to 1 in all frequency bands within the speech intervals, and 0 outside these intervals, creating essentially a gated masker.

Speech intervals were derived from the target sentences alone and were used for all mixture SNR computations by averaging target and masker energy within speech intervals only. A speech interval was defined by low-pass filtering the absolute target sample values using a first-order IIR low-pass filter with the time constant of 1 ms (for 20 kHz sample rate the transfer function was $H(z)=\lambda/(1-(1-\lambda)z^{-1})$, $\lambda=0.04877$), thresholding the result at 60 dB below the maximum value, and further designating all non-speech intervals less than 2 s as speech to include inter-word intervals in all sentences. All detected speech onsets were shifted 100 ms backward to account for forward masking effects (Wang *et al.*, 2009).

C. Subjects

A total of 15 normal-hearing, native Danish speaking subjects participated in the experiment. The subjects volunteered for the experiment and were not paid for their participation. Their age ranged from 25 to 52 with a mean age of 35. The audiograms of all subjects indicated normal hearing with hearing thresholds below 20 dB HL in the measured range of 250 Hz–8 kHz.

D. Signal processing

All target and masker signals were resampled from 44.1 to 20 kHz sampling rate. Gain factors for target and masker were computed in order to achieve a given mixture SNR and fixed mixture power. This was done by computing the signal energies of target and masker within the speech intervals previously defined. The target and masker signals were processed separately by means of a gammatone filterbank, consisting of 64 channels of 2048-tap FIR filters; each channel has the bandwidth of 1 ERB and channel center frequencies range from 2 to 33 ERBs (corresponding to 55–7743 Hz) linearly distributed on the ERB-rate scale (Patterson *et al.*, 1988; see also Wang and Brown, 2006). The filterbank response was divided into 20 ms frames with 10 ms overlap, and the total signal energy was computed within each T-F unit.

For IBM processing, a binary mask was formed by comparing the local SNR within a T-F unit against LC, assigning 1 if the local SNR was greater than LC and 0 otherwise. For TBM processing, the reference masker (i.e. the SSN masker) was processed through the filterbank, with a gain set to achieve a 0 dB mixture SNR. The TBM was formed by comparing the local SNR within a T-F unit using the reference masker against the RC threshold, assigning 1 if the local SNR was greater than RC.

The binary mask signal was then upsampled to the full 20 kHz sampling rate by means of a sample-hold scheme followed by low-pass FIR filtering using a 10 ms Hanning filter. In each band, the target-masker mixture was delayed 20 ms in time, accounting for the total delay from the T-F unit energy summation, sample-hold, and low-pass filtering, before the upsampled mask was multiplied with the mixture. Finally, the ITFS processed waveform was synthesized using time reversed gammatone filters.

The target and masker stimuli for Session I were processed through the filterbank analysis and synthesis proce-

dure (no binary mask was applied), reducing the signal bandwidth to 55 Hz–7.74 kHz in order to match processed signals in Session II.

E. Procedure

1. Session I: SRT and slope measurements

The first session consisted of an adaptive Dantale procedure for each of the four masker types. Prior to this the subjects were given a short training session consisting of 30 randomly chosen sentences using speech-shaped and cafeteria noise maskers. These maskers were chosen to let listeners familiarize themselves with the task under stationary and non-stationary noise conditions.

In the adaptive Dantale procedure, the mixture SNR was varied after each sentence according to the number of correctly identified words, and the 20% and 80% SRTs were tracked in an interleaved manner (Brand and Kollmeier, 2002). The 20% and 80% points were chosen since they were proposed by Brand and Kollmeier (2002) to be optimal for the simultaneous measurement of the logistic function parameters L_{50} and s_{50} of Eq. (2). A total of 30 sentences were presented for each masker type in the adaptive procedure. To account for learning effects, the order of masker types was balanced across subjects (Beck and Zacharov, 2006).

2. Session II: ITFS mixtures

In the second session, each subject listened to 336 offline computed ITFS sentences. The stimuli alone lasted approximately 51 min so the subjects were allowed two breaks in the middle.

Prior to the main experiment, subjects were exposed to 60 sentences of training using all four noise types. First, for each masker type ten sentences corresponded to the unprocessed condition with increasingly lower mixture SNRs. The remaining 20 training sentences corresponded to various ITFS conditions, randomly selected but increasing difficulty. We found from pilot experiments that an extended training procedure was required to reduce learning effects and subject variability.

Learning and other temporal effects were accounted for by using a balanced design: for each subject the ordering of the seven conditions was changed and for each condition the ordering of SNR levels and RC values were balanced as much as possible.

Subjects were seated in a sound treated room where sounds were presented using Sennheiser HD280 Pro headphones connected to a SoundBlaster SB0300 sound card, using a PC running MATLAB.

3. Level of presentation

All mixtures were normalized to have same broadband long-term signal power before ITFS processing, both across mixture SNR and across noise types. The SSN condition was used to calibrate the presentation level to 65 dB(A) sound pressure level, and the volume control settings were then held fixed. The calibration was done using a sound level meter coupled to an earpiece of the headphones. The resulting presentation levels were measured to 62 dB(A) for cafeteria noise, 60 dB(A) for car interior noise, and 68 dB(A) for bottling hall noise.

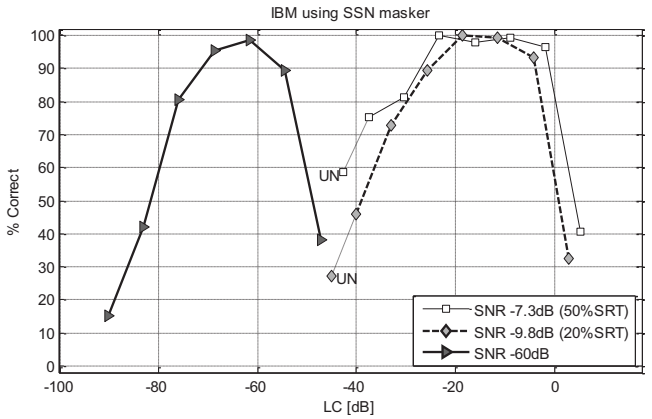


FIG. 2. Percentage of correctly identified words for IBM-processed mixtures with SSN masker as function of LC used for generating the IBM. Three mixture SNR levels are shown. The unprocessed conditions do not correspond to a particular LC value, but are inserted to the left of the respective curves, marked as "UN" and connected with dotted lines. Chance performance level is 10%.

III. RESULTS AND DISCUSSION

Figure 2 shows the percentage of correctly identified words as a function of LC for IBM segregated mixtures with the SSN masker in the three mixture SNR settings, averaged over all subjects. The unprocessed conditions do not correspond to a particular LC value, and are inserted as the leftmost points of the respective curves (marked as "UN") and connected with dotted lines to the curves.

The unprocessed data points resulted in higher performance than expected; across conditions the average scores are 25.7% and 59.5% out of 600 answers, which are larger than the 20% and 50% expected scores. This could be explained by the training that was encountered during Session I and during the training session introduced between Session I and Session II, as described in Sec. II E 2.

Each of the three curves shows a plateau or peak of very high intelligibility; for the 50% SRT (SNR of −7.3 dB), the interpolated average performance was above 95% in the interval $-25\ \text{dB} < \text{LC} < -2\ \text{dB}$, a 23 dB wide region. For 20% SRT (SNR of −9.8 dB) the interval was $-22\ \text{dB} < \text{LC} < -6\ \text{dB}$ and 16 dB wide, while for the −60 dB case the interval was $-69\ \text{dB} < \text{LC} < -59\ \text{dB}$ and 10 dB wide. The results for 20% and 50% SRT have similar profiles as those reported by Brungart *et al.* (2006) and Li and Loizou (2008). In Brungart *et al.*, 2006, the range is $-12\ \text{dB} < \text{LC} < 0\ \text{dB}$ using a multi-talker task and similar ITFS processing. The plateaus in the present study are wider than those of Brungart *et al.* (2006), due to higher scores at lower LC values, while plateau upper bounds are similar. Li and Loizou (2008) reported plateaus from −20 to +5 dB at −5 dB SNR and −20 to 0 dB at −10 dB SNR using a sentence test with a SSN masker and a T-F representation with linear frequency. The observed differences are probably due to differences in sentence material and mixture SNR.

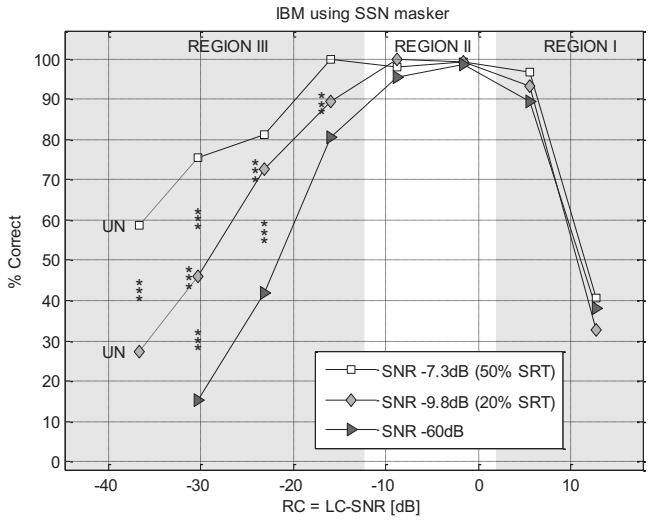


FIG. 3. Percentage of correctly identified words for IBM-processed mixtures with SSN masker as function of RC=LC−SNR. This figure gives a different plot of the same data in Fig. 3. Asterisks (*) indicate significant differences between intelligibility scores at adjacent mixture SNR levels, or when placed to the left of a diamond, between the scores at the lowest and highest mixture SNR levels, according to a Tukey HSD test. In the figure, * corresponds to $p<0.05$, ** to $p<0.01$, and *** to $p<0.001$.

The −60 dB SNR curve, however, is different. First of all, since the mask here was applied to essentially pure noise, this is consistent with the results of Wang *et al.* (2008) who demonstrated that listeners achieve nearly perfect recognition from IBM-gated noise where the mask is obtained from speech and SSN. This process of producing intelligible speech from noise may be viewed as a form of noise gating. Our results extend their findings by showing that the vocoding ability of the IBM applies to a range of LC values. This range is not much smaller than those of the performance plateaus at much higher mixture SNR levels, a finding that has not previously been reported.

Secondly, the shape of the −60 dB curve is similar to but narrower than the curves at higher SNR levels, but its position on the LC axis is very much shifted. As pointed out by Brungart *et al.*, (2006), the IBM is insensitive to covariations of LC and mixture SNR. This means that the mask pattern is a function of the difference LC-SNR, which was termed RC in Sec. I A.

A. Performance versus RC

Depicting the performance curves versus RC rather than LC brings the curves together, as shown in Fig. 3. Most notably the decline in performance at high RC values seems to be aligned well. Recall that the IBMs for the three SNR levels are equal for a fixed RC regardless of mixture SNR.

A two-way analysis of variance (ANOVA) with repeated measures was performed on the rationalized arcsine transformed subject mean percentage scores (Studebaker, 1985). The ANOVA revealed significant effect of mixture SNR, RC, and of interaction terms, as indicated in Table III. To further investigate the interaction effect, a *post hoc* Tukey HSD test was performed comparing all pairwise differences across SNR. In Fig. 3, asterisks are used to indicate significant pairwise differences, where the significance level is indicated by their number: * indicates $p<0.05$, ** indicates $p<0.01$, and *** indicates $p<0.001$. In this case, all pairwise comparisons that were significant were at the level of $p<0.001$. The significance of the difference between the upper and lower SNR performance is indicated to the left of the corresponding data point of the middle SNR curve (diamond).

In Fig. 4, plots similar to Fig. 3 are shown for the remaining conditions tested. The two rows of the plots show IBM and TBM processing, respectively. The three columns correspond to the three remaining masker types: cafeteria, car interior, and bottle noise. As shown in Table III, a two-way ANOVA in all conditions revealed significant effects of mixture SNR, RC, and of interaction terms.

The results in Fig. 4 show patterns similar to that of Fig. 3. Tukey HSD tests revealed significant differences across mixture SNR for low RC values just as was the case for the IBM/SSN condition.

1. Interpretation using regions in RC

In a manner similar to Brungart *et al.* (2006) we divide the performance curves into three distinct regions. The main difference in our analysis is that our regions are defined in terms of RC instead of LC. The purpose is to interpret the intelligibility improvement in terms of RC (Fig. 3), instead of LC (Fig. 2). While the aim of the analysis by Brungart *et al.* (2006) was to separate effects of informational and energetic masking, our analysis highlights the importance of the binary mask pattern.

TABLE III. Two way ANOVA test results using rationalized arcsine transformed mean subject scores (Studebaker, 1985) revealed significance of effects of mixture SNR, RC, and interaction terms for the measurement data shown in Figs. 3 and 4.

	Effect of mixture SNR	Effect of RC	Effect of interaction
Test statistic	$F(2,28)$	$F(7,98)$	$F(14,196)$
IBM/SSN (Fig. 3)	136.1, $p<0.000\,01$	153.1, $p<0.000\,01$	13.8, $p<0.000\,01$
IBM/cafeteria	340.5, $p<0.000\,01$	149.7, $p<0.000\,01$	17.4, $p<0.000\,01$
IBM/car noise	172.4, $p<0.000\,01$	295.5, $p<0.000\,01$	12.0, $p<0.000\,01$
IBM/bottling noise	173.0, $p<0.000\,01$	126.0, $p<0.000\,01$	12.2, $p<0.000\,01$
TBM/cafeteria	253.1, $p<0.000\,01$	95.1, $p<0.000\,01$	11.8, $p<0.000\,01$
TBM/car noise	133.1, $p<0.000\,01$	156.8, $p<0.000\,01$	12.3, $p<0.000\,01$
TBM/bottling noise	234.3, $p<0.000\,01$	146.7, $p<0.000\,01$	15.2, $p<0.000\,01$

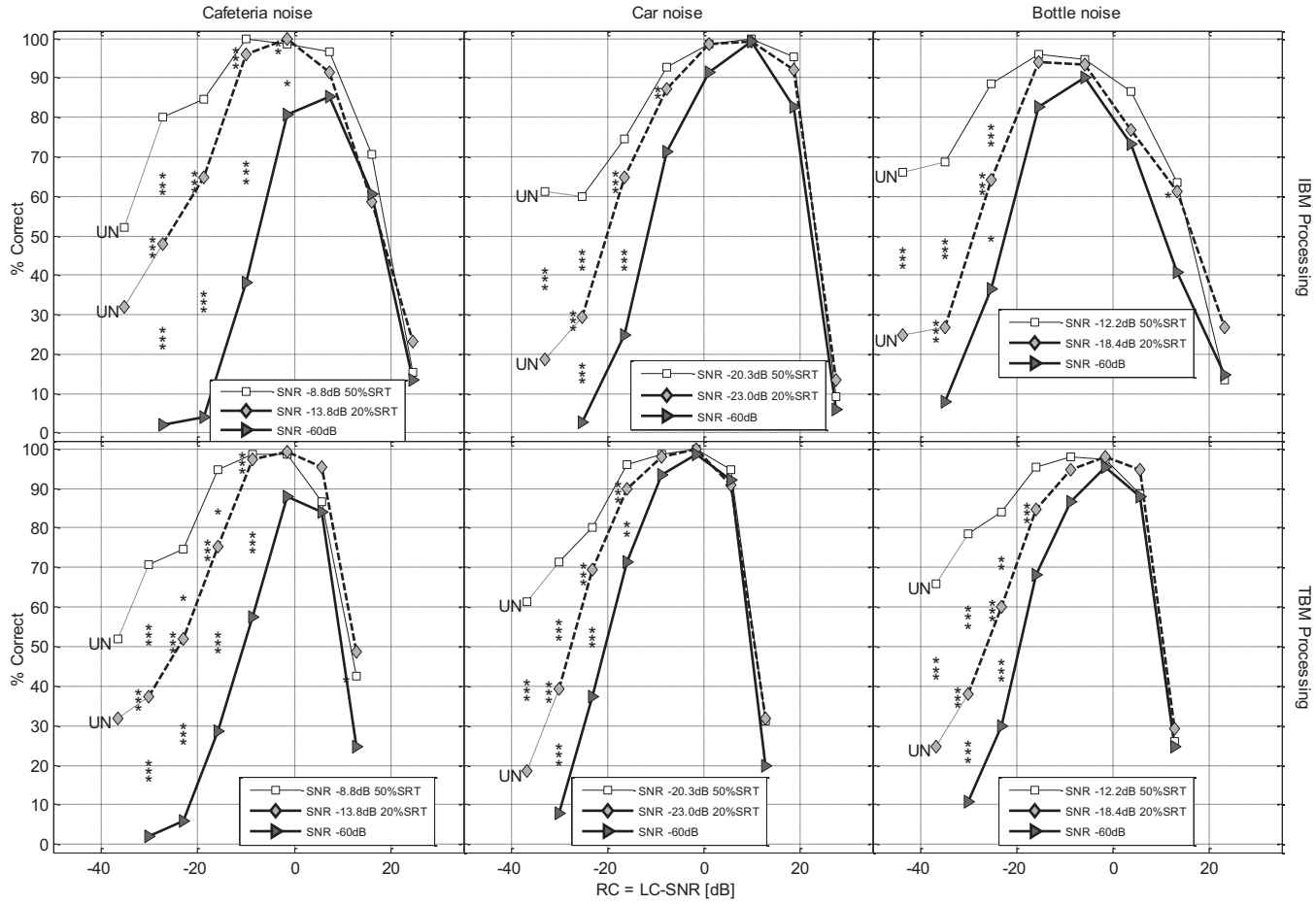


FIG. 4. Percentage of correctly identified words versus RC for IBM-processed mixtures (upper row) and TBM processed mixtures (lower row). Each column corresponds to a masker type. The three curves in each plot correspond to the mixture SNR levels (squares: 50% SRT, diamonds: 20% SRT, and triangles: −60 dB mixture SNR). Asterisks (*) indicate significant difference between adjacent mixture SNR levels, or when placed to the left of a diamond, between the lowest and highest mixture SNR levels (* corresponds to $p<0.05$, ** to $p<0.01$, and *** to $p<0.001$), according to Tukey HSD tests.

Region I corresponds to large RC values, where intelligibility decreases with increasing RC due to increasing sparseness of the ideal mask. In our results from the IBM/SSN condition, performance decreased for RC > −2 dB.

Region II corresponds to an intermediate range of RC values, with nearly perfect performance. For the IBM/SSN condition this occurred as a plateau at RC values between −8.8 and −1.6 dB, where intelligibility was above 95%.

Region III ranges below approximately RC = −10 dB in the IBM/SSN case. In this region performance decreases as RC decreases and the number of T-F units included in the IBM increases, until the performance of the unprocessed mixture is reached.

A general pattern in our data is that the influence of mixture SNR on the recognition performance decreases with increasing RC: In Regions I and II the effect was small or insignificant, while in Region III there was significant influence.

The fact that the performance in Region I (high RC values) showed only a negligible or small effect of mixture SNR level suggests that the target component of the processed mixture plays a relatively small role. Our results seem to indicate that some of the traditional cues for speech perception, such as F0, periodicity, and other temporal fine structure cues, are less important in Region II than in Region III and of even smaller importance in Region I. Otherwise one would have expected a difference in performance across mixture SNRs. So the application of the IBM seems, on the one hand, to improve the intelligibility relative to the unprocessed condition and, on the other hand, to reduce or eliminate the listener's ability to make use of speech cues other than what is carried in the binary mask. This result is of particular interest for the design of hearing aids, since reports suggest that the ability of hearing impaired subjects to make use of temporal fine structure cues is limited compared to normal listeners (Lorenzi *et al.* 2006; Hopkins *et al.* 2008), making the trade-off more favorable for the hearing impaired.

In Region III, there was an overall significant effect of mixture SNR (indicated with asterisks in Figs. 3 and 4). We further note that across all seven mask scheme/masker conditions, the increase in performance at the mixture SNR corresponding to 20% SRT from Region III to Region II is accompanied by an increasing vocoding ability at −60 dB mixture SNR.

2. Influence of masker type

The results in Fig. 4 show that the RC values beneficial to intelligibility varied across the seven mask scheme/masker

TABLE IV. Measured peak intelligibility score (in percentage) for noise gating data (at a mixture SNR of −60 dB) together with average width (in RC) of performance plateau where the interpolated performance was within 95% of the peak value, for the four masker types and two mask computation schemes.

	Speech shaped noise		Cafeteria		Car interior		Bottling noise	
IBM	98.7%	23.6 dB	85.3%	20.7 dB	99.3%	23.0 dB	90.0%	19.0 dB
TBM			88.0%	16.9 dB	98.7%	21.5 dB	95.3%	18.4 dB

conditions. While the plateau became narrower at lower mixture SNR levels, its position shifts across the seven conditions tested. As already described, mixture SNR, which factors in the definition of RC, is not a good indicator of intelligibility across masker types. For instance, in the IBM/bottle noise curve at the mixture SNR corresponding to 50% SRT, the performance plateau—the region of RC values where intelligibility is within 95% of the maximum score—ranged from −22 to −3 dB (measured on interpolated mean data), while in the IBM/car noise curve the corresponding plateau occurs in the RC range of −4 to 19 dB.

Table IV shows the average plateau width for the three mixture SNR levels for each of the seven mask scheme/masker conditions. The IBM/SSN condition produced the broadest plateau, 23.6 dB on average, and the TBM/cafeteria the narrowest plateau of 16.9 dB. Comparing mask schemes within masker signals, the IBM showed slightly wider average plateaus for all masker types. The table also gives the peak intelligibility scores of various noise gating curves.

B. Discussion of binary noise gating results

The noise gating performance curves (SNR −60 dB) form a performance lower bound for each masker type: in no case was the noise gating performance significantly greater than that for any other mixture SNR level. The measured peak value of the noise gating performance curves varied across masker type and mask computation scheme as indicated in Table IV. The effect of masker type was greater than the effect of mask computation scheme (from 85.3% for IBM/cafeteria to 99.3% for IBM/car noise).

The cafeteria noise was a relatively poor signal for vocoding, yielding maximum scores of 85% correct using IBM and 88% using TBMs, a result which may be explained by the sparse energy distribution in retained T-F units: The presence of 1 in the binary mask may coincide with a dip in the noise signal. In our data, the performance in the TBM/cafeteria condition with the −60 dB SNR was significantly lower at RC=15 dB than those with higher SNR levels. The modulation dips of the cafeteria masker made the distribution of T-F energy in the processed signal relatively sparse, a likely reason for reduced intelligibility performance.

Figure 5 shows the density of the binary mask measured as percentage ones in the mask averaged over all speech intervals (see Sec. II B) as function of channel center frequency for different masker types. The bold lines correspond to the RC value with the highest noise gating intelligibility (at mixture SNR of −60 dB). The figure shows that when the target and masker signal spectra are matched (speech-shaped and cafeteria noise) the result is a more uniform mask density compared to when the signals are not matched (bottle noise and car noise).

It should be noted that, for stationary maskers, the TBM is similar to the IBM with a LC parameter made frequency dependent in such a way that the resulting distribution of mask sparseness resembles that of the TBM (i.e. IBM with SSN masker). Since the TBM in the bottle noise case brings some intelligibility benefits over the IBM, it is possible that speech separation algorithms that estimate the IBM would also benefit from making the LC parameter frequency dependent, to ensure that enough ones are present in frequency bands relevant for speech.

C. Results from TBM

In Fig. 6, the results of applying the TBM to mixtures of the four masker types are compared. From left to right the mixture SNR level corresponds to 50% SRT, 20% SRT, and −60 dB. The curves corresponding to the four different maskers appear to align well. This is further reflected in Table V, showing the results from a two-way ANOVA with repeated measures performed on the rationalized arcsine mean subject scores, for each of the three mixture SNR levels. Compared to the previous analysis, the effects are not as strong; in fact, the noise type influence was not above the standard 5% significance level for the 20% SRT data and the interaction term for the 50% SRT data was also not significant. Tukey HSD tests revealed significance in the pairwise differences across masker type only for cafeteria noise in −60 dB SNR against all three other noise types, and only for RC values of −23.1, −15.9, and −8.7 dB as indicated with asterisks in Fig. 6.

D. Performance versus mask density

Given the importance of mask density for resulting intelligibility, the performance scores versus resulting overall mask density are plotted in Fig. 7. The mask density was measured as resulting percentage of ones in all frequency bands within speech intervals. The unprocessed condition is indicated as having 100% ones in the mask. The IBM results are connected with solid lines, and the TBM results are connected with dashed lines. Note that a nonlinear abscissa is used to better illustrate the performance differences at low percentages.

All curves show maximum performance between 15% and 60% ones in the masks. The curves all show a sharp decline toward zero at low percentages, a plateau in the middle which is wider for higher mixture SNRs and a gradual drop to the level of unprocessed mixtures, from 40%

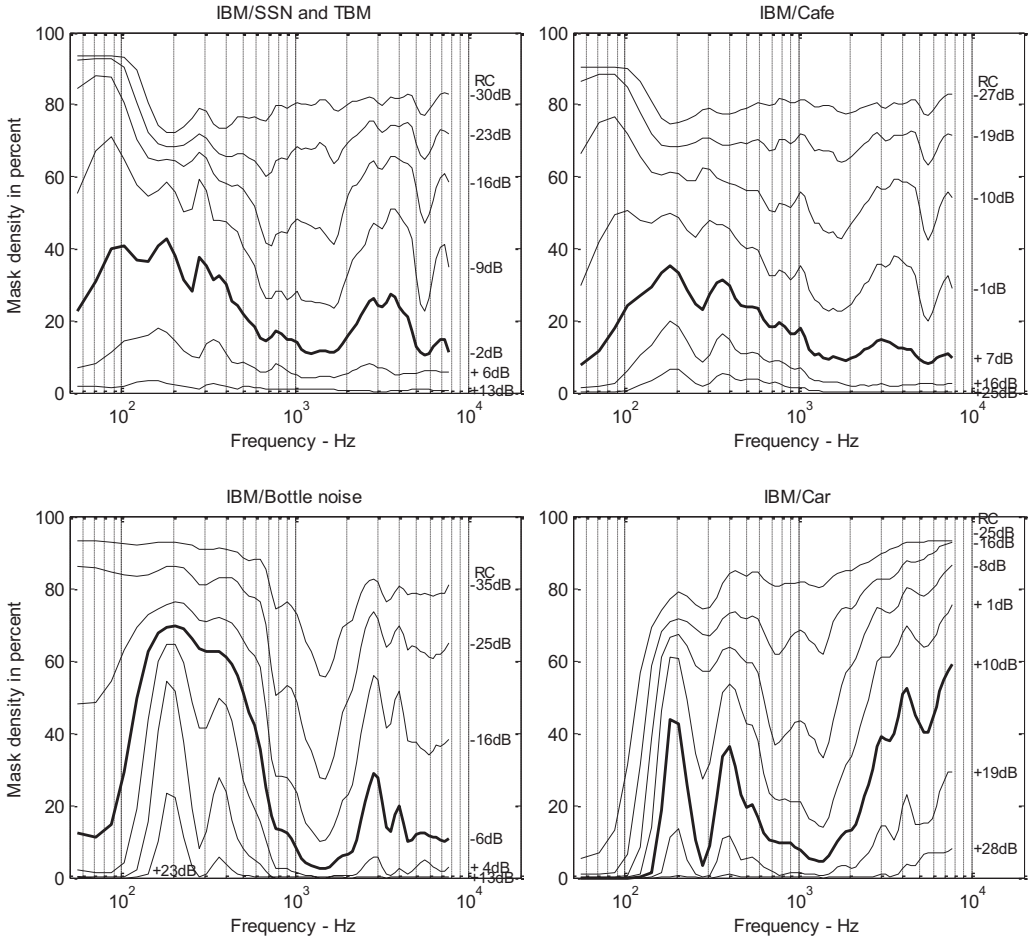


FIG. 5. Mask density (in percentage of mask value 1) as function of channel center frequency averaged over entire sentence material. The corresponding RC value used for computing the mask is indicated to the right of each curve. The bold line corresponds to the RC value with the highest intelligibility for IBM-gated noise (at mixture SNR of −60 dB). The mask densities of the TBM masks equals that of the IBM/SSN by definition.

to 100% ones in mask. The TBM and IBM curves are generally similar, with slightly larger scores for the target binary mask except for the cafeteria masker at high percentage of ones. Below 5%–10% ones, the TBM scores were higher than for the IBM for all masker types. For the exceptional case of the cafeteria noise, the IBM strategy based on mixture SNR was apparently better than the TBM scheme according to the target energy. Overall, it is rather remarkable how well the TBM and IBM results are aligned, considering their differences with respect to RC in Fig. 4.

IV. CONCLUSION

By measuring intelligibility of ideal binary-masked noisy speech, we have shown that intelligibility performance

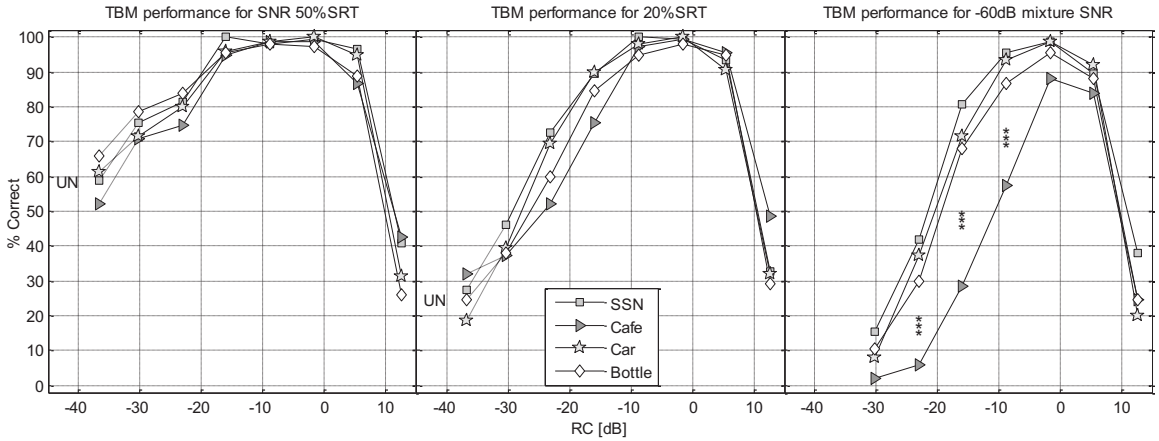


FIG. 6. Percentage of correctly identified words versus RC for TBM processed mixtures comparing the effect of noise types. Note that all curves use the same mask for a given RC. The three plots correspond to the three mixture SNR levels. The individual curves correspond to masker types. Asterisks (*) indicate significant difference between adjacent noise types (* corresponds to $p<0.05$, ** to $p<0.01$, and *** to $p<0.001$), according to a Tukey HSD test.

TABLE V. Two way ANOVA test was performed on rationalized arcsine transformed mean subject scores revealing significance of effects of noise type, RC, and interaction terms for the measurement data shown in Fig. 6.

	Effect of noise type	Effect of RC	Effect of interaction
Test statistic	$F(3,42)$	$F(7,98)$	$F(21,294)$
50% SRT data	3.80, $p<0.017$	92.3, $p<0.000\,01$	1.54, $p<0.063$
20% SRT data	2.78, $p<0.053$	147.4, $p<0.000\,01$	2.25, $p<0.001\,7$
−60 dB SNR data	87.9, $p<0.000\,01$	297.1, $p<0.000\,01$	6.19, $p<0.000\,01$

curves became aligned across a large range of mixture SNR levels when using the RC defined as the difference of LC and SNR. This alignment was demonstrated for four masker types, using the IBM as well as the proposed TBM. By fixing RC and varying the mixture SNR level, we identified three regions in RC, differentiated by intelligibility and influence of the mixture SNR level. In Regions I and II, weak or insignificant influence was found, whereas in Region III the influence was large and significant. The size and location of the regions varied with masker type.

By applying IBM processing to mixtures of low negative SNR levels, we have extended the findings of Wang *et al.* (2008) showing that the processing acts as binary noise gating and produces intelligible speech at a range of sparseness configurations parametrized by RC. We further showed that the proposed TBM based on the target signal alone was comparable to the IBM in terms of intelligibility improvements. For a given level of mask sparseness, the mean measured TBM intelligibility scores were even slightly higher than those of the IBM in some conditions.

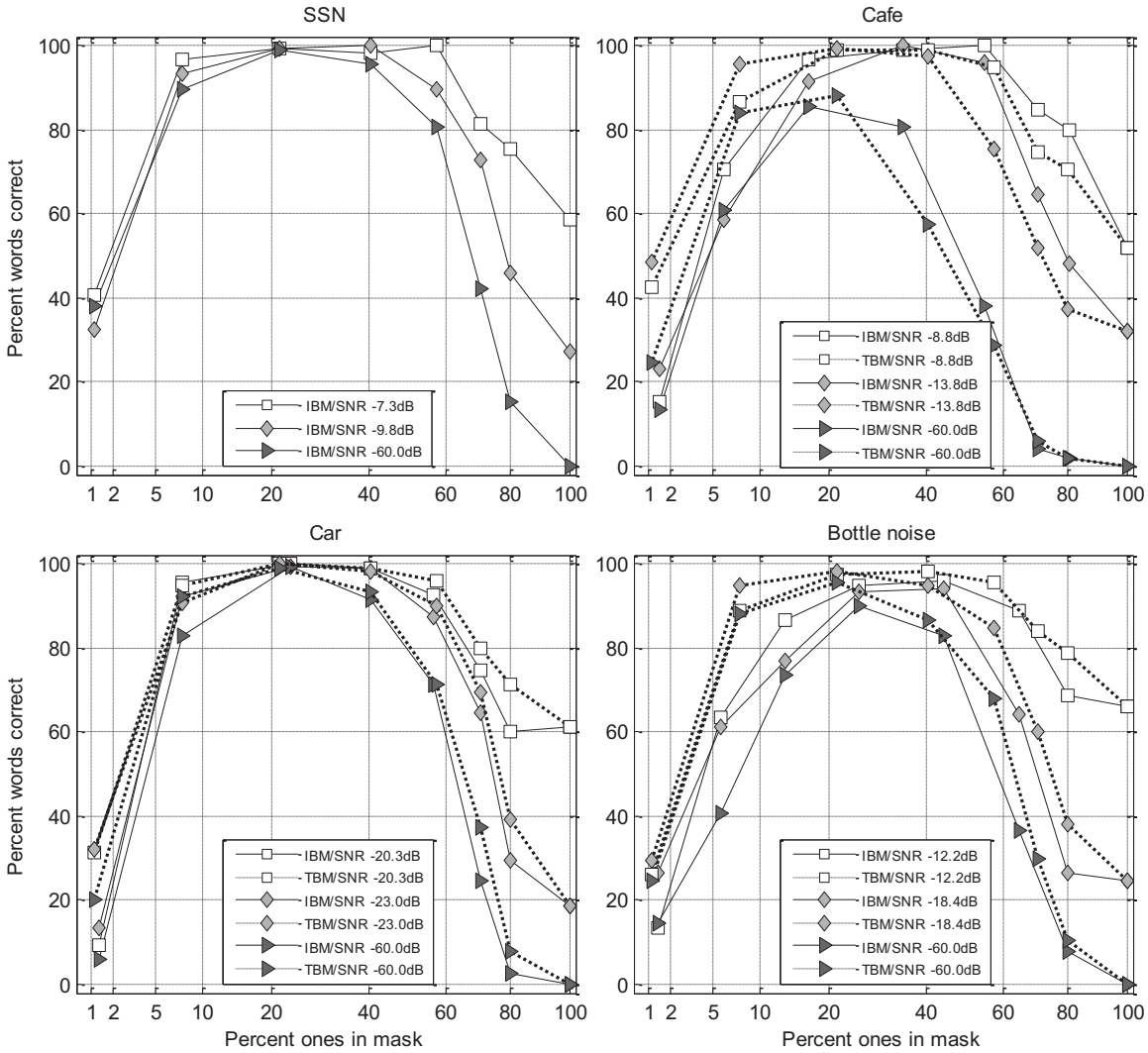


FIG. 7. Percentage of correctly identified words as function of mask density. The four plots show the four masker types: SSN, cafeteria, car noise, and bottle noise. Each plot corresponds and the three mixture SNR levels and the two mask computation schemes. The IBM results are connected with solid lines, and the TBM results with dotted lines. The unprocessed condition is marked as 100% ones in mask.

ACKNOWLEDGMENTS

The authors would like to thank Tayyib Arshad for assistance in performing the experiments, volunteering subjects for participating, and colleagues for discussions and proofreading. The research of D.W. was supported in part by an AFOSR grant (FA9550-08-1-0155) and a NSF grant (IIS-0534707).

ANSI S3.5-1997 (**1997**). "American National Standard: Methods for the calculation of the speech intelligibility index" (American National Standards Institute, New York).

Anzalone, M. C., Calandruccio, L., Doherty, K. A., and Carney, L. H. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480–492.

Beck, S., and Zacharov, N. (**2006**). *Perceptual Audio Evaluation: Theory, Method and Application* (Wiley, Chichester, UK).

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge MA).

Brungart, D., Chang, P. S., Simpson, B. D., and Wang, D. L. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Byrne, D., Dillon, H., and Tran, K. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.

Goldsworthy, R. L., and Greenberg, J. E. (**2004**). "Analysis of speech-based speech transmission index methods with implications for nonlinear operations," J. Acoust. Soc. Am. **116**, 3679–3689.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Audiol. **11**, 79–87.

Hopkins, K., Moore, B. C. J., and Stone, M. A. (**2008**). "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," J. Acoust. Soc. Am. **123**, 1140–1153.

Houtgast, T., and Steeneken, H. J. M. (**1971**). "Evaluation of speech transmission channels by using artificial signals," Acustica **25**, 355–367.

Kollmeier, B., and Wesselkamp, M. (**1997**). "Development and evaluation of a German sentence test for objective and subjective speech intelligibility assessment," J. Acoust. Soc. Am. **102**, 2412–2421.

Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.

Li, Y., and Wang, D. L. (**2009**). "On the optimality of ideal binary time-frequency masks," Speech Commun. **51**, 230–239.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. (**2006**). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," Proc. Natl. Acad. Sci. U.S.A. **103**, 18866–18869.

Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1988**). "An efficient auditory filterbank based on the gammatone function," Report No. 2341, MRC Applied Psychology Unit, Cambridge.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Srinivasan, S., and Wang, D. L. (**2008**). "A model for multitalker speech perception," J. Acoust. Soc. Am. **124**, 3213–3224.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Vestergaard, M. (**1998**). "The Eriksholm CD 01: Speech signals in various acoustical environments," Report No. 050-08-01, Oticon Research Centre Eriksholm, Snekkersten.

Wagener, K. (**2003**). "Factors Influencing Sentence Intelligibility in Noise," Ph.D. thesis, Oldenburg University, Oldenburg, Germany.

Wagener, K., Josvassen, J. L., and Ardenkjær, R. (**2003**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.

Wang, D. L. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.

Wang, D. L., and Brown, G. J. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley, Hoboken NJ).

Wang, D. L., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2008**). "Speech perception of noise with binary gains," J. Acoust. Soc. Am. **124**, 2303–2307.

Wang, D. L., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2009**). "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Am. **125**, 2336–2347.

Perception of complete and incomplete formant transitions in vowels

Pierre Divenyi[a)]
Veterans Affairs Northern California Health Care Systems and East Bay Institute for Research and Education Martinez, California 04553

(Received 31 October 2007; revised 10 June 2009; accepted 11 June 2009)

In everyday speech, formant transitions rarely reach the canonical frequencies of a target vowel. The perceptual system often compensates for such production undershoots, called vowel reduction (VR), by a perceptual overshoot of the final transition frequencies. The present investigation explored the perceptual parameters and existence region of VR. In a series of experiments a 100-ms steady-state vowel V_1 was followed by a formant transition toward a target vowel V_2. By manipulating both its duration and velocity, in most stimuli the transition was truncated and only seldom reached the target. After being presented with the vowel V_2 before each block of trials, listeners were asked to rate their confidence that the transition actually reached the V_2 target. Transitions along six trajectories connecting the three cardinal vowels /a/, /i/, and /u/ in both directions as well as the transition /ie/ (halfway along the trajectory /ia/) were examined in experiments in which either the duration of the transition was fixed and its velocity was varied or vice-versa. Results confirmed the existence of perceptual overshoot and showed that, at the point a transition short of reaching the vowel V_2 was just perceived as if it had reached the target, transition duration and transition velocity were inversely related. The amount of overshoot was found to be larger for larger V_1-V_2 distances and shorter trajectory durations. The overshoot could be reliably predicted by a linear model based on three parameters—the extent of V_1-V_2 distance, transition velocity, and transition acceleration. These findings suggest that the perceptual dynamics of speech relies on mechanisms that estimate the rate of change in the resonant characteristics of the vocal tract. [DOI: 10.1121/1.3167482]

PACS number(s): 43.71.Es, 43.66.Lj, 43.71.An [RSN] Pages: 1427–1439

I. INTRODUCTION

To say that speech communication is based on gestures, that is, speech movements, amounts to a tautology. When one individual engages in conveying information to another by way of speech, he/she has to perform a series of articulatory movements that generate a sequence of changes in his/her vocal tract. Although there are brief islands of relatively quiescent periods between consecutive gestures leading to such changes, an overwhelming portion of the information transmitted is encoded in the transitions that connect motion-free quasi-steady-state periods, i.e., in the changes themselves.

Paradoxical as it may sound, because vowels appear to constitute the bulk of such still periods, perhaps nowhere is the importance of transitions better illustrated than in the production and perception of vowels. Studies on vowel production have revealed a considerable variability in the formant values—even within a given language—across different speakers, speaking rates, speaking styles, and contexts (Peterson and Barney, 1952). One of the most intriguing aspects of vowel perception is that, despite large overlaps of the vowel categories when mapped onto the formant domain, vowels retain their perceived identity across physical changes due to variations of different origin (Nearey, 1989). Early investigations on running speech presented evidence that a non-negligible portion of formant variability may be attributed to dynamic context (Furui, 1986; Lindblom and Studdert-Kennedy, 1967; Strange *et al.*, 1979) and suggest that data obtained on the production and perception of static vowels are inadequate to describe, or model, spoken language. This conclusion is supported by a number of studies focused on the identification and discrimination of V_1-V_2, CV, VC, and CVC transitions (Andruski and Nearey, 1992; Carré *et al.*, 2007; Carré and Divenyi, 2000; Hillenbrand and Nearey, 1999; Nearey, 1989; Strange, 1989; van Son and Pols, 1993; van Wieringen, 1995; van Wieringen and Pols, 1994). Also, experiments on the perception of vowels in "silent center" CVC syllables have demonstrated that a vowel replaced by silence remains largely identifiable (Fox, 1989; Jenkins *et al.*, 1999). On the other hand, replacement of consonantal segments in sentences by noise preserves about twice the information as does replacement of vowel segments, suggesting that even vestigial CV and VC transitions in the vowels can lead to identifying the missing consonant or even consonant cluster (Kewley-Port *et al.*, 2007). Interestingly, in contrast to the well-known differences between formants of isolated vowels produced by male, female, and child talkers, formant trajectories from vowel onset to vowel center in CV diphones are remarkably stable across talkers (Hillenbrand *et al.*, 2001). Such stability is also reinforced by an essential lack of overlap between vowel categories in multitalker French V_1-V_2 utterances, observed when the vowel pair is represented by the rates of F1-F2 change instead of the actual formant frequencies (Carré *et al.*, 2007). It appears, therefore, that vowel transitions contain information

[a)]Electronic mail:pdivenyi@ebire.org

which not only complements that conveyed by static formants but which, under contextual diversity, is necessary, and may be sufficient, for signaling the vowel's identity. Patterns of inherent diphthongization of vowels provide a dynamic cue that contributes to their identification in American English (Assmann and Katz, 2005; Morrison and Nearey, 2007) while formant contours carry a great amount of phonological information in Australian English (Watson and Harrington, 1999). Important for understanding the perception of diphthongization is to establish the conditions under which altering one or several formants of a steady-state vowel will be perceived as a change to a diphthong or a different vowel. While this change, both from the productive and perceptual points of view, is influenced by the fundamental frequency f_0 (Di Benedetto, 1994), altering the transition appears to affect the percept also when f_0 is held constant, as demonstrated in a series of studies by Nabelek and co-workers. Manipulating the rate and the duration of formant transitions shifted the perceptual boundary between an initial steady-state vowel and a target vowel at the end of a transition (Nabelek *et al.*, 1993) or changed the point at which the initial vowel was heard either as steady-state or diphthongized (Nabelek *et al.*, 1996) Added noise or reverberation also influenced these boundaries (Nabelek *et al.*, 1994, 1996).

Transitions *toward* a vowel, however, whether from a preceding consonant or another vowel, often terminate at formant frequencies that fall short of those of a canonical vowel, i.e., the target vowel in isolation. Such production undershoot, termed vowel reduction (VR, Lindblom and Studdert-Kennedy, 1967), has been associated with casual speaking style ("hypo-speech," Lindblom, 1983), increased speaking rate (Flege, 1988; Gottfried *et al.*, 1990), and reduced vocal effort (Lienard and Di Benedetto, 1999), as well as with consonantal context (Stevens and House, 1963). Interestingly, listeners not only successfully identify these reduced vowels (Macchi, 1980) but also perceive them as *identical* to the canonical vowel when they are preceded by a transition, and as *different* from the canonical vowel when they are presented in isolation (Carré and Divenyi, 2000). This perceptual overshoot has been considered the perceptual system's compensation for the articulatory undershoot of VR (Lindblom and Studdert-Kennedy, 1967; Pols and van Son, 1993). Contrary to citation-form "hyper-speech" in which boundaries of the vowel triangle are expanded (Whalen *et al.*, 2004), in VR the vowel preceded by the transition may have formant values displaced toward the center of the vowel triangle or simply lying on an incomplete trajectory leading to but not reaching the target vowel. VR-type formant changes also underlie diphthongization (Andruski and Nearey, 1992) and CV and VC diphones (Hillenbrand *et al.*, 2001). Three main acoustic factors appear to contribute to VR: formant displacement, transition duration, and transition velocity (Lindblom *et al.*, 1996). Not surprisingly, the same three factors also emerge as those contributing most heavily to identification of silent center vowels in different CVC contexts (Jenkins *et al.*, 1994; Strange and Bohn, 1998), boundary shifts between /a/ and /aI/ percepts (Nabelek *et al.*, 1994), and discrimination of vocalic transitions (van Wieringen and Pols, 1995). VR perception has been observed in /wVw/ and /jVj/ contexts and has been also shown to be affected by nonlinearity, i.e., the acceleration component, of the transition (Nabelek and Ovchinnikov, 1997).

The above reports therefore suggest that the perceptual overshoot compensating for the VR undershoot represents not only a noteworthy illusion but also a possibly indispensable aspect of speech perception: it stands to guarantee perceptual invariance by keeping the perceived identity of the vowel constant despite any physical variability imposed by prosodic and contextual differences. Although work over the last 30-or-so years has uncovered characteristics of this phenomenon, it was seldom the focus of investigations but, rather, a tool to test the validity of a large-scale hypothesis—such as the overarching role of context effect (e.g., Holt *et al.*, 2000)—or an ancillary observation while studying another phenomenon—such as inherent spectral changes in vowels (Nearey and Assman, 1986). Yet, knowledge of the parametric regions of existence of VR perception could bring us closer to understanding the dynamic processes operating in speech. For one, identifying the mechanisms behind the perceptual VR overshoot would not be without interest: the controversy concerning the origin—auditory or linguistic—of the overshoot for the identification of vowels in a CVC setting has been around for well over a decade but is still not settled. The currently prevailing auditory explanation states that spectral contrast, whether intrasegmental, i.e., within a CVC syllable, or extrasegmental, i.e., a CVC test item preceded and/or followed by an anchoring sound, can fully account for the perceptual overestimation of the extent of the transition (Holt *et al.*, 2000; Lotto and Holt, 2006). The contrasting linguistic theory states that the perceptual overshoot of the transition is a direct reflection of the articulatory undershoot because speech is perceived by processes that translate the acoustics into the articulatory gestures that were used by the talker (Fowler, 1994; Vishwanathan *et al.* 2009). Both views have classic predecessors and are backed by modern-day confirmatory evidence. Early auditory explanation based on frequency-specific adaptation along formant sweeps (Lindblom and Studdert-Kennedy, 1967) is consistent with spectrotemporal modulation indices of auditory cortical responses to complex sounds (David *et al.*, 2009). The concept of gestural perception of speech is at least 300 years old[1] and it formed the basis for the analysis-by-synthesis theory proposed in the 1960s (Halle and Stevens, 1962); recently, it has received support from neuroimaging studies showing activation of precentral cortical motor areas mapped to the lips and the larynx, when listening to speech (e.g., Pulvermuller *et al.*, 2006). While testing a hypothesis in favor of either of the two contracting theories is not the objective of the present study, addressing the perception of VR may give support to either theory—or both.

The experiments reported in this article assessed listeners' perception of vowel-to-vowel transitions by asking if they perceived whether a specified target was reached, rather than by looking at conditions in which the presence of the transition resulted in the percept of a diphthong—as was the case in many studies on this topic (e.g., Nabelek *et al.*,

1994). The choice of studying V_1-V_2 rather than CVC or CV or VC transitions was dictated by a desire to also examine trajectory durations that, albeit brief, could exceed those of consonantal transitions. Aside from determining parameters of complete and incomplete transitions that lead to the percept of a vowel target reached, these experiments also represent a test of whether the Gestalt principle of continuity—the auditory mechanism shown to follow a monotonic rise or fall of the frequency of a sine wave signal (Crum and Hafter, 2008; Nakajima *et al.*, 2000)—can be also demonstrated to operate in the processing of vowels.

Specifically, the experiments reported here used VR to gauge the perception of vowels in a dynamic context by examining in detail the role of three transition parameters: formant trajectory, transition duration, and transition velocity. The specific aim of the experiments was to find for each condition a transition boundary, i.e., the extent to which the transition leading from a steady-state vowel V_1 to formant frequencies of a target vowel V_2 should be truncated so that listeners are just able to perceive the target at the end of the transition. For the first two experiments, the separation of the starting and target vowels was kept as large as possible by using the six trajectories between pairs of the three vowels /a/, /i/, and /u/. Experiment 1 was directed at determining the boundary for a variable transition velocity with its duration held constant within each condition, while experiment 2 measured the boundary for a variable transition duration with its velocity held constant within each condition[2]. In experiment 3, transition velocity and transition duration boundaries were measured for the /ia/ and the /ie/ trajectories, i.e., two trajectories that were collinear and differed only in their lengths. Finally, a control experiment tested the notion that VR-type overshoot for the perception of the extent of transitions can be observed for nonspeech complex sounds having the frequency of a single formant modulated to form a sweep.

II. METHODS

A. Stimuli

In all experiments, the stimulus consisted of an unbroken succession of a steady-state vowel V_1 and a transition toward a second vowel V_2. The transitions proceeded along a trajectory from a vowel at one corner of the vowel triangle /a/-/u/-/i/ toward a vowel at one of the other two corners. These vowels do occur in American English, although the /u/ vowel is rare in everyday spoken language in America. They were chosen because they are as far from each other as the vowel space allows, thereby facilitating the study of transitions between them. In experiments 1 and 2, the trajectories investigated consisted of the exhaustive set of the six V_1-V_2 combinations of the three corner vowels—/ai/, /ia/, /au/, /ua/, /iu/, and /ui/.[3] All stimuli were digitally synthesized at a 20 kHz sampling rate using the distinctive region model (DRM) (Mrayati *et al.*, 1988)—an articulatory model based on deformations of a vocal tract-analog tube at specific locations that generates a parameter table, which serves as input to a cascade-mode formant synthesizer. The resonators of the first four formants were excited by a train of 100-μs pulses.

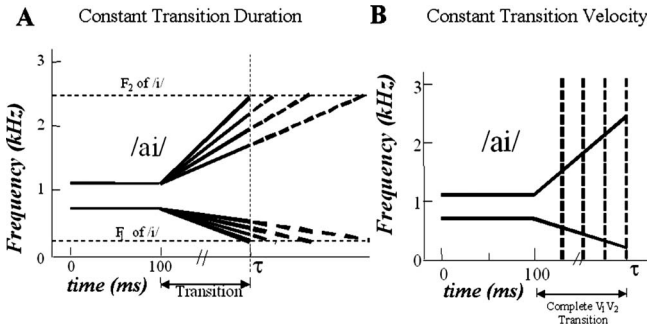


FIG. 1. Schematic time-frequency diagram of the stimuli illustrating the trajectories of the first two formants in four of the eight tokens in each experimental condition. Panel A shows tokens of different transition velocities (i.e., transition slopes) in the constant transition duration experiments (experiments 1 and 3a), while Panel B shows different transition durations for the constant transition velocity experiments (experiments 2 and 3b). The specific example in both panels illustrates the formant trajectories for the /ai/ transition. Note that in both diagrams only one of the transitions reaches the formants of the /i/ target vowel.

The fundamental frequency (f_0) contour of the pulses corresponded to that of a natural V_1-V_2 utterance by a male talker, with an initial f_0 of 120 Hz linearly rising to 140 Hz for the duration of V_1 and linearly declining back to 120 Hz for the duration of the transition. The duration of the first vowel V_1 was always 100-ms; the duration and the velocity of the transition were the two parameters of interest.

All transitions were synthesized by following, in a piecewise linear fashion, a formant frequency table generated by the DRM, updated every 2 ms. Because formant changes generated by the DRM synthesizer were a monotonic but not linear function of time, the velocity of the frequency change was also compressive-nonlinear (with an exponent between 0.6 and 0.9), with the bulk of the variation concentrated at the onset (at which the given regions of the vocal tract specified by the model are least impeded when the area functions expand or contract). The curvature of the formant change, therefore, was similar to that of the quadratic function used by Nabelek and Ovchinnikov (1997), although the rate in the present experiments was variable.

Experiment 1. In each block of experiment 1, the transition duration was held constant (30, 60, or 90 ms) while eight tokens with different transition velocities were presented 16 times each in random order. The V_2 target was reached, and thus the entire trajectory covered, by only the token having the highest transition velocity; transitions in the other seven tokens had increasingly lower velocities with trajectories covering, on the average, lengths of 7/8, 6/8, 5/8, etc. of the maximum. Figure 1(a) illustrates schematic first and second formant frequency-vs-time trajectories along the /ai/ continuum for four of the eight variable transition velocity tokens used in experiment 1.

Experiment 2. In each block of experiment 2, the transition velocity was held constant (corresponding to the entire V_1-V_2 trajectory covered in 70, 100, or 130 ms) while eight tokens with different transition durations were presented 16 times each in random order. Again, the V_2 target was reached and the entire trajectory covered by only one token, the one having the longest transition duration; the durations of the transition in the other seven tokens were 7/8, 6/8, 5/8, etc.

of that of the longest-duration token. Figure 1(b) shows frequency-time trajectory plots along the /ai/ continuum for four of the eight variable transition duration tokens used.

Experiment 3. In experiment 3, the perception of transitions along the /ie/ trajectory was investigated. Since the natural American English vowel /e/ has formant frequencies very close to those lying halfway on the trajectory between /i/ and /a/, the formant frequencies chosen for the target vowel V_2 in this experiment were chosen to be exactly at the midpoint of the /ia/ trajectory used in experiments 1 and 2. In experiment 3a, the stimulus was similar to that of experiment 1:a 100-ms initial vowel V_1, /i/, was followed by a constant-duration (30, 60, or 90 ms) transition toward the target vowel V_2, that is, /e/. Again, in each run with a given transition duration, eight tokens with different transition velocities were presented in random order. In experiment 3b, as in experiment 2, three transition velocities were used: those corresponding to a fully completed /ie/ transition having a duration of 70, 100, and 130 ms. Just as in experiment 2, in each run in the second condition there were of eight tokens having the transition portion of the stimulus truncated to different durations.

Perceptually, all stimuli in all experiments (even, to some extent, the highest velocity and longest duration tokens) sounded like truncated diphthongs with the truncation percept somewhat attenuated by windowing the envelope of all stimuli with 10-ms cosine-square on- and off- ramps.

B. Subjects

The same eight subjects between the ages of 19 and 29 participated in experiments 1, 2, and 3. A crew of six subjects, four of whom took part in the main experiments, were listeners in the control experiment. They were recruited through advertisement and were paid hourly wages. All participants had pure-tone thresholds of 15 dB HL or better at all frequencies in the range from 250 to 8000 Hz, in both ears. The subjects gave informed consent to participate in the study and the procedures conformed to those mandated by the Institutional Review Board of the VA Northern California Health Care Systems. All subjects had prior experience as listeners in other psychoacoustic tests but were unsophisticated with respect to phonetic awareness.

C. Procedure

Control of the experiments as well as collection and analysis of the data were accomplished by a PC using an integrated psychoacoustic data collection and analysis software developed at our laboratory. Listeners were tested one at a time. Prior to the beginning of each block of trials, they were presented the target vowel V_2 ten times in succession and were told that the vowel stimuli in the following block will glide toward this target. For example, if a given block of trials investigated the perception of the /au/ trajectory, the target was the /u/ vowel; this vowel was the one presented in isolation ten times before the block. The listeners' task was to rate on a four-point scale their confidence that this target was actually reached, with "1" being very sure and "4" not at all. They were encouraged to use all four categories within each block of trials. They were seated in a sound-treated testing booth and gave their rating responses by pressing one of four horizontally aligned labeled buttons on a special response box. In each trial the stimulus token was presented only once, i.e., repeated listening was not allowed. The listeners were allowed to respond any time (minimum 600 ms) after the termination of the stimulus; their key press response initiated the next trial. Presentation of the stimuli was done through earphones at a maximum instantaneous level of 80 dB sound pressure level. In all experiments, in each block of trials each of the eight tokens occurred 16 times in random order. For each condition and each subject, the results reported below are based on the average of six to eight blocks of 128 trials (8 targets times 16 repetitions per target).

D. Control experiment

Stimuli and targets in experiments 1–3 were synthesized vowels. In order to determine whether overestimates observed for the perception of VR would be restricted to stimuli recognized as speech or whether similar results could be also obtained for nonspeech stimuli, a control experiment was conducted. Stimuli for the control experiment were designed to maintain certain features of the stimuli used in the main experiments: a 100-ms steady-state first portion with a given frequency profile was followed by a glide toward a target—a sound with a different frequency profile. The target sound was explicitly identified to the listeners as such and presented ten times before the block of trials. The listeners' task was to indicate their confidence that the glide reached the target. Inspired by a classic study (Brady *et al.*, 1961), stimuli were sinusoidal complexes generated by passing a pulse train through a single resonator having its resonant frequency kept constant for 100 ms at 2452 Hz and then linearly swept toward 1078 Hz at a rate and duration that varied from condition to condition. Thus, stimuli of the control experiment were buzz-like sounds with some vowel-like characteristics; they were variants of the /ia/ stimuli used in experiments 1, 2, and 3, with only the second formant resonator excited. Two constant transition duration conditions (30 and 90 ms) with eight different glide velocities and two constant transition velocity conditions (velocities needed to reach the target in 70 and 130 ms) with eight different durations were investigated. Pilot listening tests with /ai/-like upward going single-formant transitions yielded results similar to the /ia/-like control experiments.

E. Data analysis

The objective of the data analysis was to assess the transition duration and transition velocity boundaries at which the subjects heard the presence of the target vowel, based on their confidence rating responses. Analysis of the results revealed that the subjects' rating criteria were not identical: some were tilted toward a stricter and others toward a more lax judgment regarding the perceived presence of the target. In order to determine the duration and velocity boundaries while neutralizing the effect of the subjects' idiosyncratic response criteria, the rating scale receiver-operating characteristic (ROC) method was used (Egan *et al.*, 1959). Rating

scale ROC analysis represents performance across a number of different strict-to-lax response criteria which are gauged by asking the subject to rate his/her confidence that the signal was present at a given trial. The rating ROC curve plots performance as the cumulative proportion of a certain confidence rating or stricter, given that the stimulus was the signal, as a function of the cumulative proportion of the same rating or stricter given that the stimulus was not the signal. The assumption underlying ROC analysis is that some portion (most often 50%) of the stimuli presented in a block contains a stimulus recognized by the subject as the signal, whereas the other stimuli are regarded as "noise"—in the narrow or broad sense—thus dividing the stimulus ensemble into two categories. Since in the present experiments the task was to judge the similarity of the final formant frequencies of the transition and those of the target, meaning that the "real signal" (i.e., the steady-state target vowel V_2 as presented before each block) was never actually present, a different definition of signal and noise was needed. Because in each condition the final frequencies of the eight stimulus tokens, from Token 1 to Token 8, got closer and closer to V_2, we could divide the tokens into signal and noise subsets on the basis of similarity to V_2 in seven ways, each of which yielding an ROC curve. Thus, the Token 1 plus Token 2 subset is the least similar to the target but Token 2 is more so than Token 1—i.e., in this subset Token 2 is the signal and Token 1 the noise. Similarly, in the Token 1 through Token 3 subsets Token 3 is the most similar to the target and becomes the signal while the ensemble of Tokens 1 and 2 will be the noise. As we go on this way to the last, i.e., the entire set of the eight tokens in which Token 8 will be the signal and the ensemble of all others the noise, we will have created seven partially overlapping sets in each of which the signal and the noise are defined. While it is true that the seven noise categories are increasingly overlapping, such overlap is not uncommon in signal detection analysis of categorical data where common false alarm pools are used for different hit (i.e., signal) categories (see e.g., Anderson and Neill, 2002). Thus, from the seven signal-noise sets seven ROC curves are generated. Each curve plots the cumulative proportion of a given rating being higher than or equal to "highly confident that V_2 was not reached" for a given signal-noise set; when represented on a z-scale, the relationship of these proportions can be approximated by a straight line. Since there is no basis to assume that the variance of the signal and noise distributions is identical (and therefore the ROC curve is parallel to the diagonal), we followed the method suggested by MacMillan and Creelman (1991, pp. 65ff) and calculated the ROC assuming that the slopes were other than unity. From the ROC curves we calculated, for each condition and each subject, Egan's d_e (Egan, 1975) and the response bias β. An example of the seven z-transformed ROC curves is illustrated in Fig. 2(a) (the data represent the average results of the eight subjects in one of the constant transition velocity conditions). According to signal detection theory, detection performance expressed as d_e is the distance from the origin to the point at which the ROC curve crosses the negative diagonal. Therefore, the points at which the seven ROC lines in Fig. 2(a) intersect the negative diagonal indicate the degree to which the "signal" token and the "noise" tokens could be discriminated with respect to the presence of the target vowel at the end of the transition. Distances of these points from the origin, i.e., the d_e values, are re-plotted as a function of the token number in Fig. 2(b), with the abscissa indicating the stimulus value—transition duration in this example—corresponding to the token numbers. (A nonlinear regression line fitted to these points illustrated as the curve in Fig. 2(b) accounts for 89% of the variance.) Except for error, this d_e scale, therefore, is monotonically related to the likelihood that the listener perceives the final formant frequencies of the transition trajectory as being those of the V_2 vowel. The abscissa value at which the curve in Fig. 2(b) crosses the d_e value of 1.41 was adopted as the performance level at which the listener judged, with fair certainty, the final formant frequencies of the stimulus as being those of the target.[4] As shown in Sec. III, for most experimental conditions the boundary corresponded to a token with a trajectory shorter than that of the eighth token (i.e., the only one of the eight tokens to reach the target frequencies), thus indicating an overestimation of the trajectory length—or VR, as it was described earlier. Consequently, all other trajectories with lengths between the one corresponding to the boundary and that of the eighth token are bound to lead to perceptual overestimation, i.e., the area under the curve between the bound-

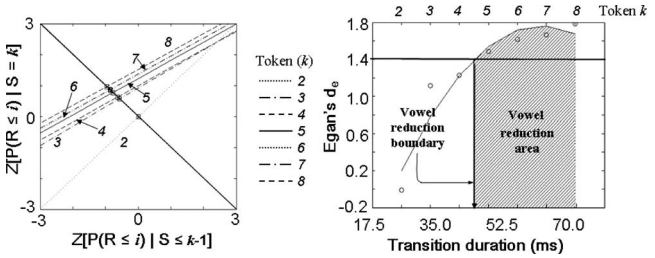


FIG. 2. Panel A shows one (constant transition velocity) condition as an example to display the seven ROC curves representing all subjects' average rating responses of the eight tokens k, truncated to different durations along the trajectory, heard as having their transition reach the target vowel. Each of the seven straight lines was fitted to three cumulative "positive" rating response probabilities as a function of the three corresponding cumulative "negative" response probabilities, where the positive responses signify the subject giving a certain rating when the token number was k and the negative responses mean that the subject gave the same rating to token numbers less than k. Panel B illustrates the points at which the seven positive response tokens intersected the negative diagonal in Panel A. The $1.41d_e$ intersection point is considered as the VR boundary, i.e., the token value corresponding to the subject saying that he/she heard the transition reaching the target vowel at the d_e=1.41 level. Thus, tokens with serial numbers higher than indicated by this boundary always led to the subject perceiving that the target vowel was, in fact, reached.

TABLE I. Steady-state formant frequencies of the four vowels used in the experiments and their corresponding BM distance from the helicotrema.

Vowel	F1 (Hz)	F2 (Hz)	F1 BM pos. (mm)	F2 BM pos. (mm)
/a/	729	1079	12.24	14.63
/i/	239	2452	6.48	20.02
/u/	271	640	4.03	11.48
/e/	533	1959	10.47	18.51

TABLE II. Trajectory lengths and F1-F2 vector sums of the four V_1-V_2 vowel pairs used in the experiments, in Hertz and in corresponding BM millimeter distance.

Trajectory	F1 (Hz)	F2 (Hz)	F1 BM distance (mm)	F2 BM distance (mm)	F1-F2 vector sum	F1-F2 BM distance vector sum
/ai/	490	1373	5.76	5.39	1458	7.89
/au/	458	439	5.20	3.16	634	6.09
/iu/	21	1812	0.55	8.55	1812	8.57
/ie/	294	493	3.96	1.51	574	4.24

ary line and the eighth is the one of VR. Since essentially for all experimental conditions this abscissa crossing occurred between two tokens, transition durations or velocities corresponding to the VR boundary were obtained by linear interpolation between those corresponding to the tokens on either side.

Similar ROC analyses were performed for the data obtained in each experimental condition to obtain estimates for boundaries of transition velocity (in experiments 1 and 3a) and transition duration (in experiments 2 and 3b). These estimates, by definition, minimize the effects of individual tendencies toward too strict or too lax certainty responses. Across all experimental conditions, intersubject variability of the d_e threshold estimates was 0.150 (0.077) while that of intersubject variability of the response bias β was 2.243 (1.012). In other words, the difference among listeners was almost 15 times larger as regards their strict-lax response criteria than as regards their ability (and proclivity) to associate a given transition duration/velocity combination with the target.

To tie the data on the perception of transitions to physical dimensions of the stimuli, the frequency change in the first two formants (F1 and F2) was selected. Although a transition between any two vowels represents simultaneous changes that involve all formants, because proportionally the largest changes occur in the first two formants (F1 and F2), and because F1 and F2 frequencies are sufficient to unambiguously specify a vowel, the results were gauged on the F1-F2 change in the transition. Since auditory sensitivity to frequency change as well as to formant frequency discrimination in vowels under ordinary listening conditions (Kewley-Port and Zheng, 1999)[5] is a constant proportion of critical bands (Moore, 1973), we represented the extents of the formant changes in the transitions as F12, the vector sum of the F1 and F2 changes on a scale analogous to the Bark scale: the basilar membrane (BM) distance between the start and the end of the transition. The algorithm used was the one by Greenwood (1990)[6] that estimates positions of maximum excitation on the BM from the helicotrema to the base. Frequencies and corresponding BM positions of the four V_1 and/or V_2 vowels included in the study are shown in Table I, whereas maximum transition lengths (i.e., trajectories between V_1 and V_2 vowels) as well as F1 and F2 vector sums are shown in Table II.

III. RESULTS

Combined results of all subjects in experiment 1, that is, transition velocities for constant-duration transitions estimated to yield VR boundaries, are illustrated in Fig. 3 for the three transition durations investigated. Panel A shows the velocity estimates for the /ai/ and /ia/ trajectories, panel B for the /au/ and /ua/ trajectories, and panel C for the /iu/ and /ui/ trajectories. With the exception of two among the eighteen

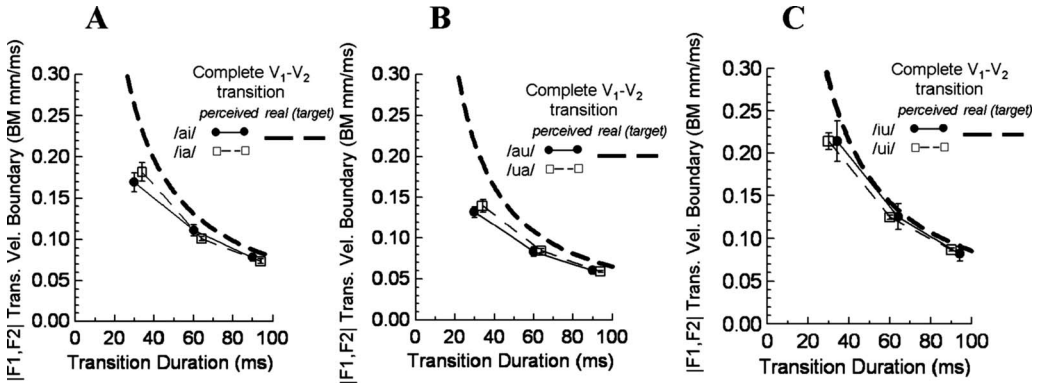


FIG. 3. Average results of the eight subjects in experiment 1, the constant transition duration experiment. On the abscissa: transition duration; on the ordinate: transition velocity boundary estimated with the ROC method illustrated in Fig. 2(b). This boundary corresponds to the lowest velocity of the |F1, F2| vector's change (in BM mm/ms) at which the listeners perceived a complete V_1-V_2 trajectory, i.e., heard the terminal formant frequencies of the transition glide as if they had reached those of the target vowel presented before the block of trials. The error bars stand for intersubject standard errors of the mean. The three panels illustrate thresholds for transitions proceeding in both directions along the three trajectories defined by the /a/-/i/-/u/ vowel triangle. The dashed line in each panel shows transition velocities across all durations at which the V_1-V_2 trajectory is complete. Data points lying below this line indicate perceptual overestimate of the length of the transition's trajectory, i.e., perceived VR, and the vertical distance between the line and any date point gives the degree of overestimation.

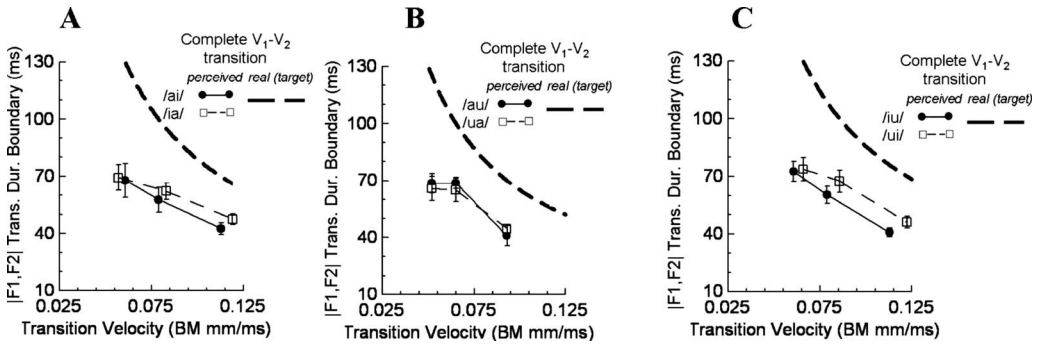


FIG. 4. Average results of the eight subjects in experiment 2, the constant transition velocity experiment. On the abscissa: transition velocity in BM mm/ms; on the ordinate: transition duration corresponding to the VR boundary. The error bars stand for intersubject standard errors of the mean. The three panels illustrate VR thresholds for transitions of either direction on the three trajectories defined by the /a/-/i/-/u/ vowel triangle. The dashed line in each panel shows transition durations across all velocities at which the V_1-V_2 trajectory is complete. Data points lying below this line indicate perceptual overestimate of the length of the transition's trajectory, i.e., perceived VR, and the vertical distance between the line and any date point gives the degree of overestimation.

conditions, intersubject variability was low to very low. The most striking feature of the data is that velocity estimates for the two opposing trajectory directions were essentially identical in each of the three panels. As shown by the heavy dashed lines in Fig. 3, at the 30-ms transition duration for all three trajectories and at the 60-ms transition duration for the /ai/-/ia/ and /au/-/ua/ trajectories, velocities estimated to be at VR boundary are below those of the trajectories actually reaching V_2. One must therefore conclude that in those conditions a perceptual overshoot, indicating VR, occurred. At the longest, 90-ms, duration and at the 60-ms duration for the /iu/-ui/ transitions the velocity estimates overlapped with the velocity of the complete V_1-V_2 trajectory's velocity, indicating that in those conditions perception of the transition was almost veridical. In other words, for all three vowel pairs, VR was largest when the duration of the transition was the shortest.

Combined results of all subjects in experiment 2, that is, transition durations for constant-velocity transitions estimated to yield VR boundaries, are illustrated in Fig. 4 for the three transition velocities investigated. In Fig. 4, panel A shows these boundaries for the /ai/ and /ia/ trajectories, panel B for the /au/ and /ua/ trajectories, and panel C for the /iu/ and /ui/ trajectories. Duration estimates were very similar for the two opposing directions along all three trajectories, although not as perfectly overlapping as it was seen in experiment 1. Intersubject variability was even lower than in experiment 1. In all three panels velocities of V_1-V_2 durations over the complete trajectory at the three constant transition velocities are also shown. Since the duration estimates always fell well below those of the durations of the complete V_1V_2 transitions (shown by the heavy dashed lines), it may be concluded that a strong perceptual overshoot leading to VR was manifest, with the strongest VR observed for the lowest of the three constant velocities.

Combined results of the eight subjects in experiment 3 are presented in Fig. 5. Panel B shows results of experiment 3a, with transition velocity at VR boundary as a function of transition duration, whereas results of experiment 3b are shown in panel C, with transition duration at VR boundary as a function of transition velocity. For comparison purposes, data for the /ia/ trajectory from experiments 1 and 2 are also shown on the same graph. Panel B of the figure suggests that when listeners had to determine which transition of different velocities was most likely to reach the /e/ target, VR was absent; i.e., the perception of the /ie/ transition was veridical for all three fixed transition durations. In contrast, as shown in panel C, VR is clearly present at all three constant transition velocities of the /ie/ trajectory, although the perceptual overshoot is much smaller than the one observed for the /ia/ trajectory. One observation in panel C worth noting is the apparent continuity of the VR boundary durations for the transitions perceived to be /ia/ (on the right) and /ie/ (on the left). This continuity may be due to the fact that, as illustrated in the vowel triangle shown in panel A, the /ia/ and /ie/ trajectories were collinear with the /ie/ trajectory being half

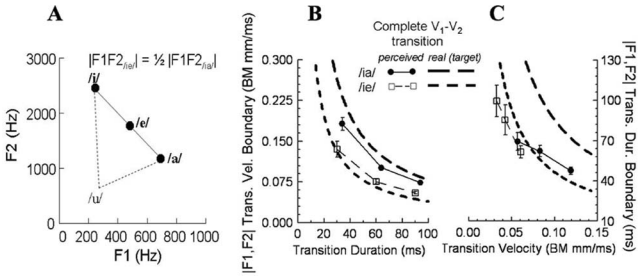


FIG. 5. Average results of the eight subjects in experiment 3, the /ie/ transition experiment. Panel A shows the formant frequencies of the /i/, /e/, and /a/ vowels in the F1-F2 plane (in linear Hz), in order to illustrate the fact that the /ie/ and /ia/ trajectories are colinear and that the length of the /ie/ trajectory is $\frac{1}{2}$ of that of the /ia/ trajectory. Panel B: results of experiment 3a. On the abscissa: transition duration; on the ordinate: transition velocity corresponding to the VR boundary shown as open squares. For comparison purposes data for the /ia/ trajectory in experiment 1 [Fig. 3(a)] are also shown as filled circles. Velocities corresponding to the complete /ie/ trajectory are shown as the line with short dashes, and those corresponding to the complete /ia/ trajectory as the line with long dashes. Panel C: results of experiment 3b. On the abscissa: transition velocity; on the ordinate: transition duration corresponding to the VR boundary shown as open squares. Data for the /ia/ trajectory in experiment 2 [Fig. 4(a)] are shown as filled circles. Data points lying below the short-dashed lines indicate perceptual overestimate of the /ie/ trajectory, whereas those lying below the long-dashed lines overestimate of the /ia/ trajectory. The vertical distance between the short-dashed line and the open squares indicates the degree of overestimation for the /ie/ trajectory, whereas the distance between the long-dashed line and the closed circles the degree of overestimation for the /ia/ trajectory. Error bars in all panels represent intersubject standard error of the mean.

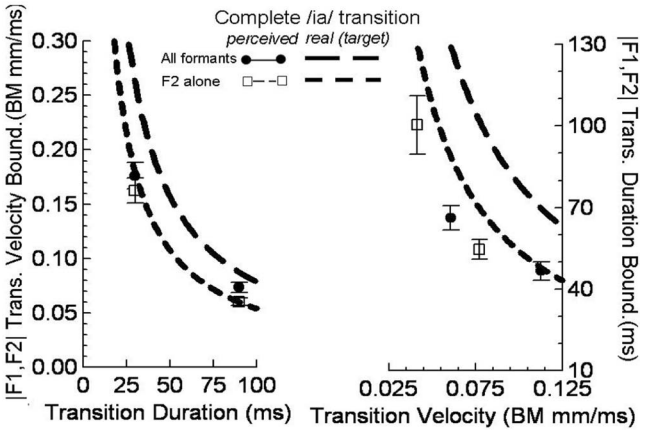


FIG. 6. Results of six listeners for the control experiment with /ia/-like glide stimuli but with only the second formant present and all others suppressed. Data from this experiment are shown as the open squares, whereas data for the corresponding full /ia/ stimuli (from experiments 1b and 2b) are represented as the solid circles. The left panel shows data for the constant transition duration, and the right panel for the constant transition velocity conditions. The short-dashed lines display velocities (left) and durations (right) corresponding to complete trajectories in the control experiment stimuli, whereas the long-dashed lines show velocities (left) and durations (right) corresponding to complete /ia/ trajectories from Fig. 3(a) and 4(a), respectively; data points below the lines indicate perceptual overestimation. Error bars represent intersubject standard errors of the mean.

of the length of the /ia/ trajectory. The interesting data points are the nearly overlapping ones in the middle: the duration estimate for the lowest /ia/ velocity (chosen to reach the /a/ target in 130 ms) is almost on top of the one for the highest /ie/ velocity (chosen to reach the /e/ target in 70, i.e., close to 130/2 ms). This overlap means that, for that particular velocity and at apparently similar (61–66 ms) durations, the transition can be perceived to lead either to the /e/ or to the /a/ vowel, depending on which one was designated as the target, played ten times pre-block, and remembered as such.

Results of six listeners in the control experiment are shown in Fig. 6 with the constant-duration conditions on the left and the constant-velocity conditions on the right as the open squares. In addition to the velocity estimates (on the left) and duration estimates (on the right) of the stimuli judged to have their terminal formant frequency reach 1078 Hz, data for the perception of the /ia/ trajectory in experiments 1 [Fig. 3(b)] and 2 [Fig. 4(b)] are also displayed for comparison. Overestimation of the trajectory's length is assessed as the difference between the velocity or duration estimates at the VR boundary and those corresponding to the complete trajectories displayed as the heavy dashed lines. While for constant-duration transitions, no overestimation of the trajectory length was observed (the difference between the velocities of full trajectory and judged trajectory at the 30-ms duration is not significant); the duration estimates for the constant-velocity conditions are significantly below those of the complete physical trajectory (the short-dashed line), thus indicating that the length of the trajectory was overestimated. Since for the /ia/ vowel stimuli the trajectory was overestimated for all conditions (illustrated by the distance between the data points and the long-dashed line that represents the target values), although less for the constant duration than the constant-velocity conditions, we have to assume that when all formants change simultaneously along their /i/ toward /a/ trajectory, the totality of simultaneous formant transitions generates a percept in which the course of frequency change is more overestimated than when the second formant alone was present. But, despite of the lack of overestimation in the constant-duration control condition, this experiment suggests that perceptual overestimation of a monotonic frequency change can also occur for complex nonspeech sounds, i.e., it is not the exclusive property of vowels.

IV. DISCUSSION

A. Perception of vowel-to-vowel transitions: A model

Listeners in the experiments reported here were given the task of finding either the velocity for constant-duration transitions or the duration for constant-velocity transitions that were perceived to lead to a target vowel specified preceding the trial block. Since knowing both the duration and the velocity of a transition can reveal the extent of the trajectory, the difference between the starting and final frequencies, i.e., the frequency trajectory traveled, represents a perceptual cue—dimension $\boldsymbol{\Delta F}$, the format change[7]—in addition to the duration $\boldsymbol{t}$ and the velocity $\boldsymbol{v}$ of the transition. The question arises: on which dimension or dimensions of the stimulus did the listeners base their decision? Because the duration of the steady-state vowel V_1 was 100-ms, i.e., long enough to form a perceptual unit (Hirsh, 1974), the transition portion of the stimuli was perceptually distinct, allowing us to focus our attention on perceptual cues of the transition alone. Since the subjects were instructed to gauge the extent of the transition trajectory and compare its final formant frequencies to those of the target, the explicitly stated cue was $\boldsymbol{\Delta F}$—a cue reported to be the most salient in the discrimination of formant transitions (van Wieringen and Pols, 1995).

Comparing the results in Figs. 3 and 5(a) with those in Figs. 4 and 5(b) suggests that, for constant-duration variable velocity transitions, the perceptual overshoot was present only when the transitions were brief, whereas for transitions of constant-velocity and variable duration it was consistent and often substantial. This discrepancy may be at least partially attributed to the different salience of these two cues, suggesting that a constant velocity $\boldsymbol{v}$ in trial blocks with randomly mixed durations $\boldsymbol{t}$ provided a more salient cue than the opposite, i.e., constant-duration stimuli with different velocities. Such larger overshoots for rapid and brief transitions were also observed by Nabelek and co-workers for the perceptual boundary of diphthongized transitions (Nabelek *et al.*, 1994). The finding is also consistent with results reported by van Wieringen and Pols (their experiment 2, 1995) as well as with a study showing that, while discrimination of frequency change velocities in the absence of other cues for sinusoidal sweeps is possible, it is also rather inefficient (Divenyi, 2005). In contrast, this latter study as well as others (e.g., van Wieringen and Pols, 1994) demonstrated that transition duration discriminability of pure tone sweeps and formants in speech stimuli are comparable (6–10% of the base duration).

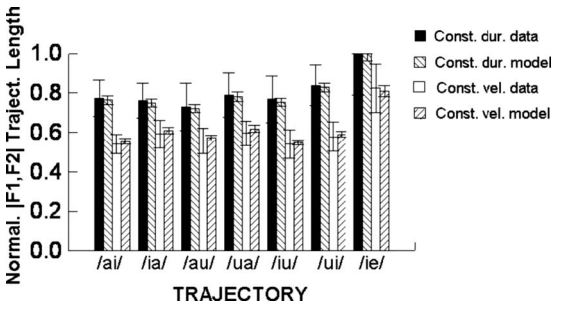


FIG. 7. Incomplete F1-F2 trajectory length necessary to generate VR normalized by the length of the complete trajectory, for the six trajectories in experiments 1 and 2 and the trajectory in experiment 3. Solid and open bars stand for the constant transition duration and constant transition velocity experiments, respectively, whereas the descending and ascending hatched bars display respective constant-duration and constant-velocity predictions of the model in Eq. (2). While the trajectory lengths are reasonably similar along the two directions of a transition between any two vowels, they differ significantly between trajectories. VR is perceived for all trajectories except the constant-duration condition for the /ie/ trajectory. Error bars represent intersubject standard errors of the mean.

With no exception, all panels in Figures 3–5 display a negative slope of the perceptual data: at the VR boundary, increasing transition durations $\boldsymbol{t}$ are associated with decreasing transition velocities $\boldsymbol{v}$ and vice versa. Such trend hints at the existence of a $\boldsymbol{t}$-$\boldsymbol{v}$ trade-off. However, a strict trade-off means that the product of transition duration and velocity, i.e., the trajectory covered, would have to be constant for all V_1-V_2 trajectories and across all conditions, which was not the case in the present data. In fact, illustrated as the dark and the empty bars in Fig. 7 for the constant-duration and constant-velocity conditions, respectively, trajectory lengths (the normalized lengths of the |F1, F2| frequency change vectors, i.e., $\boldsymbol{\Delta F}$, perceived as the complete trajectories) were not only consistently shorter for the constant-velocity than the constant-duration experiments but they were also different across the different vowel pair trajectories. Obviously, a strict trade-off $\boldsymbol{\Delta F}=\boldsymbol{tv}$ model does not account for the data. To better understand the auditory processes responsible for the perception of the complete and incomplete formant transitions in the present stimuli, a different model had to be found.

To search for an adequate but simple model, it was first assumed that we could find a linear combination of maximum trajectory (the V_1-V_2 trajectory leading to the target), velocity at VR boundary, and transition duration at VR boundary that explained a large proportion of the data variance. A stepwise regression analysis that included linear and quadratic terms found that the model

$$\boldsymbol{\Delta F}_{\text{incompl}} = \boldsymbol{C} + \boldsymbol{\Delta F}_{\text{compl}} + \boldsymbol{t} + \boldsymbol{v} \tag{1}$$

(where $\boldsymbol{C}$ is a constant, $\boldsymbol{\Delta F}_{\text{compl}}$ is the complete V_1-V_2 trajectory, $\boldsymbol{t}$ and $\boldsymbol{v}$ are the transition duration and velocity, respectively, and $\boldsymbol{\Delta F}_{\text{incompl}}$ is the incomplete trajectory length perceived to be equal to $\boldsymbol{\Delta F}_{\text{compl}}$) accounted for 0.789 of the variance $[F(3,38)=52.18,\ p<0.0001]$, although the model fit was only moderate (Durbin–Watson D statistic of 2.103). To look for a better fit, some basic assumptions of the model had been modified. One previous finding pointed to the high salience of acceleration, specifically to that of a steady-state frequency suddenly starting to change into a monotonically increasing or decreasing one (Divenyi, 2005). If such an acceleration monitor were present, it could possibly explain the large perceptual overshoot in the Fig. 3 data at the shortest fixed duration. For this, it was assumed that the stimuli were processed through a running 66-ms asymmetric window (half-power width of 34 ms)[8] and the running average of velocity and acceleration were computed. A stepwise regression analysis that included linear and quadratic terms found that the model

$$\boldsymbol{\Delta F}_{\text{incompl}} = \boldsymbol{C} + \boldsymbol{\Delta F}_{\text{compl}} + \boldsymbol{v}_{\text{runav}} + \boldsymbol{a}_{\text{runav}} \tag{2}$$

(where $\boldsymbol{v}_{\text{runav}}$ and $\boldsymbol{a}_{\text{runav}}$ are the velocity and the acceleration of the transitions calculated with the running averaging time window) accounted for 0.964 of the variance $[F(3,38)=365.02,\ p<0.0001]$ and displayed a good fit (Durbin–Watson D statistic of 2.486). The model output and the data shown in adjacent columns in Fig. 7 suggest that the target trajectory cue combined with the velocity and acceleration of the formant transition processed by a realistic window provide a reasonably good explanation of the perception of vowel-to-vowel transitions, including the perceptual overshoot. The presence of acceleration among the factors, despite the fact that the stimuli contained only linear transitions, is consistent with results by Nabelek and Ovchinnikov (1997) on overestimation of CVC trajectories using nonlinear transitions.

B. Perception of VR: An auditory explanation

The model thus shows that the extent of incomplete trajectory heard as the complete trajectory is a function of (1) the length of the complete trajectory connecting the initial and target vowels, (2) the velocity of the transition, and (3) the acceleration of the transition. The velocity and acceleration components are consistent with the existence of a putative frequency velocity detector/monitor (Divenyi, 2005; Kay and Matthews, 1972; Pollack, 1968), suggesting a frequency differentiation process. Due to the time course of the formation of excitatory and inhibitory contours in spectro-temporal receptive fields (STRFs) (David *et al.*, 2009), such velocity detector should account for an overshoot in the perception of formant transitions and the spectral asymmetry in labeling the vowel in jVj or wVw syllables, as suggested by Linblom and Studdert-Kennedy (1967) and confirmed by others (Holt *et al.*, 2000; Nabelek and Ovchinnikov, 1997). The fact that no asymmetry was found in the present data (see in Figs. 4 and 5 the practically colinear pairs of curves for either direction of the transition along the trajectories) does not invalidate the peripheral explanation, it only points out that V_1V_2 transitions may differ from CVC transitions in ways that should be examined in the future. While the build-up and decay of excitatory and inhibitory contours in STRFs is a continuous process, the sampling of frequencies at times coinciding with glottal pulses and making judgments from predictions based on a mechanism that averages instantaneous frequency samples, i.e., a frequency integration process, has also been proposed (Brady *et al.*, 1961). Thus, if duration cues are not available or are uncertain, the rate of

frequency change can be discriminated by sampling the instantaneous frequency at the beginning and end of the transition and taking the difference between them (Dooley and Moore, 1988; Pollack, 1968). In fact, a frequency sampling/integration hypothesis may be consistent with the present data: since they show larger overestimation for constant velocity and variable duration formant sweeps than for sweeps with variable velocity and constant duration, the percept of a prolonged trajectory is more likely to be induced by stimuli with a fixed rate of formant change than with a fixed duration. Chistovich as well as Feth (Chistovich, 1985; Chistovich *et al.*, 1979; Feth *et al.*, 2006) proposed a variant of the frequency sampling hypothesis, a "center of gravity" effect, consisting of averaging samples of instantaneous frequency, to account for the perception of vowels as well as CV and V_1V_2 formant transitions. Trying to account for the present results by postulating a frequency sampling/averaging runs into the problem that the length of the trajectory perceived to reach the target is at times three-to-four pulse durations shorter than the complete trajectory, i.e., frequencies contributing to the percept are not part of the sample. Thus, a frequency differentiation, i.e., velocity detector, explanation could be given more credence. Nevertheless, in either case the results appear to be interpretable by invoking auditory processing. The fact that the single-formant nonspeech stimuli of the control experiment also induced overestimation of the trajectory (although to a lesser degree than the synthesized four-formant speech stimuli in the main experiments) supports an interpretation that does not need to rely on linguistic processes.

Auditory processing also constitutes the cornerstone of the spectral contrast theory that explains effects of coarticulation and other stimulus context on phonetic labeling (Leek *et al.*, 1987; Liberman *et al.*, 1957; Lotto *et al.*, 1997). This theory could also cover the present data, except for one element: memory. Although memory in the theory is not altogether absent—spectral contrast is based on memory for spectrum within a syllabic-range time scale (as in CVC stimuli) or between different items in a perstimulatory or poststimulatory time range short enough to be part of a single trial (Holt *et al.*, 2000)—remembering the target presented to listeners prior to the block of trials in the present study would require them to have memory lasting several minutes (if the possibility of linguistic encoding is not taken into account) or be altogether part of the life-long language acquisition process (if one assumes that the target triggers the memory of a long-ago encountered phonetic exemplar). The memory issue is perhaps the most troubling when attempting to give experiment 3 a spectral contrast explanation: transitions along the /ia/ trajectory elicited different responses depending on whether the target was /a/ or /e/, which is possible only if the contrast is explicitly defined as one between the stimulus and the memory of a target heard up to 5 min earlier (128 trials times 2.5 s, the average trial duration). Therefore, spectral contrast could provide an auditory explanation for the results but only if it were to be supplemented with a long term memory process, one which is outside the realm of a purely auditory domain.

A more general perceptual explanation is also a possibility. The formant transitions having endpoints the subject must evaluate represents an instance of monotonic motion the perception of which is likely to be governed by the Gestalt principle of continuity (Koffka, 1935). The existence of such "perceptual inertia" has been demonstrated over the years mainly in the visual modality (e.g., Krekelberg and Lappe, 1999), though comparable instances in audition have also been reported. For example, when an upward frequency glide is broken by a noise burst that also marks the onset of a similar glide shifted in frequency, the first glide is perceived as if it were continuous (Nakajima *et al.*, 2000); the threshold of a probe tone forward-masked by a broad-band noise burst following a pure tone glide is lower at the frequency that lies on the extrapolated path traversing, but not physically present in, the masking noise (Crum and Hafter, 2008). Further support to the inertia notion is provided by results showing high sensitivity to acceleration of a sinusoid's frequency change (Divenyi, 2005), consistent with the significant acceleration component in the model of Eq. (2). It is therefore possible that the perceptual extrapolation of formant transitions toward an expected target is just part of an ensemble of processes through which the nervous system enables a person to function in a dynamically changing environment.

C. Perception of VR: An auditory explanation?

Despite its plausibility, there are two problems with an exclusively auditory explanation of VR perception. First, an exclusively auditory interpretation of the results of the control experiment is tenuous because the overestimation of vowel trajectories was consistently larger than overestimation of single-formant glides which, in one of the conditions, was altogether missing. Although these differences could be due to nonlinguistic auditory processes—the target, presented pre-block, could have been kept in memory regardless of whether it was a vowel or a buzz having a given spectral prominence—an auditory explanation would require the simultaneous existence of a memory that is better for vowels than for buzzes.[9] This memory requirement may become a bigger problem if one assumes that (i) compensation for VR routinely occurs and (ii) Eq. (2) representing VR magnitude also holds when a listener is presented real-life speech, for how else would the listener know what the putative target is without having access to his/her stored language references? Second, noting that the vowel-plus-transition stimuli are present in the repertoire of speech sounds generated by articulatory gestures, one cannot exclude the possibility that, at a higher cortical level the stimuli transmitted by the auditory system will evoke the very gestures that would have to be produced by the articulatory mechanism. That is, an auditory model would have to explain activity in the speech-motor cortical areas during speech perception, documented by recent fMRI and TMS studies (D'Ausilio *et al.*, 2009; Fadiga *et al.*, 2002; Pulvermuller *et al.*, 2006). Because the mechanisms of speech production and of speech perception have long been considered to be optimally matched complementary systems (de Cordemoy, 1668; Guenther *et al.*, 1998), an articulatory-gestural interpretation of the VR-bound percep-

tual overshoot should represent a complement, rather than an alternative, to an auditory-acoustic interpretation. Viewing the two not as opposite but as complementary has already led to discoveries by computational speech scientists for the development of articulatory-based low rate speech coding (Atal *et al.*, 1978) and for automatic speech recognition (Rose *et al.*, 1994). Another area of auditory-linguistic exploration of VR perception, not yet undertaken, could be the study of the phenomenon in other languages and across languages.[10]

V. CONCLUSION

The perception of VR prolongs the perceived extent of the formant transition leading to a vowel. This phenomenon affords the perceptual system to compensate for the articulatory undershoot of a target vowel's formant frequencies that often occurs in everyday conversational speech, in rapid speech, and in diverse phonetic context. The present experiments identified some rules underlying such compensation and suggest that they may be valid not only in the diphthong-type context investigated but could also apply to the perception of CV and VC transitions. The perception of VR is consistent *both* with current understanding of auditory mechanisms involved in the processing of monotonic frequency changes *and* with a mechanism tracking articulatory gestures that produce such changes, which situates the phenomenon in the general framework of speech dynamics.

ACKNOWLEDGMENTS

The author thanks Alvin Liberman, René Carré, Jean-Sylvain Liénard, Björn Lindblom, Brian Gygi, Associate Editor Rochelle Newman, and four anonymous reviewers for their helpful comments and advice. Portions of the paper were presented at the 135th meeting of the Acoustical Society of America and the 16th International Congress on Acoustics in Seattle, WA, in June 1998. The research was supported by Grant No. R01-07998 from National Institute on Aging, a grant from the Air Force Office of Scientific Research, and by the Veterans Administration Biomedical Laboratory Research.

[1]For example, de Cordemoy, 1668.

[2]Although Nabelek and co-workers studied the effect of transition duration and transition velocity on the boundary for /ɑ/ to /ɑI/ diphthongization (Nabelek *et al.*, 1994), their main focus was the effect of noise and reverberation in normal-hearing and hearing-impaired listeners and included only a limited parametric investigation of the trajectory, duration, and velocity of formant transitions.

[3]Although these vowel combinations sound like diphthongs, they do not necessarily figure in the repertory of diphthongs in American English. They were chosen because they encompass the longest possible trajectories in the vowel space and thus best allow for studying the perception of transitions along those trajectories.

[4]We say "fair certainty" because this boundary is stricter than the 50% subjective equality criterion most often used for the definition of phonetic category shifts, including the one by Nabelek and colleagues (Nabelek *et al.*, 1993) in their study on the shift of vowel identity in CVC settings.

[5]In their paper, Kewley-Port and Watson (1994) measured discrimination of either F1 or F2 in synthetic vowels. Although they proposed a two-tier function of the formant difference limen (DL) as a function of formant frequency, with the DL remaining constant up to about 1100 Hz and the Weber fraction constant thereafter, the variability of the data permit representing the DL as a linear function of formant frequency, i.e., as a constant Weber fraction.

[6]Greenwood, perhaps more than any other researcher, devoted much time and effort to defining and clarifying the auditory bandwidth's dependence on frequency, debunking on his way several widely accepted theories, such as the Mel scale (e.g., Greenwood, 1997).

[7]The symbol $\mathbf{\Delta}\boldsymbol{F}$ is to be interpreted as a matrix containing all salient formants, specifically F1 and F2.

[8]Other window durations were also tried but, when inserted in the various models, they accounted for less of the variance and resulted in a lower Durbin–Watson index of fit.

[9]Unfortunately, the subjects were not asked how they remembered the target, so the possibility exists that they could have used a phonetic image.

[10]The vowel space used in the present study may differ from those of other languages, such as the relatively restricted space of Arabic or the complex space of Swedish, to cite only two examples. It is, therefore, quite likely that changing the stimulus space to one with Mandarin vowels and repeating the study in Beijing with Mandarin speakers and listeners (L2-L2), or using native Mandarin speaking listeners with the present stimuli (L1-L2), would have yielded different results. However, since the study's aim was to characterize the *mechanism* responsible for the perception of transitions, and since such mechanism is likely to be shared by people regardless of their language background, it is likely that the conclusions to which such L2-L2 or L1-L2 results would point would be similar to those proposed here. The generalization should hold even for languages not having any V1-V2 transitions because the present results point to the duration-velocity field of CV and VC transitions present in practically every language.

Anderson, C. J., and Neill, W. T. (**2002**). "Two bs or not two Bs? A signal detection theory analysis of repetition blindness in a counting task," Percept. Psychophys. **64**, 732–740.

Andruski, J. E., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels and 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.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique," J. Acoust. Soc. Am. **63**, 1535–1555.

Brady, P. T., House, A. S., and Stevens, K. N. (**1961**). "Perception of sounds characterized by a rapidly changing resonant frequency," J. Acoust. Soc. Am. **33**, 1357–1362.

Carré, R., and Divenyi, P. L. (**2000**). "Modeling and perception of "gesture" reductions," Phonetica **57**, 152–169.

Carré, R., Pellegrino, F., and Divenyi, P. (**2007**). "Speech dynamics: Epistemological aspects," paper presented at the 16th International Congress of Phonetic Sciences, Saarbrucken, Germany.

Chistovich, L. A. (**1985**). "Central auditory processing of peripheral vowel spectra," J. Acoust. Soc. Am. **77**, 789–805.

Chistovich, L. A., Sheikin, R. L., and Lublinskaya, L. L. (**1979**). "Centres of gravity and spectral peaks as the determinants of vowel quality," in *Frontiers of Speech Communication Research*, edited by B. L. a. S. Ohman (Academic, London), pp. 143–157.

Crum, P. A., and Hafter, E. R. (**2008**). "Predicting the path of a changing sound: Velocity tracking and auditory continuity," J. Acoust. Soc. Am. **124**, 1116–1129.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (**2009**). "The motor somatotopy of speech perception," Curr. Biol. **19**, 381–385.

David, S. V., Mesgarani, N., Fritz, J. B., and Shamma, S. A. (**2009**). "Rapid synaptic depression explains nonlinear modulation of spectro-temporal tuning in primary auditory cortex by natural stimuli," J. Neurosci. **29**, 3374–3386.

de Cordemoy, G. (**1668**). *Discours physique de la parole (Physical Discourse on Speech)* (Bibliothèque nationale de France, Paris, France).

Di Benedetto, M.-G. (**1994**). "Acoustic and perceptual evidence of a complex relation between F1 and F0 in determining vowel height," J. Phonetics **22**, 205–224.

Divenyi, P. L. (**2005**). "Frequency change velocity detector: A bird or a red herring?," in *Auditory Signal Processing: Physiology, Psychology and Models*, edited by D. Pressnitzer, A. d. Cheveigné, S. McAdams, and L. Collet (Springer-Verlag, New York), pp. 176–184.

Dooley, G. J., and Moore, B. C. (**1988**). "Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change," J. Acoust. Soc. Am. **84**, 1332–1337.
Egan, J. P. (**1975**). *Signal Detection Theory and ROC-Analysis* (Academic, New York).
Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (**1959**). "Operating characteristics determined by binary decisions and by ratings," J. Acoust. Soc. Am. **31**, 778–773.
Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (**2002**). "Speech listening specifically modulates the excitability of tongue muscles: A TMS study," Eur. J. Neurosci. **15**, 399–402.
Feth, L. L., Fox, R. A., Jacevicz, E., and Iyer, N. (**2006**). "Dynamic center-of-gravity effects in consonant-vowel transitions," in *Dynamics of Speech Production and Perception*, edited by P. L. Divenyi, S. Greenberg, and G. Meyer (IOS, Amsterdam, The Netherlands), pp. 103–111.
Flege, J. E. (**1988**). "Effects of speaking rate on tongue position and velocity of movement in vowel production," J. Acoust. Soc. Am. **84**, 901–916.
Fowler, C. A. (**1994**). "Speech perception: Direct realism theory," in *Encyclopedia of Language and Linguistics*, edited by R. E. Ashler (Pergamon, New York), pp. 4199–4203.
Fox, R. A. (**1989**). "Dynamic information in the identification and discrimination of vowels," Phonetica **46**, 97–116.
Furui, S. (**1986**). "On the role of spectral transition for speech perception," J. Acoust. Soc. Am. **80**, 1016–1025.
Gottfried, T. L., Miller, J. L., and Payton, P. E. (**1990**). "Effect of speaking rate on the perception of vowels," Phonetica **47**, 155–172.
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—–29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
Greenwood, D. D. (**1997**). "The Mel scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios," Hear. Res. **103**, 199–224.
Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). "A theoretical investigation of reference frames for the planning of speech movements," Psychol. Rev. **105**, 611–633.
Halle, M., and Stevens, K. N. (**1962**). "Analysis by synthesis," *Proceedings on the Seminar on Speech Compression and Processing* (USAF Cambridge Research Center, Cambridge, MA), AFCRC-TR-59-198, Vol. **2**, p. D7.
Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.
Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.
Hirsh, I. J. (**1974**). "Temporal order and auditory perception," in *Sensation and Measurement: Papers in Honor of S. S. Stevens*, edited by H. R. Moskowitz, B. Scharf, and J. C. Stevens (D. Reidel, Dordrecht, The Netherlands), pp. 251–258.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2000**). "Neighboring spectral content influences vowel identification," J. Acoust. Soc. Am. **108**, 710–722.
Jenkins, J. J., Strange, W., and Miranda, S. (**1994**). "Vowel identification in mixed-speaker silent-center syllables," J. Acoust. Soc. Am. **95**, 1030–1694.
Jenkins, J. J., Strange, W., and Trent, S. A. (**1999**). "Context-independent dynamic information for the perception of coarticulated vowels," J. Acoust. Soc. Am. **106**, 438–448.
Kay, R. H., and Matthews, D. R. (**1972**). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," J. Physiol. (London) **225**, 657–677.
Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.
Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**, 2945–2958.
Kewley-Port, D., Burkle, T. Z., and Lee, J. H. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.
Koffka, K. (**1935**). *Principles of Gestalt Psychology* (Harcourt Brace Jovanovich, New York).
Krekelberg, B., and Lappe, M. (**1999**). "Temporal recruitment along the trajectory of moving objects and the perception of position," Vision Res. **39**, 2669–2679.
Leek, M. R., Dorman, M. F., and Summerfield, Q. (**1987**). "Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 148–154.
Liberman, A. L., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," J. Exp. Psychol. **54**, 358–368.
Lienard, J. S., and Di Benedetto, M. G. (**1999**). "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422.
Lindblom, B. (**1983**). "Economy of speech gesture," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 217–246.
Lindblom, B., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.
Lindblom, B., Brownlee, S. A., and Lindgren, R. (**1996**). "Formant undershoot and speaking styles: An attempt to resolve some controversial issues," in *Sound Patterns of Connected Speech, Models and Explanation*, edited by A. P. Simpson and M. Pätzold (Institut für Phonetik und Digitalsprachverarbeitung, Universität Kiel, Kiel, Germany), pp. 119–130.
Lotto, A. J., and Holt, L. L. (**2006**). "Putting phonetic context effects into context: a commentary on Fowler (2006)," Percept. Psychophys. **68**, 178–183.
Lotto, A. J., Holt, L. L., and Kluender, K. R. (**1997**). "Effect of voice quality on perceived height of English vowels," Phonetica **54**, 76–93.
Macchi, M. J. (**1980**). "Identification of vowels spoken in isolation versus vowels spoken in consonantal context," J. Acoust. Soc. Am. **68**, 1636–1642.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge, UK).
Moore, B. C. J. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.
Morrison, G. S., and Nearey, T. M. (**2007**). "Testing theories of vowel inherent spectral change," J. Acoust. Soc. Am. **122**, EL15-EL22.
Mrayati, M., Carré, R., and Guérin, B. (**1988**). "Distinctive Region and Modes: a new theory of Speech Production," Speech Commun. **7**, 257–286.
Nabelek, A. K., Czyzewski, Z., and Crowley, H. J. (**1993**). "Vowel boundaries for steady-state and linear formant trajectories," J. Acoust. Soc. Am. **94**, 675–687.
Nabelek, A. K., Czyzewski, Z., and Crowley, H. J. (**1994**). "Cues for perception of the diphthong /aI/ in either noise or reverberation. Part I. Duration of the transition," J. Acoust. Soc. Am. **95**, 2681–2693.
Nabelek, A. K., and Ovchinnikov, A. (**1997**). "Perception of nonlinear and linear formant trajectories," J. Acoust. Soc. Am. **101**, 488–497.
Nabelek, A. K., Ovchinnikov, A., Czyzewski, Z., and Crowley, H. J. (**1996**). "Cues for perception of synthetic and natural diphthongs in either noise or reverberation," J. Acoust. Soc. Am. **99**, 1742–1753.
Nakajima, Y., Sasaki, T., Kanafuka, K., Miyamoto, A., Remijn, G., and ten Hoopen, G. (**2000**). "Illusory recouplings of onsets and terminations of glide tone components," Percept. Psychophys. **62**, 1413–1425.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Nearey, T., and Assman, P. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Peterson, G., and Barney, H. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Pollack, I. (**1968**). "Detection of rate of change of auditory frequency," J. Exp. Psychol. **77**, 535–541.
Pols, L. C. W., and van Son, R. J. J. H. (**1993**). "Acoustics and perception of dynamic vowel segments," Speech Commun. **13**, 135–147.
Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., and Shtyrov, Y. (**2006**). "Motor cortex maps articulatory features of speech sounds," Proc. Natl. Acad. Sci. U.S.A. **103**, 7865–7870.
Rose, R. C., Schroeter, J., and Sondhi, M. M. (**1994**). "An investigation of the potential role of speech production models in automatic speech recognition," paper presented at the ICSLP-1994, Yokohama, Japan.
Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: an acoustical study," J. Speech Hear. Res. **6**, 111–128.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Strange, W., and Bohn, O. S. (**1998**). "Dynamic specification of coarticulated German vowels: perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.
Strange, W., Edman, T. R., and Jenkins, J. J. (**1979**). "Acoustic and phono-

logical factors in vowel identification," J. Exp. Psychol. Hum. Percept. Perform. **5**, 643–656.
van Son, R. J. J. H., and Pols, L. C. W. (**1993**). "Vowel identification as influenced by vowel duration and formant track shape," paper presented at the Proceedings of Eurospeech-93, Berlin.
van Wieringen, A. (**1995**). "Perceiving dynamic speechlike sounds," Doctoral dissertation, University of Amsterdam, the Netherlands.
van Wieringen, A., and Pols, L. C. W. (**1994**). "Frequency and duration discrimination of short first-formant speechlike transitions," J. Acoust. Soc. Am. **95**, 502–511.
van Wieringen, A., and Pols, L. C. W. (**1995**). "Discrimination of single and complex consonant-vowel- and vowel-consonant-like formant transitions," J. Acoust. Soc. Am. **98**, 1304–1312.
Vishwanathan, N., Fowler, C. A., and Magnuson, J. S. (**2009**). "A critical examination of the spectral contrast account of compensation for coarticulation," Psychon. Bull. Rev. **16**, 74–79.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., and Iskarous, K. (**2004**). "Vowel production and perception: Hyperarticulation without a hyperspace effect," Lang Speech **47**, 155–174.

A perceptual equivalent of the labial-coronal effect in the first year of life

Thierry Nazzi,[a] Josiane Bertoncini, and Ranka Bijeljac-Babic
Laboratoire Psychologie de la Perception, CNRS, Université Paris Descartes, 75006 Paris, France

(Received 11 March 2009; revised 13 May 2009; accepted 30 May 2009)

Several studies have investigated infants' acquisition of the phonological (prosodic or phonotactic) regularities of their native language at the lexical level, by showing that infants around 9/10 months of age start preferring lists of words that have a more versus less frequent phonological structure. The present study investigates whether a similar acquisition pattern of preferences can be found for labial-coronal (LC) words over coronal-labial (CL) words, a bias classically interpreted in terms of production constraints but that could also be explained in terms of relative frequency of frequent LC and less frequent CL words in many languages including French, the language used here. Results show that a preference for bisyllabic LC words emerges between 6 and 10 months of age in French-learning infants (Experiment 1), and that the non-preference at 6 months is not due to the infants' inability to discriminate the two lists of words (Experiment 2). The present study thus establishes an early perceptual equivalent of the LC bias initially found at the onset of word production. Implications of this finding for an understanding of the perception-production relationship are discussed. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158931]

PACS number(s): 43.71.Ft [RSN] Pages: 1440–1446

I. INTRODUCTION

A great deal of attention has been devoted in the past decades to specify how infants' initial language-general abilities change into abilities that are attuned to the language they are acquiring. These studies have given a better picture of the main steps in phonological acquisition (for a review see Jusczyk, 1997). The present study was conducted in this perspective, and aimed at establishing whether, early in life, infants have a perceptual equivalent of the labial-coronal (LC) bias. This LC bias was first found in early word production studies, infants initially producing more LC than coronal-labial (CL) words. While initially interpreted in terms of phonological constraints of markedness (Ingram, 1974), this bias is now usually interpreted in terms of production constraints (MacNeilage and Davis, 2000). The present study will evaluate the possibility that perceptual learning might contribute to the explanation of this bias. Before presenting data on the LC bias and the rationale of the present study in more detail, they first review the literature on early phonological acquisition.

Numerous studies attest the acquisition of native language properties during the first year of life. Prosodic acquisition is suggested by the fact that language discrimination for languages of the same rhythmic class is limited to pairs including the native language around 4/5 months of age (Bosch and Sebastian-Galles, 1997; Nazzi *et al.*, 2000). It is also reflected in the emergence of various preference biases around the same age: for words in the native language over foreign words when the words have different prosodic properties (by 6 months, Jusczyk *et al.*, 1993b), and for words with the predominant trochaic stress pattern of their native language (between 6 and 9 months for English, Jusczyk *et al.*, 1993a; between 4 and 6 months for German, Höhle *et al.*, 2009). Finally, it is also attested by a decline in sensitivity to tone contrasts between 6 and 9 months of age in infants learning a non-tonal language, English, but not in infants learning tonal (Mandarin or Cantonese) Chinese (Mattock and Burnham, 2006).

At the segmental level, a similar acquisition pattern has been observed. Effects of the native language appear around 6 months of age for vowel perception (Kuhl *et al.*, 1992; Polka and Bohn, 1996; Polka and Werker, 1994) and around 10 months of age for consonant perception (Rivera-Gaxiola *et al.* 2005; Werker and Tees, 1984).

Lastly, a few studies have shown that between 6 and 9 months of age, infants become sensitive to the phonotactic properties of their native language, that is, to the constraints on the possible order of consecutive phonemes within words. Dutch- and English-learning infants have been shown to start preferring words in their native language (English or Dutch) than in the other language when the words presented differed only by their phonotactic properties between those two ages (Jusczyk *et al.*, 1993b). Dutch 9-month-olds have also been found to listen longer to words containing phonotactically legal clusters than illegal ones (Friederici and Wessels, 1993). A similar result was found at 10 months with Catalan-dominant Catalan-Spanish bilingual infants (Sebastián-Gallés and Bosch, 2002). Moreover, English 9- but not 6-month-olds have been found to prefer to listen to words containing frequently-occurring rather than infrequent sequences of phonemes (Jusczyk *et al.*, 1994). Phonotactic knowledge also appears to impact on the ability to discriminate word forms, as shown by crosslinguistic data on Japanese- and English-learning 6-, 12-, and 18-month-olds (Kajikawa *et al.* 2006; Mugitani *et al.*, 2007).

[a]Author to whom correspondence should be addressed. Electronic mail: thierry.nazzi@parisdescartes.fr

The present study is based on the above perceptual findings. The authors' goal was to re-explore the phenomenon of the LC bias by taking this perceptual perspective. The LC bias was first pointed out in a production study on two infants, one learning English and one learning French (Ingram, 1974). It was later confirmed in a study of a group of English-learning infants showing that they tend to produce 2.55 times more LC words than CL words, this pattern being present in nine out of the ten 12-to18-month-old infants tested, the remaining infant presenting no bias (MacNeilage *et al.*, 1999). This bias, which is not present in babbling, was interpreted in the context of the frame/content theory as evidence of biomechanical constraints on early production that would make LC words easier to produce than CL words, because LC words require minimal articulatory movements: a simple mandibular oscillation to produce the labial and then a tongue movement to produce the coronal (MacNeilage and Davis, 2000). However, more recent results suggest that this bias may not be limited to production, and raise the issue of the involvement of perception in determining this bias.

First, two studies have explored the typology of languages in order to determine whether there is an asymmetry in the number of LC and CL structures present in the lexicon of these languages. In both studies, a wide range of languages from different linguistic families were investigated: English, Estonian, French, German, Hebrew, Japanese, Maon, Quichua, Spanish, and Swahili (MacNeilage *et al.*, 1999), and Afar, Finnish, French, Kannada, Kwakw'ala, Navaho, Ngizim, Quichua, Sora, and Yup'ik (Vallée *et al.*, 2001). The results of both studies converge in showing that LC structures are more frequent than CL structures in most of these languages.

In particular, the analyses of Vallée *et al.* (2001) for French (based on the BDLex corpus, Pérennou and de Calmès, 2002), the language of the infants tested in this study, show that the LC/CL asymmetry is pervasive. It appears to be present both across onsets of consecutive syllables (ratio of 1.69 at word onset; ratio of 1.56 in all lexical positions) and between the onset and the coda of the same syllable (ratio of 2.9 for word-initial syllables; ratio of 2.29 for all syllables). Given the studies on early phonological acquisition findings reported above, that have shown that infants start preferring to listen to the phonological structures which are more frequent in their native language, it appears possible that the LC asymmetry in the French language could give rise to the emergence of a preference for LC structures in young French-learning infants.

Second, a recent perception study found that French adults hearing the continuous alternation of a labial-initial syllable and a coronal-initial syllable tend to perceive them as LC rather than CL bisyllabic sequences (Sato *et al.*, 2007). These phenomena in adult perception further raise the possibility of the existence of an equivalent of the LC bias in early perception. Finding such an early perceptual LC bias would have implications regarding the determinants of the LC bias, challenging its classic interpretation in terms of motor constraints and raising the possibility that it (partly) arises from perceptual learning (that is, the acquisition of the predominant sound patterns present in the linguistic input).

Given the data on the acquisition of the phonological properties of the native language reviewed above showing the emergence of preferences for more typical, more frequent structures in the second half of the first year of life (for the most part, between 6 and 9/10 months of age), a perceptual equivalent of the LC bias might translate into a preference for LC words over CL words in infancy. Accordingly, the goals of Experiment 1 were to determine (a) whether a perceptual LC bias can be found in early infancy and (b) whether this bias is present early in life or whether it emerges during development as a reflection of the acquisition of native language properties. In order to do so, Experiment 1 explored whether French-learning infants prefer to listen to more frequent LC sequences over less frequent CL sequences at two different ages: 6 and 10 months. The stimuli used were CVCV bisyllabic words.

II. EXPERIMENT 1

A. Method

1. Participants

Thirty-two infants from French-speaking families were tested and their data included in the analyses: 16 6-month-olds (mean age=6.18 months; range: 6.01–7.09; 10 girls, 6 boys) and 16 10-month-olds (mean age =10.16 months; range: 10.05–11.05; 10 girls, 6 boys). The data of 2 additional 6-month-olds were not included in the analyses, due to fussiness or crying. The data of 5 additional 10-month-olds were not included in the analyses: 3 infants for fussiness or crying, and 2 infants for having at least 3 orientation times in the test phase shorter than 1.5 s (this criterion was used to ensure that infants heard at least one or two words of the list).

2. Stimuli

Recordings were made in a sound-attenuated booth. A female native speaker of French recorded several tokens of 24 French bisyllabic words. Twelve words had a LC structure: 3 bVdV words (/bödo/, /bode/, /bude/), 3 pVtV words (/pote/, /patĩ/, /pitö/), 3 bVtV words (/bato/, /byte/, /butõ/), and 3 pVdV words (/padi/, /pedã/, /pãda/); and 12 words had a CL structure: 3 dVbV words (/deby/, /döbu/, /dobe/), 3 tVpV words (/tapi/, /tupe/, /topĩ/), 3 tVbV words (/tõbe/, /tabu/, /tyba/), and 3 dVpV words (/depi/, /depo/, /dopã/). Words in both lists were made up of exactly the same consonants, and vowels were almost completely balanced across lists. Two tokens of each word were selected. Overall, the duration of the LC and CL tokens was similar [541 versus 537 ms, $t(46)<1$, n.s.].

Four lists were made up: two lists with the 12 LC words (different tokens, the order of the words in the two lists being reversed) and two lists with the 12 CL words (different tokens, the order of the words in the two lists being reversed). The duration of all the lists was 18.00 s.

3. Procedure and apparatus

The experiment was conducted in a three-sided test booth made of pegboard panels. Except for a small section of

pre-existing holes in the front panel used for monitoring the infant's headturns, the panels were backed with white cardboard to prevent the infant from seeing behind the panels. The test booth had a red light and a loudspeaker (SONY xs-F1722) mounted at eye level on each of the side panels and a green light mounted on the center panel. Directly below the center light a 5 cm hole accommodated the lens of a video camera used to record each test session. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A PC computer terminal (COD) and response box were located behind the center panel, out of view of the infant. The response box, which was connected to the computer, was equipped with a series of buttons. The box was controlled by an observer hidden behind the center panel, who looked through a peephole and pressed the buttons of the response box according to the direction of the infant's headturns, thus starting and stopping the flashing of the lights and the presentation of the sounds. The observer, and also the infant's caregiver, wore earplugs and listened to masking music over tight-fitting closed headphones, which prevented them from hearing the stimuli presented. Information about the direction and duration of the headturns and the total trial duration were stored in a data file on the computer.

The classic version of the headturn preference procedure (HPP) was used in the present study (c.f. Jusczyk *et al.*, 1993a). Each infant was held on a caregiver's lap. The caregiver was seated in a chair in the center of the test booth. Each trial began with the green light on the center panel blinking until the infant had oriented in that direction. Then, the center light was extinguished and the red light above the loudspeaker on one of the side panels began to flash. When the infant made a turn of at least 30° in the direction of the loudspeaker, the stimulus for that trial began to play. Each stimulus was played to completion (i.e., when the 12 words had been presented) or stopped immediately after the infant failed to maintain the 30° headturn for 2 consecutive seconds (200 ms fade-out). The stimuli were stored in digitized form on the computer, and were delivered by the loudspeakers via an audio amplifier (Marantz PM4000). If the infant turned away from the target by 30° in any direction for less than 2 s and then turned back again, the trial continued but the time spent looking away was not included in the orientation time. Thus, the maximum orientation time for a given trial was the duration of the entire speech sample. The flashing red light remained on for the entire duration of the trial.

Each experimental session began with two musical trials, one on each side (randomly ordered) to give infants' an opportunity to practice one headturn to each side before the test session itself. The test phase consisted of three test blocks (in each of which the two LC and the two CL lists were presented). The order of the different lists within each block was randomized.

B. Results and discussion

Mean orientation times to the LC and the CL lists were calculated for each infant. The data for the two age groups are presented in Fig. 1. A two-way analysis of variance with

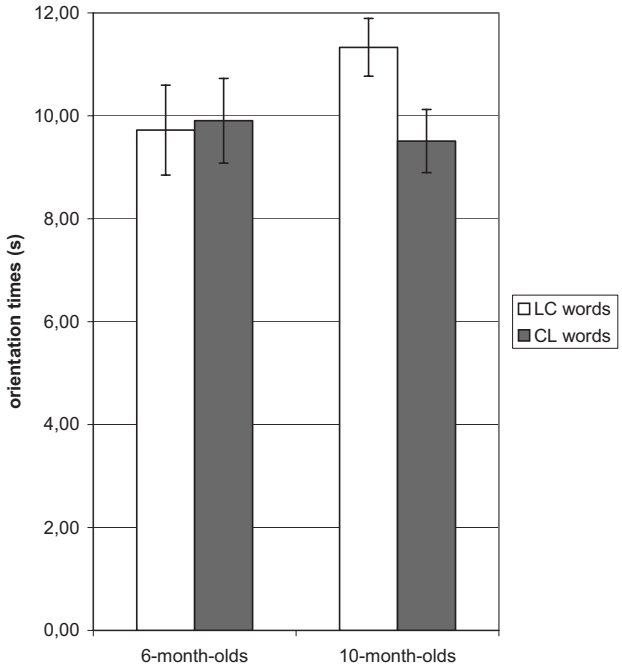


FIG. 1. Mean orientation times (s) to the bisyllabic LC words versus the bisyllabic CL words in Experiment 1. The error bars indicate the standard error of the mean. Left panel: 6-month-old infants; Right panel: 10-month-old infants.

the main between-subject factor of age (6 and 10 months) and the main within-subject factor of lexical structure (LC versus CL words) was conducted. The effect of lexical structure approached significance, $F(1,30)=4.05$, $p=.053$, indicating that the infants tended to have longer orientation times to the LC words than to the CL words. However, there was a significant age × lexical structure interaction, $F(1,30)=6.05$, $p=.020$, indicating that the effect of lexical structure changed with age. The effect of age was not significant, $F(1,30)<1$.

In order to specify the age × lexical structure interaction, planned comparisons were conducted. The effect of lexical structure failed to reach significance at 6 months $F(1,30)<1$, indicating that these infants had similar orientation times to the LC words ($M=9.72$ s, SD=3.50) and the CL words ($M=9.90$ s, SD=3.29). Only half of the 16 6-month-olds oriented longer to the LC words. However, the effect of familiarity was significant at 10 months, $F(1,30)=9.99$, $p=.004$, indicating that 10-month-old infants had longer orientation times to the LC words ($M=11.33$ s, SD =2.24) than to the CL words ($M=9.51$ s, SD=2.46). This pattern of LC preference was found for 13 of the 16 infants ($p=.01$, binomial test).

The first goal of the present experiment was to determine whether a perceptual equivalent to the LC bias is present in infancy. Accordingly, infants heard two lists of CVCV bisyllabic words, a list of varied LC words and a list of varied CL words. The present results show that 10-month-old infants prefer to listen to the LC words. Since LC words are predominant in French, this pattern is similar to that found in previous studies showing that infants around 9–10 months of age prefer to listen to types of words that are more frequent in their native language (Friederici and

Wessels, 1993; Höhle *et al.*, 2009; Jusczyk *et al.*, 1993a, 1993b, 1994; Sebastián-Gallés and Bosch, 2002). Given that these effects are found before the onset of word production, and given that the LC bias in production is found in early word production but not in babbling, this result suggests that the LC bias in production might not necessarily or at least not solely arise from production constraints, an issue further discussed later.

The second goal of the present experiment was to determine whether, if present, such a LC bias is present early in life, or whether it emerges during development. The present data support the latter possibility given the lack of a preference in 6-month-olds, and the significant difference in behavior between the two age groups. However, the significance of this developmental pattern of change would be stronger if it were possible to specify that the failure at 6 months is really due to a lack of preference between the two structures, rather than difficulties with the HPP task or the discrimination between two phonotactic patterns each exemplified by 12 phonetically-varied words.

In order to better understand the lack of a LC preference in 6-month-olds, Experiment 2 was conducted to evaluate whether they are nevertheless able to discriminate between the lists of LC and CL words used in the present experiment. This was done by familiarizing infants with one type of structure (counterbalanced across infants) for 1 min, before presenting them with trials of LC versus CL stimuli (see Höhle *et al.*, 2009, for a similar use of HPP). Note that if infants succeed in this discrimination paradigm, this will unambiguously show that the lack of a bias in the present experiment was not due to methodological or perceptual issues; however, if they fail, the interpretation of the lack of a bias will remain ambiguous.

III. EXPERIMENT 2

A. Method

1. Participants

Sixteen 6-month-old infants (mean age=6.18 months; range: 6.01–7.09; 10 girls, 6 boys) from French-speaking families were tested and their data included in the analyses. The data of one additional infant were not included in the analyses, due to fussiness.

2. Stimuli

The stimuli were the same as those used in Experiment 1 (2 files made up of different tokens of 12 LC words, and 2 files made up of different tokens of 12 CL words).

3. Procedure and apparatus

The apparatus was the same as in Experiment 1. The procedure was the same as in Experiment 1, with one crucial modification that changed the experiment from a preference to a discrimination experiment. Infants were familiarized with either a LC or a CL file until they reached a familiarization criterion of 60 s of orientation times. Half of the infants were familiarized with the LC words, the other half being familiarized with the CL words. Once the familiarization criterion was reached, infants were tested with the other file of the same structure (same words repeated in a different order), and one of the other two files of the opposite structure. These two files were repeated four times in randomized order, leading to the presentation of eight test trials, half of the same and half of the opposite structure.

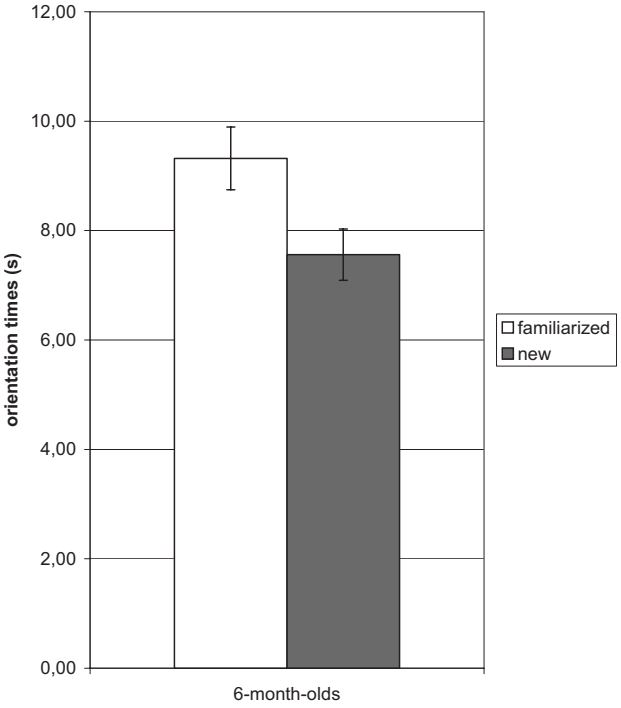


FIG. 2. Mean orientation times (s) at 6 months to the familiarized versus non-familiarized words in Experiment 2.

B. Results and discussion

Mean orientation times to the files with the familiarized and new structures were calculated for each infant. The data are presented in Fig. 2. On average, infants oriented to the sequences with the familiarized structure for 9.32 s (SD =2.30) while they oriented to the sequences with the new structure for 7.56 s (SD=1.89). This difference was significant, $t(15)=3.12$; $p=0.007$, two-tailed. 12 of the 16 infants had longer orientation times to the sequences with the familiarized structure ($p=0.038$, binomial test).

The present experiment shows that 6-month-old infants presented with a list of 12 words later prefer to listen to a different list of the same words (different tokens, different order of presentation) than to a list of 12 different words. This familiarity effect (similar to that found by Höhle *et al.*, 2009, using the exact same procedure) suggests that 6-month-old infants discriminate between LC and CL words in spite of all the phonetic variability present. Although it remains unclear whether the present effect would emerge if a different list of words with the structure used during familiarization had been presented, the present results nevertheless rule out the possibility that the lack of a preference for LC words found at 6 months in Experiment 1 was due to methodological or perceptual issues. More likely, 6-month-old infants do not have a preference for LC over CL words at that age (although preferences can be found at that age, if they

are based on prosody, Jusczyk *et al.*, 1993b, Höhle *et al.*, 2009), and the preference emerges between 6 and 10 months due to exposure to the native language.

IV. GENERAL DISCUSSION

The present results first establish that infants' preferences for some kinds of word patterns over others emerge between the ages of 6 and 10 months. Indeed, a preference for LC words over CL ones is found at 10 months of age, but not at 6 months (Exp. 1). Moreover, the fact that 6-month-olds were able to discriminate the two lists of words (Exp. 2) suggests that their lack of a preference (Exp. 1) is not due to methodological or perceptual issues. Given that LC words are more frequent than CL words in French, this pattern of results is in line with previous results showing that during the second part of the first year of life, infants start preferring the structures that are more frequent in their native language. In those previous studies, infants around 6- to -10 months of age were found to start preferring words having the predominant stress pattern of their native language (for English, Jusczyk *et al.*, 1993a, 1993b; for German, Höhle *et al.*, 2009), or being made up of more frequent phonotactic sequences (for Dutch. Friederici and Wessels, 1993; for English. Jusczyk *et al.*, 1993b, 1994; for Catalan; Sebastián-Gallés and Bosch, 2002). The present study is the first to provide data establishing the emergence of a preference for words with a structure that is predominant in the native language for French-learning infants.

The timing of this emergence is in line with results showing that infants start learning the phonetic inventory of their native language during that time period. But, like the data contrasting words on their phonotactic properties, the present results go beyond these phonetic acquisition findings. The emergence of a preference for phonotactically legal over illegal, phonotactically frequent over infrequent, and, here, LC over CL words between 6 and 10 months of age is likely to reflect some phonological acquisition regarding the constraints on how phonemes are ordered within word-form units in the native language. Thus, taken together, these results suggest that infants start preferring structures that follow a predominant pattern in their native language once they have learned that this pattern is predominant. These acquisitions are probably made possible by the onset of word-form segmentation abilities between these two ages (Gout, 2001; Höhle and Weissenborn, 2003; Houston *et al.*, 2004; Jusczyk and Aslin, 1995; Nazzi *et al.*, 2005; Nazzi *et al.*, 2006; Nazzi *et al.*, 2008).[1]

The present results also raise the issue of how the present early perception bias impacts on the classical interpretation of the LC bias in terms of articulatory constraints. On a strong version of this proposal (denying any role of perceptual processes), one would have predicted that a perceptual preference would not have emerged before infants start producing LC (and CL) structures. In theory, it might not have emerged before the onset of word production, as the LC bias has been found in early word production but not in babbling. These predictions were not supported by the present results. Indeed, 10-month-olds preferred to listen to LC words over CL words. Yet, the infants in the present study did not produce words. Moreover, an ongoing follow-up of the present study established for another group of 10-month-old infants that infants at that age do not even produce babbling in which consonants vary, that is, they do not produce LC and CL sequences (Nazzi *et al.*, 2009).

Does this mean that the LC bias is actually the sole product of the perceptual properties of the auditory/speech perception system? This position (denying any role of production processes) is also unlikely given evidence that producing LC words is easier than producing CL words. Rather, what seems to be at play here is the result of a production-perception loop in which types of words that are easier to produce due to articulatory constraints end up being more frequent in (most) languages, and consequently, become preferred at the perceptual level due to their higher frequency (an idea to be related to the notion, put forward in the motor theory of speech perception, of a coevolution of perception and production skills, c.f. Galantucci *et al.*, 2006, for a recent review).

Accordingly, the LC bias found at 10 months in the present study could be due to perceptual acquisition of input regularities reflecting articulatory constraints. Thus, the LC bias found here would be a direct perceptual effect, reflecting an indirect production effect. In the future, it would be interesting to explore this early perceptual LC bias in languages such as Japanese that, contrary to most languages, has been reported to have a small, but significant, input bias in the opposite CL direction (0.84, c.f. MacNeilage *et al.*, 1999). If this bias were to be confirmed in the input (MacNeilage *et al.*, 1999, only looked at a subset of the Japanese lexicon), then exploring the existence of a perceptual LC bias in Japanese 10-month-olds would provide a means to evaluate the respective contribution of production constraints and perceptual learning. Indeed, if production constraints predominate, then a LC bias should also be found in Japanese infants. But if the properties of the input (and the perceptual learning that follows) are the main drivers of the bias, then one predicts either no bias (due to the small size of the bias) or a coronal-labial perceptual bias in Japanese infants.

To finish, note that although the present study reveals a perceptual equivalent of the LC bias previously described in early production, the perceptual factors that give raise to the observed bias remain unspecified at this point. First, although vowels were matched as much as possible across the two lists of phonotactic patterns, perfect match was not possible due to the use of real bisyllabic words, so that the possibility that these subtle differences might have contributed to the effect observed cannot be entirely ruled out. Syllables were also different in the two lists of words, and frequency analyses later conducted on the Lexique 3 database (New *et al.*, 2004) revealed that the second syllables of the LC words were more frequent than those of the CL words, which again might have contributed to the bias observed. Note, however, that the rationale for using lists of phonetically-varied items is to elicit processing at levels beyond the acoustic/phonetic level (Bertoncini *et al.*, 1995; Bijeljac-Babic *et al.*, 1993; Jusczyk *et al.*, 1993a, 1993b). Second, although the effects observed have been discussed in

terms of the relative input frequency of the order of appearance of the labial and coronal consonants in the words presented, the present results on bisyllabic words leave open the question of whether the perceptual bias is best described in terms of the order of the consonants themselves, or the syllables they are embedded in. In the later case, the effect would be determined by the relative order of adjacent syllables, not by the relative order of non-adjacent consonants. Third, the present study does not rule out the possibility that the LC bias stems from a general preference for labial-initial words, even though labial-initial words are not more frequent than coronal-initial words in French (Lexique 3 database analyses, c.f. Nazzi *et al.*, 2009). All of these issues are currently being addressed in a parallel study exploring the emergence of a LC effect in French-learning infants using monosyllabic items, which allow for perfect vocalic match, frequency control, and can only be interpreted at the phonetic level (Nazzi *et al.*, 2009). The data show a bias for LC CVC words over CL CVC words at 10 months. Moreover, the results of a control condition in which infants were presented with recordings of the sole CV portion of the CVC items used in the main experiment failed to show any preference, reinforcing the present interpretation of the perceptual LC bias in terms of sensitivity to the relative frequency of the order of labial and coronal consonants within words.

V. CONCLUSIONS

In conclusion, the present study establishes the emergence between 6 and 10 months of age of a perceptual equivalent of the LC bias previously reported in production. This finding contributes to a growing literature showing infants' early acquisition of the phonological properties of their native language during the second half of the first year of life. The present findings thus show that the LC bias is unlikely to be the sole product of production constraints, but more likely to result from complex interactions between production constraints, the structure of the input, and early perceptual learning.

ACKNOWLEDGMENTS

This research was supported by an ACI grant "systèmes complexes" to T.N. and J.B. We would like to thank Jean-Luc Schwartz and Nathalie Vallée for fruitful discussions and help in selecting the stimuli, and the participants and their families for their time and cooperation.

[1]A similar claim has been made for the emergence of the trochaic bias (Jusczyk *et al.*, 1993a), although more recent data suggest that the trochaic bias might appear before the onset of word segmentation (Höhle *et al.*, 2009) and might therefore result from rhythmic acquisition at the sentence level rather than at the word level (Nazzi *et al.*, 2006).

Bertoncini, J., Floccia, C., Nazzi, T., and Mehler, J. (**1995**). "Morae and syllables: Rhythmical basis of speech representation in neonates," Lang Speech **38**, 311–329.

Bijeljac-Babic, R., Bertoncini, J., and Mehler, J. (**1993**). "How do 4-day-old infants categorize multisyllabic utterances?," Int. J. Geriatr. Psychiatry **29**, 711–721.

Bosch, L., and Sebastián-Gallés, N. (**1997**). "Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments," Cognition **65**, 33–69.

Friederici, A. D., and Wessels, J. M. I. (**1993**). "Phonotactic knowledge and its use in infant speech perception," Percept. Psychophys. **54**, 287–295.

Galantucci, B., Fowler, C. A., and Turvey, M. T. (**2006**). "The motor theory of speech perception reviewed," Psychon. Bull. Rev. **13**, 742–742.

Gout, A. (**2001**). "Etapes précoces de l'acquisition du lexique (Early steps in lexical acquisition)," Ph.D. thesis, Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

Höhle, B., and Weissenborn, J. (**2003**). "German-learning infants' ability to detect unstressed closed-class elements in continuous speech," Dev. Sci. **6**, 122–127.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., and Nazzi, T. (**2009**). "Language specific prosodic preferences during the first half year of life: Evidence from German and French infants," Infant Behav. Dev. **32**, 262–274.

Houston, D. M., Santelmann, L. M., and Jusczyk, P. W. (**2004**). "English-learning infants' segmentation of trisyllabic words from fluent speech," Lang. Cognit. Processes **19**, 97–136.

Ingram, D. (**1974**). "Fronting in child phonology," J. Child Lang **i**, 233–241.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Jusczyk, P. W., and Aslin, R. N. (**1995**). "Infants' detection of the sound patterns of words in fluent speech," Cogn. Psychol. **29**, 1–23.

Jusczyk, P. W., Cutler, A., and Redanz, N. (**1993a**). "Preference for the predominant stress patterns of English words," Child Dev. **64**, 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., and Jusczyk, A. M. (**1993b**). "Infants' sensitivity to the sound patterns of native language words," J. Mem. Lang. **32**, 402–420.

Jusczyk, P. W., Luce, P. A., and Charles-Luce, J. (**1994**). "Infants' sensitivity to phonotactic patterns in the native language," J. Mem. Lang. **33**, 630–645.

Kajikawa, S., Fais, L., Mugitani, R., Werker, J. F., and Amano, S. (**2006**). "Cross-language sensitivity to phonotactic patterns in infants," J. Acoust. Soc. Am. **120**, 2278–2284.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindbolm, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

MacNeilage, P. F., and Davis, B. L. (**2000**). "The motor core of speech: A comparison of serial organization patterns in Infants and languages," Child Dev. **71**, 153–163.

MacNeilage, P. F., Davis, B. L., Kinney, A., and Matyear, C. L. (**1999**). "Origin of serial-output complexity in speech," Psychol. Sci. **10**, 459–460.

Mattock, K. and Burnham, D. (**2006**). "Chinese and English infants' tone perception: Evidence for perceptual reorganization," Infancy **10**, 241–265.

Mugitani, R., Fais, L., Kajikawa, S., Werker, J. F., and Amano, S. (**2007**). "Age-related changes in sensitivity to native phonotactics in Japanese infants," J. Acoust. Soc. Am. **122**, 1332–1335.

Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (**2000**). "Language discrimination by English learning 5-month-olds: Effects of rhythm and familiarity," J. Mem. Lang. **43**, 1–19.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hunagel, S., and Jusczyk, P. W. (**2005**). "English-learning infants' segmentation of verbs from fluent speech," Lang Speech **48**, 279–298.

Nazzi, T., Gonzalez-Gomez, N., Bijeljac-Babic, R., and Bertoncini, J. (**2009**). "Learning of non-adjacent phonological dependencies in the first year of life," poster presented at the *Biennal Meeting of Society for the Research on Child Development*, Denver CO.

Nazzi, T., Iakimova, I., Bertoncini, J., Frédonie, S., and Alcantara, C. (**2006**). "Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences," J. Mem. Lang. **54**, 283–299.

Nazzi, T., Mersad, K., Iakimova, G., Sundara, M., and Polka, L. (**2008**). "Differences in the development of speech segmentation abilities in two French dialects," paper presented at the *11th International Congress for the Study of Child Language (IASCL)*, Edinburgh, United Kingdom.

New, B., Pallier, C., Brysbaert, M., and Ferrand, L. (**2004**). "Lexique 2: A New French Lexical Database," Behav. Res. Methods Instrum. Comput. **36**, 516–524.

Perennou, G., and de Calmes, M. (**2001**). "Ressources lexicales texte et parole à l'IRIT (Text and speech lexical resources at the Institut de Recherche en Informatique de Toulouse)," La lettre d'information ELRA **6**, 8–10.

Polka, L., and Bohn, O. S. (**1996**). "Cross-language comparison of vowel perception in English-learning and German-learning infants," J. Acoust. Soc. Am. **100**, 577–592.

Polka, L., and Werker, J. F. (**1994**). "Developmental changes in perception of nonnative vowel contrasts," J. Exp. Psychol. Hum. Percept. Perform. **20**, 421–435.
Rivera-Gaxiola, M., Silva-Pereyra, J., and Kuhl, P. K. (**2005**). "Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants," Dev. Sci. **8**, 162–172.
Sato, M., Vallée, N., Schwartz, J.-L., and Rousset, I. (**2007**). "A perceptual correlate of the labial-coronal effect," J. Speech Lang. Hear. Res. **50**, 1466–1480.
Sebastián-Gallés, N., and Bosch, L. (**2002**). "Building phonotactic knowledge in bilinguals: Role of early exposure," J. Exp. Psychol. Hum. Percept. Perform. **28**, 974–989.
Vallée, N., Rousset, I., and Boë, L. J. (**2001**). "Des lexiques aux syllabes des langues du monde. Typologies, tendances et organisations structurelles (From the lexicons to the syllables of the languages of the world. Typologies, tendencies and structural organizations)," Linx **45**, 37–50.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels

Ron I. Thomson[a)]
Department of Applied Linguistics, Brock University, St. Catharines, Ontario L2S 3A1, Canada

Terrance M. Nearey
Department of Linguistics, University of Alberta, 4-32 Assiniboia Hall, Edmonton, Alberta T6G 2E7, Canada

Tracey M. Derwing
Department of Educational Psychology, University of Alberta, 6-102 Education North, Edmonton, Alberta T6G 2G5, Canada

(Received 17 June 2008; revised 15 May 2009; accepted 15 June 2009)

This study describes a statistical approach to measuring crosslinguistic vowel similarity and assesses its efficacy in predicting L2 learner behavior. In the first experiment, using linear discriminant analysis, relevant acoustic variables from vowel productions of L1 Mandarin and L1 English speakers were used to train a statistical pattern recognition model that simultaneously comprised both Mandarin and English vowel categories. The resulting model was then used to determine what categories novel Mandarin and English vowel productions most resembled. The extent to which novel cases were classified as members of a competing language category provided a means for assessing the crosslinguistic similarity of Mandarin and English vowels. In a second experiment, L2 English learners imitated English vowels produced by a native speaker of English. The statistically defined similarity between Mandarin and English vowels quite accurately predicted L2 learner behavior; the English vowel elicitation stimuli deemed most similar to Mandarin vowels were more likely to elicit L2 productions that were recognized as a Mandarin category; English stimuli that were less similar to Mandarin vowels were more likely to elicit L2 productions that were recognized as new or emerging categories. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3177260]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.An, 43.71.Bp [AJ] Pages: 1447–1460

I. INTRODUCTION

Given the importance of crosslinguistic similarity in dominant models of L2 speech perception (Best, 1995; Flege, 1995), it is surprising that the operationalization of this construct has not received more deliberate attention. In this study, the authors introduce a statistical approach to measuring crosslinguistic vowel similarity that provides a useful complement to other methods. Not only is this approach capable of empirically capturing the degree to which individual instantiations of L2 vowel categories from one language are phonetically similar to those in another, but it can also determine the extent to which L2 accented productions fit L1 category distributional properties vis-à-vis distributional properties of target L2 categories. This method extends the statistical modeling of L1 vowel perception by Nearey and Assmann (1986), versions of which have also been employed in L2 studies by Strange *et al.* (2004), Thomson (2005), Morrison (2006), and Strange (2007).

A. Models of crosslinguistic speech perception

Best's (1995) perceptual assimilation model (PAM) and Flege's (1995) speech learning model (SLM) have provided the impetus for many studies investigating crosslinguistic speech perception and production. While both models are concerned with how adults process sounds in a nonnative language, they differ in their foci (Best and Tyler, 2007). PAM is a static model of crosslinguistic speech perception, primarily addressing how monolingual speakers of one language perceive sounds from an unfamiliar language. In contrast, the SLM describes dynamic patterns of L2 phonological learnability, predicting which L2 categories will be easier to learn on the basis of their interaction with L1 categories.

Both models assume that the relative similarity of L1 and L2 categories predicts listeners' perception of L2 categories. PAM predicts that L2 categories most similar to L1 categories will perceptually assimilate to the corresponding L1 category. Sometimes two or more L2 categories may assimilate to the same L1 category, leading to confusion. PAM also predicts that when L2 categories are sufficiently different from L1 categories, they will be perceived as uncategorizable speech sounds. The SLM makes predictions in terms of "similar" and "new" L2 sounds. Similar sounds are perceived as members of an L1 category, thus providing no motivation for learning a new L2 category. In contrast, new sounds are perceived as different from any L1 categories, meaning learners will be motivated to develop a new L2 category. Both PAM and the SLM acknowledge that different

[a)]Author to whom correspondence should be addressed. Electronic mail: rthomson@brocku.ca

tokens of the same L2 sound category may be recognized as better or worse members of the L1 category to which they assimilate. In practice, however, research within both paradigms usually groups target L2 tokens in terms of entire categories, averaging raw phonetic differences across all tokens being tested or learned.

An alternative view is that phonetic variability within target L2 categories is of central importance and deserves greater attention. Phonetic variation in L2 input is important because each target L2 production interacts uniquely with learners' relatively stable L1 categories. One way of incorporating these interactions into models of L2 speech perception is to define crosslinguistic similarity as the proportion of tokens from a target L2 category that assimilate to L1 categories. While on a macrolevel the resulting predictions are similar to those proposed in previous research, the interactions of individual L2 tokens with L1 categories are emphasized. For example, where PAM describes direct category assimilation, this view predicts that a very high proportion of L2 tokens will be assimilated by an L1 category. Likewise, using SLM terminology, this view predicts that a high proportion of L2 tokens from similar categories will be perceived as corresponding L1 categories, while a very small proportion of L2 tokens from new categories will be perceived as L1 categories. Furthermore, it is the proportion of L2 tokens that are not perceived as belonging to an L1 category that determines the learnability of L2 categories. When few tokens of an L2 category are perceived as being outside the distribution of an L1 category, learning is difficult, because the learner has limited evidence for not relying on the closest L1 category. In cases where many or most tokens of an L2 category are recognizably distinct from any L1 category, the learner will ultimately develop a new category that reflects the distributional properties found in the target L2.

B. Measuring crosslinguistic similarity

In order to investigate individual differences between L2 tokens and their unique interactions with learners' L1 categories, a precise means of measurement is essential. Researchers evaluating PAM and the SLM typically use empirical approaches such as perceptual mapping to measure crosslinguistic similarity. In this approach, native speakers of one language listen to stimuli from another language and identify them in terms of L1 categories. Listeners may also provide scalar judgments regarding how well tokens fit the nearest L1 category. Perceptual mapping has been used to measure English listeners' perception of German vowels (Polka, 1995), Japanese speakers' perception of English vowels (Strange *et al.*, 1998) and English consonants (Guion *et al.*, 2000), Korean speakers' perception of English consonants (Schmidt, 1996), and English speakers' perception of Korean consonants (Schmidt, 2007).

In the study of Guion *et al.* (2000), predictions stemming from a perceptual mapping experiment of English consonants by Japanese listeners were tested against L2 learner behavior using an AXB discrimination task. In this task, listeners hear a sequence of three stimuli and are asked to indicate whether the second item (X) is more similar to the first (A) or the third (B) item. Guion *et al.* (2000) expected that experienced learners of English, having had far more exposure, would better discriminate perceptually less similar English consonants than inexperienced learners. Instead, they found no group differences. To account for this, the researchers speculated that perceptual mapping might not be sensitive enough to measure crosslinguistic similarity, meaning some predictions were invalid. Another possible explanation for the findings of Guion *et al.* (2000) may be that perceptual mapping experiments and AXB discrimination tasks are incommensurable. In a recent study, Wayland (2007) found that the A and B tokens in an AXB design influenced perception of the X token, but when presented in isolation, as is usual in an identification task, the same X token was perceived differently.

Beyond the complexities just outlined, the perceptual mapping approach to measuring crosslinguistic similarity also suffers from practical limitations. It requires a large number of speaker productions in order to reflect the variability found in the population. In practice, the number of speakers used has been quite small, from as low as three per language (Flege *et al.*, 1994) to nine (Guion *et al.*, 2000). However, using larger numbers would make identification, rating, and discrimination tasks extremely onerous for listeners.

Given the theoretical and practical constraints of using human judges, another method, used for measuring crosslinguistic similarity of vowels, is to determine the extent of overlap between the spectral properties of vowels produced in the L1 and the L2. For example, Flege (1995) reported that Spanish speakers' productions of English /i/ exhibit substantial spectral overlap in an F1-F2 space with native English /i/ and /ɪ/. This evidence is used to support claims that Spanish /i/ subsumes both English /i/ and /ɪ/. Likewise, Bohn and Flege (1992) compared English /i/, /ɪ/, and /ɛ/ with German /i/, /ɪ/, /ɛ/, and /ɛː/ by plotting the distributions of these vowels produced by native speakers of each language.

As with perceptual mapping experiments, relying on rough spectral properties to determine the crosslinguistic similarity of vowels has yielded informative results. When applied in the manner just described, however, this method lacks precision. For example, it usually represents vowels in a two-dimensional space, meaning only a limited number of acoustic correlates of vowel perception are incorporated (e.g., only F1 and F2 at midpoints). Consequently, spectral change within vowels is typically ignored, despite its importance in perception (see Nearey and Assmann, 1986). In addition, such approaches often exclude F3, pitch, and vowel duration, all of which can influence vowel perception.

One way of incorporating multidimensionality into phonetic measurement of L1 vowels is to use discriminant analysis, in which acoustic measures (e.g., F1, F2, F3, pitch, and duration) from known vowels are used to train a statistical pattern recognition model that predicts group membership of novel cases. During the training phase, multivariate normal distributions (with a common covariance matrix) are fitted to known vowel categories in the training set. Using these fitted distributions, acoustic measures of new tokens can be assessed to determine to what category the new tokens are

most similar. The results of such models have been shown to correlate highly with vowel categorizations by human listeners of several English dialects (Assmann *et al.*, 1982; Assmann and Katz, 2000; Hillenbrand and Nearey, 1999; Nearey and Assmann, 1986), suggesting that this approach offers a parametric representation of English vowels that reasonably approximates human vowel perception under ideal circumstances. It may not, however, account for a human listener's response to speech in noisy conditions, or a lexically or contextually motivated bias for listeners to categorize a production in a particular direction.

The statistical pattern recognition approach has also been applied to the measurement of crosslinguistic vowel similarity. For example, Morrison (2006) used discriminant analysis to measure the similarity of Spanish and English vowels. First, using acoustic measurements from L1 English vowel productions, he built a statistical model that defined English vowel categories and then tested Spanish vowel productions against this model. He also reversed the process to determine how English vowels would be classified by a Spanish pattern recognition model. The results may reflect how Spanish and English learners might perceive the opposing language's vowels. Strange *et al.* (2004) and Strange (2007) used a similar approach to classify French and German vowel productions in terms of American English categories, while Thomson (2007) applied this approach to a comparison of English and Mandarin vowels.

Unfortunately, such pattern recognition models have crucial limitations for measuring crosslinguistic similarity. First, while they classify vowels in one language into categories in a competing language, they do not indicate *how well* a production of a vowel in one language fits a category in the other. That is, all productions that are recognized as a particular category are treated as though they equally belong. These models also assume that L2 learners perceive all L2 vowel productions in terms of L1 categories, which is not necessarily the case. Such limitations may partially explain why Strange *et al.* (2004) found that the classification of North German vowels by their American English pattern recognition model did not uniformly predict the perceptual identification of the same North German vowels by American English listeners.

C. A statistical "metamodel" approach to measuring L1/L2 vowel similarity

There are undoubtedly a myriad of ways the discriminant analysis model might be enhanced. This study uses a modified statistical pattern recognition approach (henceforth referred to as the metamodel) that the authors believe allows for more accurate measurement of crosslinguistic similarity. Unlike the single-language pattern recognition models described in Sec. I B, the metamodel incorporates categories from *both* languages being contrasted. After training the metamodel on all relevant categories from both languages, production values of new cases from each language are tested. The extent to which new cases from one language are misclassified as members of a category in the opposing language provides a means of operationalizing crosslinguistic similarity. For example, if the general distribution of a vowel category in one language is truly *identical* to the distribution of a category in the competing language, the authors would expect the metamodel to classify 50% of new cases of the category in one language as members of the corresponding category in the competing language, and vice versa. If the distribution of two categories is not identical, but very *similar*, the authors would expect some misclassification of tokens as members of the competing language category, but less than 50%.

In addition to crisp, absolute classification of new cases, statistics used in the metamodel can also reveal how well a new case fits a competing language category, providing a goodness of fit measure. In discriminant analysis, the assignment of a new case to a specific category is based on the *a posteriori probability* (APP) of membership. For example, a new case may be classified as a member of the intended language category with an APP of only 0.51, while its likelihood of belonging to a similar category in the competing language may be 0.49. The APP scores reveal that such a case, despite being classified as the intended vowel, does not fit the intended category as well as a case with an APP of 0.99, for example. When comparing two vowel systems, the authors would expect that vowel tokens that are accurately classified as the intended language category in absolute terms should also have some probability of being members of similar categories in the opposing language. (See Hillenbrand and Nearey, 1999 for more details of APP scores.)

As well as providing a measure of crosslinguistic similarity, the metamodel may also be used to measure the degree to which accented L2 productions approach native speaker targets. Studies of L2 phonological development have typically relied on intelligibility scores obtained from native speaker listeners (e.g., Munro *et al.*, 2003; Munro and Derwing, 2008). While this approach indisputably reflects native speakers' conscious responses to accent, it can also lack precision. For example, human listeners may perceptually assimilate some accented speech sounds to their own phonetically similar L1 categories. Levi *et al.* (2007) demonstrated that lexical context may also cause some L2 segments to be perceived as less accented than they really are. In such cases, although phonetic differences are present, they may be perceptually undetectable to listeners. Testing L2 productions against the metamodel may indicate whether L2 learners are simply substituting L1 sounds for L2 categories or producing sounds that reflect developing L2 categories.

II. EXPERIMENT 1

In this experiment, the authors apply the metamodel approach to the measurement of Mandarin and English vowel similarity to make predictions regarding the L2 English vowel productions of Mandarin speakers.

A. Method

1. Target contrasts

The SLM argues that the perception of an L2 category depends on the specific phonetic context in which a token from that category is found. That is, the ability to perceive and produce an L2 sound in one context (e.g., a vowel fol-

lowing a stop) does not necessarily entail the ability to perceive and produce the same category in another context (e.g., the same vowel category following a fricative). Rather, L2 sounds in each context may be perceived differently in relation to a listener's L1 categories. This stems from the fact that acoustic characteristics of sound categories can differ quite radically, depending on the surrounding phonetic and prosodic environments (e.g., Strange, 2007).

To avoid dealing with large context effects, the authors limited phonetic and prosodic contexts across Mandarin and English. However, to test for a modest degree of generalization across contexts, the authors included two minimally different consonantal contexts. The authors selected vowel categories for comparison by identifying all Mandarin and Canadian English vowels found in post-bilabial stop contexts. Mandarin has a voiceless unaspirated and voiceless aspirated bilabial stop contrast, while English makes a contrast between voiced and voiceless aspirated bilabial stops. In keeping with conventions established by Romanized Mandarin orthography (Pinyin) and English orthography, the authors will broadly label the contrasts in both languages /b/ and /p/, despite some differences at the phonetic level. Through reference to several descriptive studies (Chen, 1976; Duanmu, 2003; Lee and Zee, 2003; Maddieson, 1984), the authors determined that five Mandarin vowels occur in /bV/ or /pV/ contexts: /i/, /e/, /a/, /uə/,[1] and /u/. In addition, the authors included Mandarin /o/ and /ɤ/ in their model. Although these vowels do not occur post-bilabially, they do occur in related post-alveolar and post-velar stop contexts. While Mandarin /y/ might be considered another potential candidate for comparison, the authors excluded it for two reasons. First, this vowel does not occur after oral stops in Mandarin. Second, earlier research (Thomson, 2005) found that it was not identified as phonetically similar to English vowels in /bV/ and /pV/ contexts. The English vowel categories chosen for comparison were /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɒ/, /ʌ/, /o/, /ʊ/, and /u/ the same ten Canadian English vowels investigated by Nearey and Assmann (1986).

Since the authors' ultimate objective was to make predictions regarding the L2 English vowel productions of Mandarin speakers, the authors felt that adhering to the Mandarin phonotactic preference for open syllables was more important than adhering to the English phonotactic constraint against the occurrence of lax vowels in isolated, open syllables. For Mandarin speakers, the presence of a coda consonant might introduce an additional confound. Thomson (2005) found evidence that Mandarin speakers' L2 productions of English vowels in closed syllables (i.e., [bVt] and [pVt]) resulted in errors that seemed to reflect the speakers' inability to produce some familiar vowels in closed syllables, rather than an inability to perceive or produce the vowels themselves. Consequently, the authors decided that vowels in Mandarin and English should all be produced in open, /pV/ and /bV/ syllables.

To control for prosodic contexts across languages, the authors compared Mandarin vowels produced with the fourth tone, a high falling tone contour, to English vowels spoken in phrase final position, which has falling intonation. This resulted in the best prosodic match across languages.

With the exception of [bo], [bɤ], [po], and [pɤ], the resulting Mandarin syllables were real Mandarin words. These four exceptions contain vowels that do not occur post-bilabially in Mandarin. Of the resulting English syllables, ten contain vowels that violate the English phonotactic constraint prohibiting lax vowels in isolated, open syllables. Prior experience at the University of Alberta showed that English speakers have little difficulty producing these vowels in open syllables. Thus, Nearey and Assmann (1986) and Andruski and Nearey (1992) both used isolated vowels. The latter study also found strong parallels between the acoustic patterns of lax vowels in /V/ and /bVb/ contexts.

2. Speakers

The Mandarin vowel production data were obtained from 20 native speakers (10 males, 10 females; ages 20–46, $M=28.2$) from Mainland China who reported speaking a standard variety of Mandarin. All were current or recent students at a Canadian university. Their length of residence in Canada ranged from 3 months to 6 years ($M=2.5$ years), and their age of arrival was between 18 and 44 years of age ($M=25.5$ years). Ten reported having had English as a second language (ESL) instruction in Canada (range 0–1 year, $M=6$ months). All reported normal hearing.

The English vowel production data were obtained from 20 native speaker undergraduate students at a Canadian university (10 males, 10 females; ages 18–50, $M=28.5$). All had resided in western Canada since childhood. Although several reported advanced knowledge of a second language, English was their primary language. All reported normal hearing.

3. Mandarin and English vowel production elicitation procedure

L1 vowel productions of the Mandarin and English speakers were recorded individually in a quiet room using a high quality Marantz digital recorder with a sampling rate of 44 100 Hz. Participants listened to and repeated a series of /bV/ and /pV/ stimuli containing the target vowels spoken by a female native speaker of their L1, Mandarin or English.

All Mandarin stimuli were presented in a Mandarin carrier phrase, "Xia yige zi shi _____" meaning "The next word is _____" and participants responded by repeating the word in a Mandarin carrier, "Xianzi wo shuo ____," meaning "Now I say _____." In addition to the pronunciation model provided by the auditory prompt, each Mandarin word was also shown in written form, both in Pinyin, and Chinese characters for real words.

To ensure that Mandarin participants understood that they were to put two of the Mandarin vowels in unfamiliar contexts, the authors first provided participants with real word prompts that rhymed with the target nonsense syllable and told them they wanted them to produce the same sound after bilabials. The real word rhyming prompts were gè, tè, kè, dòu, gòu, tòu, and kòu. The Mandarin speakers demonstrated no difficulty understanding or carrying out this task after hearing the auditory prompt. Two L1 Mandarin speak-

TABLE I. Classifications (%) by Mandarin model trained and tested on L1 Mandarin productions. Correct identifications are in bold.

Mandarin vowels repeated in response to auditory stimuli	Vowel identified by Mandarin pattern recognition model						
	/i/	/e/	/a/	/uə/	/o/	/ɤ/	/u/
/i/	**100**	…	…	…	…	…	…
/e/	2.5	**97.5**	…	…	…	…	…
/a/	…	…	**100**	…	…	…	…
/uə/	…	…	…	**87.5**	7.5	5	…
/o/	…	…	…	…	**90**	…	10
/ɤ/	…	…	…	20	…	**80**	…
/u/	…	…	…	…	…	…	**100**

ing colleagues verified that the productions of Mandarin /o/ and /ɤ/ in the target labial context mirrored productions in the rhyming Mandarin contexts.

All English stimuli were presented in the carrier phrase, "The next word is _____" and participants responded by repeating the word in the carrier, "Now I say _____." Since there is no straightforward way to represent English nonce words to participants who are unfamiliar with a phonetic alphabet, the authors did not provide them with a written form of the target syllables. Stimuli were evaluated by two native speakers to ensure the lax vowels were identifiable as the target vowels. Participants were instructed to pay particular attention to the vowel portion of each syllable. As expected, the vast majority of speakers had no difficulty producing lax vowels in open /bV/ and /pV/ syllables.

The entire recording procedure was repeated to obtain two recordings of each item. The resulting target syllables were extracted, down-sampled to 22 050 Hz, normalized across tokens to peak amplitude, and saved as separate sound files.

4. Analysis

Only one repetition of each CV syllable was analyzed. Prior to analysis, the first author screened the first repetition of each target syllable produced by each speaker to ensure that the vowel was a reasonable approximation of the intended target. One Mandarin /uə/ target from the first 280 productions (20 speakers × 14 target syllables) sounded more like Mandarin /ɤ/; it was replaced with the second production of that target syllable from the same speaker. 7 of the 400 first English targets (20 speakers ×20 target syllables), 4 /ɛ/ and 3 /ʌ/ productions, were replaced with second productions. In each instance, the first production was ambiguous, or perceived as an adjacent category. Some repetitions were only marginally better than the original. It was the authors' impression that a number of other English productions of /ɛ/ and /ʌ/ were also ambiguous with respect to adjacent categories. These were not replaced with second productions, however, because the second productions were equally ambiguous.

Using a suite of speech analysis tools[2] created with MATLAB®, vowel boundaries were marked for each sound file (400 for English and 280 for Mandarin). Next, a semiautomatic formant tracking procedure (Nearey *et al.*, 2002) was used to estimate the first three formant frequencies for each vowel.[3] F1, F2, and F3 values at 20% and 70% of each vowel's duration were then extracted. These points were chosen because they excluded formant transitions from preceding consonants as well as the edges of vowel tails. An autocorrelation pitch-tracking algorithm yielded mean F0 measures for each token. All formant frequency, pitch, and duration values were converted to a log scale to provide a more accurate reflection of the human auditory system (see Hillenbrand and Nearey, 1999).

Using these spectral measures, three pattern recognition models were trained and tested using discriminant function analysis:[4] (1) a Mandarin pattern recognition model (Mandarin model) comprising only Mandarin vowel categories, (2) an English pattern recognition model (English model) comprising only English vowel categories, and (3) a metamodel, treating Mandarin and English vowels as 17 separate categories within a single system. The Mandarin and English models were primarily used to verify that the L1 data were recognized as the intended categories, since accuracy of the underlying spectral measures must be assumed for the metamodel results to be meaningful. Because the pattern recognition models were trained and tested on the same production data, the authors used a round-robin cross-validation approach whereby each speaker to be tested was excluded from the training set against which his or her productions were measured.

B. Results

First, the accuracy of the L1 English and L1 Mandarin pattern recognition models was assessed to ensure productions were recognized as the intended categories. Results are shown in Tables I and II.

The Mandarin model (Table I) recognized 94% of the tokens as the intended vowel when vowel duration was excluded as a variable and 91% of tokens when duration was included. Some Mandarin categories were recognized less accurately than others. When examining APP scores by item, however, most misclassified tokens were recognized as somewhat similar to the intended Mandarin category. For example, of the eight misidentified Mandarin /ɤ/ tokens, seven had at least some probability of belonging to the intended category (mean APP=0.23; range 0.02–0.39), as did

TABLE II. Classifications (%) by English model trained and tested on L1 English productions. Correct identifications are in bold.

English vowels repeated in response to auditory stimuli	Vowel identified by English pattern recognition model									
	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɒ/	/ʌ/	/o/	/ʊ/	/u/
/i/	**95**	⋯	5	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/ɪ/	⋯	**92.5**	⋯	7.5	⋯	⋯	⋯	⋯	⋯	⋯
/e/	⋯	⋯	**100**	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/ɛ/	⋯	12.5	⋯	**77.5**	10	⋯	⋯	⋯	⋯	⋯
/æ/	⋯	⋯	⋯	7.5	**87.5**	2.5	2.5	⋯	⋯	⋯
/ɒ/	⋯	⋯	⋯	⋯	5	**90**	5	⋯	⋯	⋯
/ʌ/	⋯	⋯	⋯	⋯	5	2.5	**80**	⋯	12.5	⋯
/o/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**97.5**	⋯	2.5
/ʊ/	⋯	⋯	⋯	5	⋯	⋯	5	⋯	**90**	⋯
/u/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**100**

all five misclassified tokens of /uə/ (mean APP=0.21; range 0.01–0.42), and all four misclassified tokens of /o/ (mean APP=0.34; range 0.20–0.47).

The English model (Table II) recognized 91% of tokens as the intended category when vowel duration was included as a variable, and 86% of tokens when duration was excluded. Nearly all the misclassified English tokens were still identified by their APP scores as similar to the intended vowel. In particular, eight of nine misclassified tokens of English /ɛ/ had some probability of being the intended category (mean APP=0.16; range 0.01–0.41), as did all eight misclassified tokens of /ʌ/ (mean APP, 0.28; range 0.02–0.48), and all five misclassified tokens of /æ/ (mean APP =0.36; range 0.23–0.46).

As should be expected, the metamodel's recognition of Mandarin and English vowels, shown in Table III, is far less accurate (73%) than the monolingual models in recognizing the data as the intended categories. Accuracy rates were only marginally better (75.4%) when vowel duration was included. Since the authors' goal was to identify possible Mandarin-English interactions, and since the Mandarin model indicated that for Mandarin speakers, duration did not serve as a useful cue, the authors based their subsequent analysis on metamodel results that excluded vowel duration. Additionally, the difference in stimuli used to elicit data for each language may introduce an undesirable bias, where results reflect the relative duration of the elicitation stimuli rather than meaningful crosslinguistic differences in vowel duration. For example, the Mandarin and English production stimuli may have been produced with a slightly different speech rate.

From the metamodel results, the authors used the average degree of overlap between categories from each language as a quantitative measure of similarity. For example, Mandarin /i/ is classified as English /i/ 27.5% of the time, while English /i/ is classified as Mandarin /i/ 30% of the

TABLE III. Classifications (%) by metamodel trained and tested on L1 English and L1 Mandarin productions. Italicized numbers indicate misclassifications in opposing language. Subscripts e and m indicate English and Mandarin vowels, respectively.

	Intended vowels produced in English or Mandarin	Vowel recognized by the metamodel																
		English										Mandarin						
		/i/$_e$	/ɪ/$_e$	/e/$_e$	/ɛ/$_e$	/æ/$_e$	/ɒ/$_e$	/ʌ/$_e$	/o/$_e$	/ʊ/$_e$	/u/$_e$	/i/$_m$	/e/$_m$	/a/$_m$	/uə/$_m$	/o/$_m$	/ɤ/$_m$	/u/$_m$
English	/i/$_e$	65	⋯	5	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*30*	⋯	⋯	⋯	⋯	⋯	⋯
	/ɪ/$_e$	⋯	85	⋯	15	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
	/e/$_e$	⋯	2.5	85	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*12.5*	⋯	⋯	⋯	⋯	⋯
	/ɛ/$_e$	⋯	10	⋯	82.5	7.5	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
	/æ/$_e$	⋯	⋯	⋯	7.5	77.5	⋯	10	⋯	⋯	⋯	⋯	⋯	*5*	⋯	⋯	⋯	⋯
	/ɒ/$_e$	⋯	⋯	⋯	⋯	⋯	60	10	⋯	⋯	⋯	⋯	⋯	*30*	⋯	⋯	⋯	⋯
	/ʌ/$_e$	⋯	⋯	⋯	⋯	5.1	10.3	61.5	⋯	7.7	⋯	⋯	⋯	*15.4*	⋯	⋯	⋯	⋯
	/o/$_e$	⋯	⋯	⋯	⋯	⋯	⋯	⋯	60	⋯	2.5	⋯	⋯	⋯	⋯	*37.5*	⋯	⋯
	/ʊ/$_e$	⋯	⋯	⋯	2.5	2.5	⋯	2.5	⋯	67.5	⋯	⋯	⋯	⋯	⋯	⋯	*25*	⋯
	/u/$_e$	⋯	2.5	⋯	⋯	⋯	⋯	⋯	⋯	⋯	95	⋯	⋯	⋯	⋯	⋯	⋯	*2.5*
Mandarin	/i/$_m$	*27.5*	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	72.5	⋯	⋯	⋯	⋯	⋯	⋯
	/e/$_m$	⋯	⋯	*25*	⋯	⋯	⋯	⋯	⋯	⋯	⋯	2.5	72.5	⋯	⋯	⋯	⋯	⋯
	/a/$_m$	⋯	⋯	⋯	⋯	*2.5*	*30*	*5*	⋯	⋯	⋯	⋯	⋯	62.5	⋯	⋯	⋯	⋯
	/uə/$_m$	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	87.5	5	5	2.5
	/o/$_m$	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*30*	⋯	⋯	⋯	⋯	⋯	⋯	62.5	2.5	5
	/ɤ/$_m$	⋯	⋯	⋯	*5*	⋯	⋯	⋯	⋯	*27.5*	⋯	⋯	⋯	⋯	22.5	⋯	45	⋯
	/u/$_m$	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	100

TABLE IV. Degree of overlap between English and Mandarin vowel categories, ranked from highest to lowest.

English vowel	Closest Mandarin vowel	Degree of overlap between two categories (%)
/o/$_e$	/o/$_m$	33.75
/ɒ/$_e$	/a/$_m$	30.00
/i/$_e$	/i/$_m$	28.75
/ʊ/$_e$	/ɤ/$_m$	26.25
/e/$_e$	/e/$_m$	18.75
/ʌ/$_e$	/a/$_m$	10.20
/æ/$_e$	/a/$_m$	3.75
/ɛ/$_e$	/ɤ/$_m$	2.50
/u/$_e$	/u/$_m$	1.25
/ɪ/$_e$	n/a	0.00

time. This results in an average overlap of 28.75%. The degree of overlap for all English categories in terms of their closest Mandarin counterpart is provided in Table IV.

In Fig. 1, the authors provide a more traditional representation of crosslinguistic similarity using rough spectral properties of vowels, plotted in a two-dimensional space. To account for differences in vocal tract length, ratio measures (F1-F0 and F2-F1) were calculated using Bark transformed values (see Flege *et al.*, 1997). Ellipses represent a region containing 95% of the population of observations from a multivariate normal distribution, with the same mean and covariance parameters as those estimated from the authors' sample of native speaker productions of the vowel in question.

This approach yields different information than the metamodel analysis. First, each English and Mandarin category exhibits overlap with other categories within the same language. This suggests that most L1 categories are quite ambiguous, even for monolingual speakers of the language. The obvious implausibility of such a suggestion demonstrates that this method lacks the precision of the metamodel. Since the assessment of categories within a single language lacks precision, the authors can assume that using this approach to assess crosslinguistic similarity is equally imprecise. Comparing results from this traditional approach with those of the metamodel, the authors find both commonalities

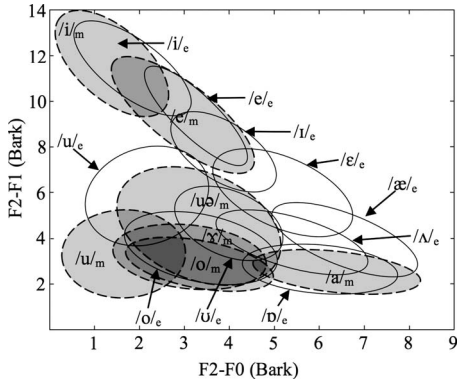


FIG. 1. Ellipses representing Bark transformed midpoint values of L1 Mandarin (enclosed in broken lines with subscript m) and L1 English (enclosed in solid lines with subscript e) vowels.

and differences. For example, this approach seems to indicate that English /i/, /e/, /ɒ/, and /o/ are very similar to their Mandarin counterparts, while English /æ/ and /ʌ/ are slightly less similar. English /ɛ/ is shown to be quite distinct from Mandarin categories. These findings, although difficult to quantify from a two-dimensional graphic representation, are consistent with the metamodel results. In contrast, this approach indicates that English /ʊ/ is less similar to Mandarin /ɤ/ than the metamodel suggests. English /ɪ/ and /u/ are shown to be somewhat similar to Mandarin /e/ and /uə/, respectively, which was also not shown by the metamodel.

C. Discussion

The Mandarin and English models' recognition of L1 vowels verifies that the spectral measures used were sufficiently accurate for extension to the metamodel, and that the authors' results are more precise than traditional two-dimensional approaches to assessing category similarity. The degree of overlap found in the Mandarin and English pattern recognition models was far less severe than that found using the traditional two-dimensional approach.

The authors recognize that their decision to use English lax vowels in isolated, open syllables may seem questionable to some researchers. They also acknowledge that using open syllables may impose a limitation on the generalizability of their results. However, they know of no empirical basis for believing that the use of open syllables was the cause of substantial artifactual inflation of pattern recognition errors. While it is true that the error rate of the L1 English model is somewhat higher than that found in methodologically incommensurate studies (e.g., Hillenbrand *et al.*, 1995), it is not out of keeping with results observed by Hillenbrand *et al.* (2001) who used similar measures and a similar round-robin cross-validation approach to examine English vowel productions in /CVC/ contexts. Furthermore, where the English pattern recognition model in the current study did find moderate confusion with adjacent categories, it was for English vowels that are known to sometimes be more difficult to perceive by human listeners, even in natural contexts. For example, Assmann and Katz (2000) reported that correct identification rates for the vowels /ɪ/ and /ɛ/ in /hVd/ contexts were only 65% and 85%, respectively. These relatively weak identification rates provide compelling evidence that even some natural English vowels (i.e., vowels in real words) can be ambiguous when listeners are required to process them phonetically, as was the case in Assmann and Katz's study (2000). Benkí (2003) suggested that in other studies, lexical context may facilitate recognition of some otherwise phonetically ambiguous vowel productions. In the authors' task, as Assmann and Katz's (2000), lexical context was unavailable, demanding greater reliance on phonetic processing.

A methodological difference likely explains the lower accuracy of the authors' English model compared to Nearey and Assmann's (1986) English model, which was based on productions of the same ten vowels, but in isolation (i.e., /V/). In that study, phonetically trained speakers were asked

to read phonetic transcriptions of the target vowels. Furthermore, the researchers discarded all tokens they deemed inaccurate.

Finally, a sociolinguistic phenomenon may also explain some of the error patterns found in the authors' results. Clarke *et al.* (1995) found evidence for what they describe as a Canadian vowel shift. Of greatest relevance, they reported that a number of speakers of a standard variety of Canadian English produced tokens of /ɪ/ and /ɛ/ that were impressionistically and acoustically identified as /ɛ/ and /æ/, respectively. Despite the fact that the vowels in the study of Clarke *et al.* (1995) were produced in real word contexts, the error patterns for the lax vowels are very similar to those observed in the authors' results. Interestingly, the speaker in their study who evidenced the most advanced degree of vowel shift was from the same Canadian city as the authors' speakers.

In spite of the minor limitations just outlined, the metamodel analysis appears to provide valuable information regarding the phonetic similarity of Mandarin and English vowels. For example, Mandarin /o/ and English /o/ categories are shown to be the most similar, although not identical. Other English vowels whose distributional characteristics are phonetically quite similar to Mandarin categories are /ɒ/, /i/, and /ʊ/ and to a slightly lesser extent, /e/. The SLM predicts that the learning of these more similar English vowels by Mandarin speakers will be the most difficult, since there is less evidence for learning new distributions. English /ʌ/ and /æ/ are somewhat less similar to Mandarin vowels, predicting they have a better chance of being learned than the more similar vowels. Learning English /æ/ may be easier than learning English /ʌ/, however, since the distribution of English /æ/ is less similar to the Mandarin /a/ category than is the distribution of English /ʌ/. The metamodel never misclassified English /ɪ/ and /ɛ/ as Mandarin vowels, and rarely misclassified English /u/ as a Mandarin vowel. Consequently, for Mandarin L1 speakers, English /ɪ/, /ɛ/, and /u/ are least likely to be confused with any Mandarin categories.

III. EXPERIMENT 2

In this experiment, the authors test predictions stemming from the metamodel's assessment of Mandarin and English vowel similarity against L2 English vowels produced by Mandarin L1 speakers. The metamodel comparison of Mandarin and English vowels in experiment 1 predicts a strong relationship between crosslinguistic similarity and the degree to which these speakers will assimilate English vowels to their L1 categories. Specifically, English vowel categories that are statistically more similar to Mandarin vowel categories will be more likely to assimilate to Mandarin categories than those that are statistically less similar. In addition, L2 English speakers will be more likely to evidence learning for the statistically less similar English vowels. This will be reflected by emerging, English-like distributional properties for these vowels. The authors used low proficiency English speakers because it was felt that they were more likely to produce L2 accented speech that would exhibit differing degrees of English category formation relative to each English category's degree of similarity to Mandarin.

To test these predictions, the metamodel will be used to evaluate the statistical similarity of Mandarin speakers' L2 English productions relative to the metamodel's recognition of L1 English and L1 Mandarin categories in experiment 1 (see Table III). If the L2 English speakers substitute a Mandarin L1 category for a similar English category, the metamodel's recognition of those L2 productions should resemble its recognition of a Mandarin L1 category. For example, if the learners substitute Mandarin /i/ for English /i/, English L2 productions should resemble the productions of the L1 Mandarin speakers and, hence, on the basis of row 11 in Table III, approximately 70% of the L2 English productions should be recognized as Mandarin /i/, while approximately 30% should be recognized as English /i/. For statistically dissimilar English categories, L2 English productions should not mirror any L1 category, but rather, evidence English category learning.

A. Method

1. L2 speakers

The L2 English vowel productions were obtained from 22 standard Mandarin speakers (14 women, 8 men; ages 27–50, $M=36.4$) from Mainland China who were recruited from a local ESL program. Most had arrived within the previous year (M LOR=10.7 months, range=4–48 months). Four who arrived earlier reported having had little interaction with Canadians, and no exposure to English on a daily basis prior to enrolling in the ESL program. Participants had been studying beginner level ESL for an average of 4 months (range =1–13 months). They had received little or no explicit pronunciation instruction. All participants reported normal hearing.

2. English L2 vowel elicitation

Elicitation and vowel editing procedures were identical to those used with the L1 English speakers in experiment 1. Of crucial importance, the authors used the same elicitation stimuli.

3. Analysis

The suite of MATLAB speech tools used to obtain spectral measures of the production data in experiment 1 was again used. The spectral measures from the native speaker English stimuli were tested against the English model to verify that they clearly reflected the intended categories. The English stimuli were also tested against the metamodel to determine their relative similarity to Mandarin categories. Finally, the L2 production data were tested against the metamodel to compare their distributional properties with those of Mandarin and English categories.

B. Results

Table V provides the English model APP scores for individual speech tokens produced by the female English na-

TABLE V. Probabilities (APPs) of native speaker English elicitation stimuli being members of specific English model categories. Largest APPs are in bold.

Intended English stimuli vowels produced by female native speaker	APP scores of belonging to each English category									
	/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɒ/	/ʌ/	/o/	/ʊ/	/u/
/bi/	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/bɪ/	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/be/	⋯	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/bɛ/	⋯	0.14	⋯	**0.86**	⋯	⋯	⋯	⋯	⋯	⋯
/bæ/	⋯	⋯	⋯	⋯	**0.99**	⋯	⋯	⋯	⋯	⋯
/bɒ/	⋯	⋯	⋯	⋯	⋯	**0.99**	⋯	⋯	⋯	⋯
/bʌ/	⋯	⋯	⋯	⋯	⋯	0.01	**0.74**	⋯	0.25	⋯
/bo/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯	⋯
/bʊ/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯
/bu/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**
/pi/	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/pɪ/	⋯	**0.99**	⋯	0.01	⋯	⋯	⋯	⋯	⋯	⋯
/pe/	⋯	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/pɛ/	⋯	⋯	⋯	**0.89**	0.01	⋯	0.10	⋯	⋯	⋯
/pæ/	⋯	⋯	⋯	⋯	**0.98**	0.01	0.01	⋯	⋯	⋯
/pɒ/	⋯	⋯	⋯	⋯	⋯	**0.96**	0.04	⋯	⋯	⋯
/pʌ/	⋯	⋯	⋯	⋯	⋯	0.01	**0.99**	⋯	⋯	⋯
/po/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯	⋯
/pʊ/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯
/pu/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**

tive speaker. Although APP scores indicate that some tokens are slightly similar to competing English categories, each of these novel tokens (not used in training the English L1 or metamodels) is clearly recognized as most like the intended category. Table VI provides metamodel APP scores for the same speech tokens. These results reveal that three of the clearly identifiable English vowels, [i], [ɒ], and [ʊ], in both /bV/ and /pV/ contexts, are spectrally closer to prototypical

TABLE VI. Probabilities (APPs) of native speaker English elicitation stimuli being members of specific metamodel categories. Largest English APPs are in bold. Largest Mandarin competitor APPs are in italics.

Intended English stimuli vowels produced by female native speaker	Vowel recognized by the metamodel																
	English										Mandarin						
	/i/$_e$	/ɪ/$_e$	/e/$_e$	/ɛ/$_e$	/æ/$_e$	/ɒ/$_e$	/ʌ/$_e$	/o/$_e$	/ʊ/$_e$	/u/$_e$	/i/$_m$	/e/$_m$	/a/$_m$	/uə/$_m$	/o/$_m$	/ɤ/$_m$	/u/$_m$
/bi/	**0.36**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*0.64*	⋯	⋯	⋯	⋯	⋯	⋯
/bɪ/	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/be/	⋯	⋯	**0.68**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*0.32*	⋯	⋯	⋯	⋯	⋯
/bɛ/	⋯	0.08	⋯	**0.91**	⋯	⋯	⋯	⋯	0.01	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/bæ/	⋯	⋯	⋯	0.02	**0.79**	⋯	0.09	⋯	⋯	⋯	⋯	⋯	*0.08*	⋯	⋯	⋯	⋯
/bɒ/	⋯	⋯	⋯	⋯	⋯	**0.17**	0.06	⋯	⋯	⋯	⋯	⋯	*0.77*	⋯	⋯	⋯	⋯
/bʌ/	⋯	⋯	⋯	⋯	0.01	0.03	**0.60**	⋯	0.23	⋯	⋯	⋯	*0.09*	⋯	⋯	0.03	⋯
/bo/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**0.59**	⋯	⋯	⋯	⋯	⋯	⋯	*0.41*	⋯	⋯
/bʊ/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**0.03**	⋯	⋯	⋯	⋯	0.04	⋯	*0.92*	⋯
/bu/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/pi/	**0.33**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*0.67*	⋯	⋯	⋯	⋯	⋯	⋯
/pɪ/	⋯	**0.96**	⋯	0.04	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/pe/	⋯	⋯	**0.65**	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*0.35*	⋯	⋯	⋯	⋯	⋯
/pɛ/	⋯	⋯	⋯	**0.46**	0.17	⋯	0.35	⋯	0.01	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/pæ/	⋯	⋯	⋯	⋯	**0.47**	0.04	0.12	⋯	⋯	⋯	⋯	⋯	*0.37*	⋯	⋯	⋯	⋯
/pɒ/	⋯	⋯	⋯	⋯	⋯	**0.11**	0.05	⋯	⋯	⋯	⋯	⋯	*0.84*	⋯	⋯	⋯	⋯
/pʌ/	⋯	⋯	⋯	⋯	0.01	0.13	**0.31**	⋯	⋯	⋯	⋯	⋯	*0.55*	⋯	⋯	⋯	⋯
/po/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**0.60**	⋯	⋯	⋯	⋯	⋯	⋯	*0.40*	⋯	⋯
/pʊ/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**0.22**	⋯	⋯	⋯	⋯	⋯	⋯	*0.78*	⋯
/pu/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**1.00**	⋯	⋯	⋯	⋯	⋯	⋯	⋯

TABLE VII. Classifications (%) of L2 English productions by metamodel trained on L1 Mandarin and L1 English categories. Winning categories are in bold. Largest competitor categories in the opposing language are in italics.

Intended L2 English vowels	Vowel recognized by the metamodel: English /i/$_e$	/ɪ/$_e$	/e/$_e$	/ɛ/$_e$	/æ/$_e$	/ɒ/$_e$	/ʌ/$_e$	/o/$_e$	/ʊ/$_e$	/u/$_e$	Mandarin /i/$_m$	/e/$_m$	/a/$_m$	/uə/$_m$	/o/$_m$	/ɤ/$_m$	/u/$_m$
(a) Responses to /bV/ stimuli																	
/bi/	*31.8*	⋯	2.3	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**65.9**	⋯	⋯	⋯	⋯	⋯	⋯
/bɪ/	2.3	**54.5**	4.5	34.1	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*4.5*	⋯	⋯	⋯	⋯	⋯
/be	4.5	⋯	*25*	2.3	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**68.2**	⋯	⋯	⋯	⋯	⋯
/bɛ/	⋯	13.6	⋯	**75**	4.5	2.3	4.5	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
/bæ/	⋯	2.3	⋯	**43.2**	34.1	2.3	11.4	⋯	⋯	⋯	⋯	⋯	6.8	⋯	⋯	⋯	⋯
/bɒ/	⋯	⋯	⋯	⋯	⋯	*22.7*	6.8	⋯	⋯	⋯	⋯	⋯	**65.9**	2.3	⋯	2.3	⋯
/bʌ/	⋯	⋯	⋯	2.3	⋯	13.6	*15.9*	2.3	2.3	⋯	⋯	⋯	**63.6**	⋯	⋯	⋯	⋯
/bo/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*18.2*	⋯	⋯	⋯	⋯	⋯	2.3	**72.7**	4.5	2.3
/bʊ/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	2.3	*9.1*	⋯	⋯	⋯	⋯	25	13.6	**47.7**	2.3
/bu/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	2.3	*20.5*	⋯	⋯	⋯	9.1	⋯	20.5	**47.7**
(b) Responses to /pV/ stimuli																	
/pi/	*34.1*	⋯	2.3	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**63.6**	⋯	⋯	⋯	⋯	⋯	⋯
/pɪ/	⋯	**54.5**	13.6	27.3	⋯	⋯	2.3	⋯	⋯	⋯	⋯	*2.3*	⋯	⋯	⋯	⋯	⋯
/pe/	2.3	2.3	*18.2*	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**77.3**	⋯	⋯	⋯	⋯	⋯
/pɛ/	⋯	⋯	⋯	**79.5**	15.9	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*4.5*	⋯	⋯	⋯	⋯
/pæ/	⋯	⋯	⋯	18.2	**29.5**	6.8	18.2	⋯	⋯	⋯	⋯	⋯	*27.3*	⋯	⋯	⋯	⋯
/pɒ/	⋯	⋯	⋯	⋯	4.5	*15.9*	6.8	⋯	⋯	⋯	⋯	⋯	**70.5**	⋯	2.3	⋯	⋯
/pʌ/	⋯	⋯	⋯	⋯	2.3	4.5	*9.1*	⋯	⋯	⋯	⋯	⋯	**84.1**	⋯	⋯	⋯	⋯
/po/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*20.5*	⋯	⋯	⋯	⋯	⋯	6.8	**68.2**	⋯	4.5
/pʊ/	⋯	⋯	⋯	⋯	⋯	4.5	4.5	⋯	*22.7*	⋯	⋯	⋯	2.3	20.5	4.5	**38.6**	2.3
/pu/	⋯	⋯	2.3	⋯	⋯	⋯	⋯	⋯	4.5	*15.9*	⋯	⋯	⋯	13.6	11.4	18.2	**34.1**
(c) Responses pooled across /bV/ and /pV/ contexts																	
/i/	*33*	⋯	2.3	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**64.8**	⋯	⋯	⋯	⋯	⋯	⋯
/ɪ/	12	**54.5**	9.1	30.7	⋯	⋯	1.2	⋯	⋯	⋯	⋯	*3.4*	⋯	⋯	⋯	⋯	⋯
/e/	3.4	1.2	*21.6*	1.2	⋯	⋯	⋯	⋯	⋯	⋯	⋯	**72.8**	⋯	⋯	⋯	⋯	⋯
/ɛ/	⋯	6.8	⋯	**77.3**	10.2	1.2	2.3	⋯	⋯	⋯	⋯	⋯	*2.3*	⋯	⋯	⋯	⋯
/æ/	⋯	1.2	⋯	30.7	**31.8**	4.6	14.8	⋯	⋯	⋯	⋯	⋯	*17.1*	⋯	⋯	⋯	⋯
/ɒ/	⋯	⋯	⋯	⋯	2.3	*19.3*	6.8	⋯	⋯	⋯	⋯	⋯	**68.2**	1.2	1.2	1.2	⋯
/ʌ/	⋯	⋯	⋯	1.2	1.2	9.1	*12.5*	1.2	1.2	⋯	⋯	⋯	**73.9**	⋯	⋯	⋯	⋯
/o/	⋯	⋯	⋯	⋯	⋯	⋯	⋯	*19.4*	⋯	⋯	⋯	⋯	⋯	4.6	**70.5**	2.3	3.4
/ʊ/	⋯	⋯	⋯	⋯	⋯	2.3	2.3	1.2	*15.9*	⋯	⋯	⋯	1.2	22.8	9.1	**43.2**	2.3
/u/	⋯	⋯	1.15	⋯	⋯	⋯	⋯	⋯	3.4	*18.2*	⋯	⋯	⋯	11.4	5.7	19.4	**40.9**

Mandarin /i/, /a/, and /ɤ/ categories, respectively, than they are to the prototypical English categories. One instance of English /ʌ/, in the [pʌ] syllable, is also closer to Mandarin /a/. The remaining tokens are all recognized as more English-like, but many others have some probability of being a member of a Mandarin category. These results should not be taken to mean that some of the English stimuli are not perfectly good English tokens. Rather, the results indicate that despite being acceptable English productions, falling well within the boundaries of the English categories, they are simultaneously closer to the center of statistically similar Mandarin categories.

The rows of Table VII show what the authors will henceforth call metamodel category profiles for L2 English productions. Numbers in each row indicate the percentage of L2 productions of each English vowel that were recognized as being members of each metamodel category. Results for each L2 English vowel are shown for /bV/ and /pV/ contexts separately in Table VII, as well as for the average over both contexts.

A study of corresponding rows of Tables III and VII suggests that the metamodel evaluates productions of some L2 vowels very similarly to certain L1 Mandarin vowels. For example, consider again the metamodel profile for L2 English /i/ (row 1 of Table VII(c)), where roughly one-third of productions were recognized as English /i/, two-thirds as Mandarin /i/, and recognition as other categories at or near zero. Contrasting this L2 English /i/ category profile with the profiles for L1 English /i/ (row 1 of Table III) and L1 Mandarin /i/ (row 11 of Table III), it is immediately apparent that the metamodel profile of L2 English /i/ is extremely similar

TABLE VIII. Correlations between L2 category distributions and the two most similar categories in L1 English and/or L1 Mandarin.

English vowels repeated in response to auditory stimuli	Most similar category	Correlation	Second most similar category	Correlation
/i/	Mandarin /i/	0.993	English /i/	0.758
/ɪ/	English /ɪ/	0.930	English /ɛ/	0.533
/e/	Mandarin /e/	0.997	English /e/	0.364
/ɛ/	English /ɛ/	0.998	English /æ/	0.146
/æ/	English /æ/	0.718	English /ɛ/	0.631
/ɒ/	Mandarin /a/	0.983	English /ɒ/	0.632
/ʌ/	Mandarin /a/	0.937	English /ɒ/	0.507
/o/	Mandarin /o/	0.980	English /o/	0.699
/ʊ/	Mandarin /ɤ/	0.953	English /ʊ/	0.499
/u/	Mandarin /u/	0.814	English /u/	0.305

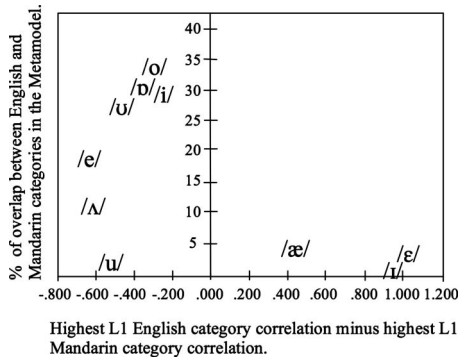


FIG. 2. Correlations between the metamodel category profile for each L2 English category and the closest L1 English category, minus the correlation between the metamodel category profile for each L2 English category and the closest L1 Mandarin category, plotted against degree of crosslinguistic similarity.

to the profile of L1 Mandarin /i/, but is less similar to the profile of L1 English /i/.

The (Pearson) correlation coefficient between any two rows of numbers serves as an efficient (purely descriptive) quantitative summary for the degree of similarity between the category profiles.[5] Thus for example, the correlation between the L2 English /i/ and L1 Mandarin /i/ profiles is 0.993, while the correlation between L2 English /i/ and L1 English /i/ profiles is 0.758. We calculated correlations of every row of Table VII(c) with every row of Table III and use the rankings to characterize the relative similarity of the metamodel category profiles of L2 productions to those of L1 categories. The ordering of the correlation summary measure corresponds well with the impressionistic assessment of the similarity of pairs of rows.

The authors summarize key aspects of these comparisons in Table VIII, which reports the two highest correlations between each of the L2 English metamodel category profiles of Table VII and the L1 metamodel category profiles of Table III. In some cases, L2 English vowel production profiles were closest to two L1 English vowel profiles, and further from the profile of any L1 Mandarin vowel.

The results in Table VIII indicate that L2 English /i/, /e/, /ɒ/, /o/, and /ʊ/ productions are most highly correlated with the similar Mandarin vowels /i/, /e/, /a/, /o/, and /ɤ/, respectively. L2 English /ʌ/ is also most highly correlated with the Mandarin category /a/. In contrast, L2 English /æ/ is most highly correlated with English /æ/ and not highly correlated with any Mandarin category. Examining the raw recognition scores in Table VII, however, there is evidence that L2 productions of /æ/ were sometimes recognized as Mandarin /a/. These results reflect the experiment 1 finding that English /ʌ/ is statistically more similar to Mandarin /a/ than is English /æ/. Hence, while some confusion of English /æ/ with Mandarin /a/ is evident, the smaller degree of confusion implies that English /æ/ should be more easily learned than English /ʌ/.

The L2 productions of the statistically less similar English vowels /ɪ/ and /ɛ/ were not recognized by the metamodel as having distributional properties that were strongly correlated with any Mandarin vowel. Rather, the L2 English /ɪ/ and /ɛ/ productions were most highly correlated with the intended English categories.

In contradiction of the authors' predictions, L2 productions of the statistically less similar English vowel /u/ were found to be quite highly correlated with L1 Mandarin /u/. However, the strength of the correlation was not nearly as high as for the statistically more similar vowels. Furthermore, looking at the raw metamodel category profiles for L2 English /bu/ and /pu/ in Table VII, it is evident that the L2 English /u/ productions are not clearly recognized as belonging to any one category, English or Mandarin. Some L2 productions were recognized as English /u/, but many were recognized as Mandarin /uə/, /o/, /ɤ/, and /u/.

The metamodel assessment of the L2 productions also shows interesting patterns with regard to the speakers' responses to individual items produced by the English native speaker. In the case of four of the five most statistically similar vowels, /e/, /ɒ/, /o/, and /ʊ/, the stimulus item in each /bV/-/pV/ pair that contained the vowel with the highest APP of belonging to a Mandarin category was also the item that resulted in L2 productions with a more Mandarin-like profile. For example, the English speaker's [bʊ] had a higher APP of being a member of the Mandarin category than the [pʊ]. In the same way, the profile of L2 /ʊ/ productions in response to the [bʊ] stimulus was more Mandarin-like than the profile of L2 /ʊ/ productions in response to [pʊ]. Among the most statistically similar vowels, only L2 productions in response to the [bi]-[pi] stimulus pair did not follow this pattern. However, these stimuli were quite equally Mandarin-like, with APP scores of 0.64 and 0.67, respectively. The resulting L2 profiles were also equally Mandarin-like, 65.9% and 63.6%, respectively.

For the two somewhat similar English vowels, /æ/ and /ʌ/, the degree of similarity between the individual stimulus items and the similar Mandarin category is also reflected in the L2 productions. The sounds in each /bV/-/pV/ pair with the higher APP of being a Mandarin category elicited the more Mandarin-like metamodel category profile.

Finally, Fig. 2 (suggested by an anonymous reviewer) provides a rough quantitative assessment of the relationship between crosslinguistic similarity as operationalized in ex-

periment 1, and the extent to which English vowels assimilate to L1 Mandarin categories in experiment 2. In this figure, the correlation between each L2 category and the closest L1 English category minus the correlation between each L2 category and the closest L1 Mandarin category is plotted by the degree of crosslinguistic similarity established in experiment 1 (see Table IV). L2 English categories falling to the left of the vertical line are more Mandarin-like, while categories to the right of the line are more English-like. It is clear that if the authors exclude L2 English /u/ as an outlier, all L1 English vowels that are above a certain similarity threshold (in this case, approximately 10% overlap) are strongly assimilated to their Mandarin L1 counterparts. In contrast, those with more negligible degrees of overlap (i.e., less than 5%) do not assimilate to L1 Mandarin categories.

C. Discussion

The general confusion patterns resulting from application of the metamodel to the L2 production data largely support the authors' predictions. According to this analysis, when English vowel categories are statistically very similar to L1 Mandarin categories, learners tend to substitute the relevant L1 category for the L2 category. When English vowel categories are statistically less similar to L1 Mandarin categories, learning may occur, as was evidently the case with English /æ/. The fact that the L2 English learners substituted Mandarin /a/ for English /ʌ/ may be because the distribution of English /ʌ/ is either too similar to Mandarin /a/ for the learners to have detected a difference, or because the learners had not had sufficient experience with English /ʌ/ to begin acquiring this category. The metamodel also correctly predicted L2 responses to different stimuli containing the same English vowel. In six of seven pairs of very similar or somewhat similar vowels, the stimulus item with the more Mandarin-like APP score elicited the more Mandarin-like distribution of L2 productions.

When English vowel categories are statistically new relative to L1 Mandarin categories, L2 production distributions do not reflect the distribution of any Mandarin category, although they may also not yet be English-like. When errors in production occur, they are always in the direction of phonetically adjacent categories. Hence, L2 production errors for English /ɪ/ were most often in the direction of English /ɛ/, while L2 production errors for English /ɛ/ were in the direction of either /ɪ/ or /æ/. Although it may appear that the metamodel's recognition of many L2 productions of English /u/ as more like Mandarin L1 categories contradicts the prediction that this vowel would not be confused with a Mandarin category, an alternate account can be offered. If, in fact, English /u/ is confused with a single Mandarin category, the metamodel category profile of L2 English /u/ productions should resemble the category profile of an L1 Mandarin category as represented by the metamodel in experiment 1 (see Table III). This is clearly not the case. A more plausible explanation for the error patterns observed in the L2 English /u/ data is that as with the other statistically new English vowels, /ɪ/ and /ɛ/, developmental errors are in the direction of adjacent categories. In the case of /ɪ/ and /ɛ/, adjacent categories are English categories. In the case of English /u/, adjacent categories happen to be L1 Mandarin categories. Therefore, any undershoot or overshoot in the productions of L2 English /u/ are recognized as those categories.

IV. GENERAL CONCLUSION

At the outset, the authors argued that more research is needed to determine how best to quantify crosslinguistic similarity. The metamodel approach demonstrated in experiments 1 and 2 provides some useful insights. It offers a strong complement to the use of perceptual mapping experiments and seems to offer greater precision than other methods that rely on spectral measures to define phonetic similarity.

We do not intend to suggest that the metamodel is representative of real-world perceivers as is an L1 pattern recognition model (e.g., Nearey and Assmann, 1986). Rather, the metamodel views the process of L2 category learning from the perspective of a putative end state, representing what an idealized bilingual human listener would perceive in a perfect interlanguage, where small phonetic differences between similar L1 and L2 categories are actually discernable. As such, the metamodel is able to characterize how categories in two languages compete for territory in the crosslinguistic space. To the extent that low or intermediate proficiency L2 learners have managed to begin developing some tentative approximation of new L2 categories, the metamodel stands a better chance of providing relevant information than do pure L1 discriminative models.

Because the metamodel currently lacks any proven ontological status with respect to an interlanguage, its purpose is currently limited to (1) comparing the distributional properties of sounds in two languages and (2) assessing the distributional properties of L2 productions. These capabilities are important to research within the PAM and SLM paradigms. While some may debate whether production ability reflects underlying perceptual categories, the authors take Flege's (1995) position that the two skills are closely related. While it might be the case that production ability in this study lags perceptual ability that would only mean that the authors' results might reflect perceptual ability at a slightly earlier point in time. In fact, Thomson (2008) found that the production and identification of English vowels by these learners were not significantly different. Furthermore, a training study (Thomson, 2007) found that improvements in perception resulted in immediate improvement in production by these learners, providing clear evidence that for them, the two skills are closely related.

For PAM, the extent of overlap between distributions of opposing language categories in the metamodel predicts the likelihood that naive listeners will identify sounds in an unfamiliar language as native language categories. When the distribution of a category in the unfamiliar language does not overlap with any native language category, sounds from that category will be perceived as uncategorizable speech sounds. On the basis of APP scores, the metamodel also seems capable of determining the extent of similarity between individual instantiations of sounds in one language and similar

categories in an opposing language. For the most statistically similar speech sounds, however (e.g., English /o/ and Mandarin /o/), relating relative phonetic similarity to learner behavior may lack meaning. The pre-existing L1 perceptual space may render small phonetic differences undetectable. For example, Kuhl and Iverson's (1995) "perceptual magnet effect" is said to warp speech perception such that phonetic differences near category centers are impossible to perceive. Although the exact nature of this effect is disputed (e.g., Lotto *et al.*, 1998), it is agreed that listeners are better able to discriminate differences between tokens of a given category when those tokens are distant from categorical centers. Consequently, the authors should not necessarily expect Mandarin or English speakers to be able to perceive a difference between Mandarin and English /o/, even if it is acoustically measurable, because sounds from one language are still relatively close to the center of the distribution of the opposing language category.

For the SLM, the extent of overlap between distributions of L1 and L2 sound categories in the metamodel predicts the likelihood that L2 learners will acquire L2 sound category distributions rather than simply substitute L1 categories for L2 ones. The learners in this study were relative beginners, which may explain why the English distribution of many categories had not yet been well learned. The metamodel approach can also be used to test SLM predictions with more advanced learners. For example, L1 Mandarin speakers with advanced knowledge of English might be expected to produce less similar English vowels more accurately than the beginner L2 speakers in this study, because they would have had more opportunity to learn their distributions.

The metamodel approach can also test Flege's (1995) claims regarding the process of dissimilation. He argued that L2 learners may be unable to perfectly develop some new L2 categories when those categories are particularly close to others in shared L1/L2 perceptual space. Instead, such categories develop in such a way as to remain maximally distinct from pre-existing L1 categories. In the case of Mandarin learners of English, this seems most likely to occur with English /u/, which must develop within what appears to be a dense phonetic space. In future studies, it would be interesting to trace the development of this vowel in more advanced L2 learners to see if it settles in a region of the space more remote from Mandarin categories.

Another SLM prediction that can be directly assessed against the metamodel is the phenomenon of category merger. Flege (1995) argued that in some cases, where L1 and L2 categories are very similar, maintaining separation of the two categories will be difficult. An example from this study is the Mandarin /o/ and English /o/ categories. According to the SLM, these two categories are likely to merge, resulting in Mandarin and English distributions that are identical. The metamodel can confirm this prediction by assessing the L2 productions of speakers who have differing degrees of L2 experience. The probability that L2 productions will be recognized as a member of one language over the other should approach 0.5 as the learners' experience with the L2 increases.

The results of this study also have implications for L2 speech learning in general. They suggest that learning that a particular L2 category is different from any L1 category may depend on the number of tokens in the L2 category that are noticeably different from L1 categories. In the case of statistically similar categories, even if a learner perceives some evidence for establishing a new category center, strengthening it will be difficult, because most input will be assimilated to the similar L1 category. For categories whose distributions are only somewhat similar, learning may be possible, but still difficult. In such cases, some tokens may provide learners with evidence for L2 category formation, but many tokens provide counterevidence because they are perceived to be like L1 categories. For categories whose distributions are statistically new, a fair number of novel L2 tokens should be correctly identified, thereby strengthening the emerging category. From this perspective then, the metamodel serves to implicitly specify areas of phonetic space where new category centers might receive the reinforcement needed to persist.

When L1 categories are so similar that they can be substituted for an L2 category without a loss of intelligibility for native listeners (e.g., Mandarin /i/ will be perceived as English /i/), learning is unnecessary and category merger is likely. In cases where somewhat similar L1 categories will not suffice (e.g., Mandarin /a/ will not be perceived as English /æ/ and /ʌ/), learning is more important for intelligibility. For such categories, it may be better to initially train learners using computer-mediated approaches that present only those tokens that are most likely to provide positive evidence. Once these L2 categories begin to emerge, more variability in training stimuli can be incorporated, including training tokens that are somewhat more Mandarin-like. For the statistically new categories, most of the English tokens that Mandarin speakers hear provide positive evidence, and, therefore, high variability training may be of benefit from the start.

In conclusion, the metamodel approach demonstrated in this paper contributes additional insights to L2 speech perception and production research. It not only confirms general PAM and SLM predictions, but it also identifies interactions between L2 sounds and L1 categories at the level of individual stimuli. While the metamodel by no means provides an end point in the operationalization of crosslinguistic similarity, the authors hope that information gleaned from its use here will motivate further investigations into how best to define this construct.

ACKNOWLEDGMENTS

The authors acknowledge Murray Munro for his very insightful comments on early drafts of this paper, and Geoff Morrison for his helpful feedback. They also greatly appreciate the extensive feedback of three anonymous reviewers. The authors are grateful for financial support provided by an Izaak Walton Killam Memorial Scholarship, a University of Alberta Doctoral Fellowship, a Social Sciences and Humanities Research Council of Canada (SSHRC) Doctoral Fellowship, and a Brock University, Humanities Research Institute

grant awarded to R.I.T. This research was also supported by SSHRC grants awarded to T.M.N. and T.M.D.

[1]Mandarin /uə/ listed by Lee and Zee (2003) as a diphthong is not the same as the Mandarin monophthong /u/, which they also listed and which was also selected for analysis.

[2]These tools were designed by G. Morrison and T. Nearey.

[3]Raw formant tracks were derived from a set of candidate analyses using 100 ms $\cos^4$ windows (Talkin, 1987) with a 2 ms frame advance rate. Each candidate analysis was calculated with a different high-frequency cutoff, computed via selective Linear Predictive Coding (LPC) (Markel and Gray, 1976). There were eight cutoff frequencies, ranging from 3000 to 4500 Hz in equal log steps. The "best" cutoff is tentatively determined automatically, by evaluating the eight candidate track sets and ranking alternate solutions using a complex heuristic procedure sketched in Nearey *et al.*, 2002. The candidate analyses were then screened manually and verified by resynthesizing vowels and comparing the results auditorily. In many cases, the best automatic analysis was deemed acceptable. In other cases, one of the alternative analyses was selected. Finally, in some cases (particularly where breathiness or laryngealization occurred near the ends of vowels), a graphic hand editing procedure was used to correct the automatic tracks.

[4]A single covariance Gaussian model with no dimensionality reduction was used (see Nearey and Assmann, 1986).

[5]An ordinary (Pearson) correlation coefficient provides a plausible descriptive index of the similarity of the predicted response profiles for L2E productions in Table VII and those of the native speaker's productions in Table V. While other measures of association might be considered, there seems to be no compelling reason to abandon this simple measure for descriptive purposes.

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

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

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

Benkí, J. R. (**2003**). "Analysis of English nonsense syllable recognition in noise," Phonetica **60**, 129–157.

Best, C. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Philadelphia, PA), pp. 13–34.

Bohn, O. S., and Flege, J. E. (**1992**). "The production of new and similar vowels by adult German learners of English," Stud. Second Lang. Acquis. **14**, 131–158.

Chen, M. Y. (**1976**). "From middle Chinese to modern Peking," J. Chin. Linguist. **4**, 113–277.

Clarke, S., Elms, F., and Youssef, A. (**1995**). "The third dialect of English: Some Canadian evidence," Language Variation and Change **7**, 209–228.

Duanmu, S. (**2003**). *The Phonology of Standard Chinese* (Oxford University Press, Oxford).

Flege, J. E. (**1995**). "Second-language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 229–273.

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "Effects of experience on non-native speakers' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., Munro, M. J., and Fox, R. (**1994**). "Auditory and categorical effects on cross-language vowel perception," J. Acoust. Soc. Am. **95**, 3623–3641.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (**2000**). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," J. Acoust. Soc. Am. **107**, 2711–2724.

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

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

Kuhl, P. K., and Iverson, P. (**1995**). "Linguistic experience and the "perceptual magnet effect"," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 121–154.

Lee, W. S., and Zee, E. (**2003**). "Standard Chinese (Beijing)," J. Int. Phonetic Assoc. **33**, 109–112.

Lotto, A. J., Kluender, K. R., and Holt, L. L. (**1998**). "Depolarizing the perceptual magnet effect," J. Acoust. Soc. Am. **103**, 3648–3655.

Levi, S. V., Winters, S. J., and Pisoni, D. B. (**2007**). "Speaker-independent factors affecting the perception of foreign accent in a second language," J. Acoust. Soc. Am. **121**, 2327–2338.

Maddieson, I. (**1984**). *Patterns of Sounds* (Cambridge University Press, Cambridge).

Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, Berlin).

Morrison, G. S. (**2006**). "L1 and L2 production and perception of English and Spanish vowels: A statistical modeling approach," Ph.D. thesis, University of Alberta, Edmonton, Alberta, Canada.

Munro, M. J., and Derwing, T. M. (**2008**). "A developmental study of Mandarin and Russian speakers' English vowels," Lang. Learn. **58**, 479–502.

Munro, M. J., Derwing, T. M., and Thomson, R. I. (**2003**). "A longitudinal examination of English vowel learning by Mandarin speakers," Can. Acoust. **31**, 32–33.

Nearey, T. M., and Assmann, P. (**1986**). "Modeling the role of vowel inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.

Nearey, T. M., Assmann, P. F., and Hillenbrand, J. M. (**2002**). "Evaluation of a strategy for automatic formant tracking," J. Acoust. Soc. Am. **112**, 2323.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Schmidt, A. M. (**1996**). "Cross-language identification of consonants. Part 1: Korean perception of English," J. Acoust. Soc. Am. **99**, 3201–3211.

Schmidt, A. M. (**2007**). "Cross-language consonant identification," in *Language Experience in Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production: A Festschrift in Honour of James E. Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 185–200.

Strange, W. (**2007**). "Cross-language phonetic similarity of vowels: Theoretical and methodological issues," in *Language Experience in Second Language Learning: In honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Philadelphia, PA), pp. 35–55.

Strange, W., Akahana-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). "Perceptual assimilation of American English vowels by Japanese listeners," J. Phonetics **26**, 311–344.

Strange, W., Bohn, O. S., Trent, S. A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–1807.

Talkin, D. (**1987**). "Speech formant trajectory estimation using dynamic programming with modulated transition costs," J. Acoust. Soc. Am. **82**, S55.

Thomson, R. I. (**2005**). "English vowel learning by speakers of Mandarin," J. Acoust. Soc. Am. **117**, 2400.

Thomson, R. I. (**2007**). "Modeling L1/L2 interactions in the perception and production of English vowels by Mandarin L1 speakers: A training study," Ph.D. thesis, University of Alberta, Edmonton, Alberta, Canada.

Thomson, R. I. (**2008**). "L2 English vowel learning by Mandarin speakers: Does perception precede production?," Can. Acoust. **36**, 134–135.

Wayland, R. P. (**2007**). "The relationship between identification and discrimination in crosslanguage perception: The case of Korean and Thai," in *Language Experience in Second Language Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Philadelphia, PA), pp. 201–218.

Cross-language categorization of French and German vowels by naïve American listeners

Winifred Strange[a)]
Ph.D. Program in Speech, Language, and Hearing Sciences, The City University of New York-Graduate School and University Center, 365 Fifth Avenue, New York, New York 10016-4309

Erika S. Levy
Department of Biobehavioral Sciences, Program in Speech and Language Pathology, Teachers College, Columbia University, 525 W 120th Street, Box 180, New York, New York 10027

Franzo F. Law II
Ph.D. Program in Speech, Language, and Hearing Sciences, The City University of New York-Graduate School and University Center, 365 Fifth Avenue, New York, New York 10016-4309

(Received 21 April 2008; revised 15 June 2009; accepted 18 June 2009)

American English (AE) speakers' perceptual assimilation of 14 North German (NG) and 9 Parisian French (PF) vowels was examined in two studies using citation-form disyllables (study 1) and sentences with vowels surrounded by labial and alveolar consonants in multisyllabic nonsense words (study 2). Listeners categorized multiple tokens of each NG and PF vowel as most similar to selected AE vowels and rated their category "goodness" on a nine-point Likert scale. Front, rounded vowels were assimilated primarily to back AE vowels, despite their acoustic similarity to front AE vowels. In study 1, they were considered poorer exemplars of AE vowels than were NG and PF back, rounded vowels; in study 2, front and back, rounded vowels were perceived as similar to each other. Assimilation of some front, unrounded and back, rounded NG and PF vowels varied with language, speaking style, and consonantal context. Differences in perceived similarity often could not be predicted from context-specific cross-language spectral similarities. Results suggest that listeners can access context-specific, phonetic details when listening to citation-form materials, but assimilate non-native vowels on the basis of context-independent phonological equivalence categories when processing continuous speech. Results are interpreted within the Automatic Selective Perception model of speech perception.
[DOI: 10.1121/1.3179666]

PACS number(s): 43.71.Hw, 43.71.Es [PEI] Pages: 1461–1476

I. INTRODUCTION

A theoretical tenet of current models of non-native and L2 speech perception (Best, 1995; Best and Tyler, 2007; Flege, 1995) is that phonetic similarities of L1 and L2 segments can, in significant part, predict relative difficulties that naïve listeners and L2 learners will have in distinguishing non-native phonetic segments. However, researchers differ in how they determine L1/L2 vowel similarities empirically. Some studies have used cross-language comparisons of the acoustic structure of vowels, primarily spectral similarity defined by relative locations in a target formant-frequency vowel space (e.g., Flege *et al.*, 1994), whereas others (e.g., Best *et al.*, 2003) refer to (abstract) articulatory-phonetic similarities of L1 and L2 vowels. A growing number of studies have employed direct measures of *perceived* L1/L2 similarities, referred to as cross-language categorization or perceptual assimilation tasks (see Strange, 2007b, for a critique of these techniques).

The present research employed a perceptual assimilation task to examine the perceived similarities of North German (NG) and Parisian French (PF) vowels to American English (AE) vowels by monolingual speakers of New York English (NYE). Of special interest was the perceptual assimilation of NG and PF front, rounded vowels, which are phonologically distinctive in both PF and NG (contrasting with both front, unrounded vowels and back, rounded vowels), but occur only as allophonic variants of back, rounded vowels in AE (Hillenbrand *et al.*, 2001; Strange *et al.*, 2007). Previous research (reviewed below) has produced conflicting results about AE listeners' relative perceptual difficulty with contrasts involving front, rounded vowels in French and German. In addition, research has also documented perceptual problems by AE listeners with other French and German contrasts that are also phonologically distinctive in AE, but that differ phonetically across languages. Thus, in the two studies reported here, naïve listeners were presented the full (oral) vowel inventories of PF (9 vowels) and NG (14 vowels).

Previous cross-language research on French and German vowels has shown that perception by AE listeners varies significantly as a function of both the speaking style in which the stimuli are produced and presented (e.g., lists vs sentences) and the consonantal context in which the vowels occur (Gottfried, 1984; Levy, 2009; Levy and Strange, 2008;

[a)]Author to whom correspondence should be addressed. Electronic mail: strangepin@aol.com

Strange *et al.*, 2004, 2005). Recent research on the acoustic structure of distributions of PF, NG, and AE vowels produced in different prosodic and phonetic contexts (Strange *et al.*, 2007) has documented language-specific patterns of contextual variation in spectral and temporal structure that lead to significant differences across contexts in the acoustic similarity of many PF and NG vowels relative to AE vowels. Thus, it was expected that L1/L2 perceived similarity relationships would also vary with prosodic and phonetic context. This contextual variation might account, at least in part, for the conflicting results of earlier research on AE listeners' perception of French and German vowel contrasts. In study 1, vowels produced in citation-form disyllables served as the stimuli. In study 2, the stimuli were vowels produced and presented in multisyllabic nonsense words embedded medially in short carrier sentences, with the vowels surrounded by labial and alveolar consonants.

Most studies of perceptual assimilation of non-native vowels by naïve listeners have reported group data for each set of L1 listeners. However, previous studies of AE listeners' perception of NG and PF vowels suggest that there may be significant differences among AE individuals in how NG and PF vowels are perceptually assimilated (Levy, 2009; Levy and Strange, 2008; Strange *et al.*, 2004, 2005). Thus, in both studies presented here, the data are reported in two ways: (1) overall categorization distributions and group median goodness ratings and (2) patterns of perceptual assimilation by individual listeners. The latter analysis allowed the authors to ask questions about differences in assimilation patterns across the two languages (and across contexts in study 2) using repeated measures analyses. However, due to practical considerations, independent groups of AE listeners served as participants in studies 1 and 2; both were drawn from the same NYE dialect group.

II. STUDY 1: GERMAN AND FRENCH VOWELS IN CITATION-FORM UTTERANCES

In the first study, NG and PF vowels were produced and presented in a "neutral" context ([hVbə] for NG vowels and [Vb(ə)] for PF vowels) that minimized coarticulatory influences of preceding and following consonants, while presenting closed syllables in which both tense and lax AE vowels are allowed phonologically. Most previous studies of AE listeners' perception of German and French contrasts presented citation-form monosyllables of the form #V#, CV, or CVC (Best *et al.*, 1996; Flege and Hillenbrand, 1984; Gottfried, 1984; Gottfried and Beddor, 1988; Polka, 1995; Strange *et al.*, 2004) or synthetically-generated #V# or CVC syllable continua (Gottfried and Beddor, 1988; Rochet, 1995). In the studies using CV or CVC syllables, the vowels were preceded and/or followed by alveolar consonants /d, t, s/. Our recent work on the acoustic variability of AE, NG, and PF vowels (Strange *et al.*, 2005, 2007) suggests that perception by AE listeners of back vs front, rounded NG and PF vowels may differ markedly in coronal and non-coronal contexts (see also Levy, 2009; Levy and Strange, 2008) due to the extreme allophonic fronting of AE [u:, ʊ, oʊ] in coronal contexts in most dialects of AE (cf., Hillenbrand *et al.*, 2001). Thus, results of the previous studies may not be representative of assimilation of front, rounded vowels (or indeed other non-native vowels) in general. The results of study 1 using vowels produced in a non-coronal context provided a basis for establishing cross-language perceived similarities of "canonical" NG and PF vowels (cf., Strange *et al.*, 2007) and may be contrasted with results of earlier studies using vowels surrounded by alveolar consonants.

Current models of cross-language and L2 speech perception (Best, 1995; Best and Tyler, 2007; Flege, 1995) are in agreement that, to be predictive of L2 perceptual difficulties, cross-language similarity relationships must be established at a level of description of phonetic segments that includes more detail than transcriptional equivalences or distinctive-feature characterizations of phoneme inventories. In his speech learning model (SLM), Flege (1995) posited that consonant and vowel categories in L1 and L2 are represented at the level of position-sensitive allophones. Best and Tyler (2007) claimed that listeners can be responsive both to phonologically-relevant phonetic information and to within-L1-category phonetic variation. In a perceptual assimilation task such as the one utilized here, categorization of L2 vowels as exemplars of the most similar L1 category by naïve listeners may reflect perceived similarity based on (L1) phonological categories, whereas ratings of category goodness may reflect perceptual attunement to gradient (perhaps language-universal) phonetic differences between L1 and L2 phones.

In the Processing Rich Information from Multidimension Interactive Representations (PRIMIR) model of the development of L1 speech processing, Werker and Curtin (2005) pointed out that performance by infants, children, and adults in speech perception experiments reflects the requirements of the language processing task as well as the initial biases and developmental level of the listener. Under the appropriate stimulus and task conditions, fine-grained, within-category phonetic information is available to adult listeners, even though they demonstrate highly over-learned language-specific patterns of perceptual processing in most online perception situations. In her Automatic Selective Perception model of speech perception, Strange (2006, 2007a, 2009; see also Strange and Shafer, 2008) outlined a similar account of the role of selective perception and attention in L1 and L2 speech perception. By this account, online L1 speech perception by adults is normally accomplished using highly over-learned selective perceptual routines (SPRs). These L1 SPRs enable the listener to extract the most reliable phonologically-relevant information rapidly from the incoming speech stream in order to recover the intended message (i.e., words specified by phonetic sequences), with few or no attentional resources required. However, when asked to differentiate segments that are not phonologically distinctive in the L1, performance may suffer because listeners' automatic L1 SPRs are not attuned to the appropriate phonetic information. Thus, listeners must resort to an attentional mode of perception in order to (learn to) detect the phonetic information that reliably distinguishes the non-native contrasts. This may be especially difficult when an L2 contrast is differentiated along phonetic dimensions that constitute allophonic variations in the L1. However, under relatively con-

strained stimulus and task conditions, naïve listeners are able to attend to within-L1-category phonetic details and make accurate discriminations. As the stimuli and tasks become more complex, listeners' performance may only reflect categorization on the basis of L1 SPRs.

According to this conceptual scheme, the task given naïve listeners in a perceptual assimilation study of nonnative contrasts explicitly directs them to attend to the phonetic details *in relation to* native phonological categories in making their responses. A question of interest, then, is how to characterize the nature of the phonetic information being used and whether it reflects a language-general or a language-specific mode of phonetic processing. If it is the former, the authors might expect that context-specific acoustic and underlying articulatory similarity relationships would be predictive of perceptual similarity patterns. Alternatively, listeners may respond on the basis of systematic patterns of language-specific phonetic variation characterized by L1 allophonic realization rules. The authors might expect, therefore, that categorization responses will reflect the relationship of the non-native segments to L1 phonological categories, in which (noncontrastive) allophonic variants are considered "equivalent" in terms of lexical specification. On the other hand, category goodness ratings might reflect detailed phonetic knowledge about the appropriateness of particular phonetic variants in particular contexts in the L1.

A previous study of AE listeners' perceptual assimilation of NG vowels in [hVp] syllables supports the latter hypothesis (Strange *et al.*, 2004). Front, rounded NG vowels were assimilated primarily to back AE vowels, despite the fact that they were acoustically more similar to AE front vowels or intermediate between front and back AE vowels in this context. However, front, rounded NG vowels were judged to be poorer exemplars of back AE categories than were the back NG vowels. [See Polka (1995) for similar results for Canadian English listeners' assimilation of front vs back, rounded vowels in dVt syllables.] In addition, Strange *et al.* (2004) reported that assimilation patterns for NG vowels with transcriptional counterparts in AE (so-called "similar" vowels) suggested that listeners were sensitive to cross-language differences in the phonetic realization of some (e.g., NG [e:]), whereas others were assimilated as good exemplars of their AE counterparts (e.g., [o:, ɛ]), despite cross-language differences in their relative locations in F1/F2/F3 vowel space.

Fewer data are available on the perceptual assimilation of citation-form French vowels by naïve AE listeners. Based on the *productions* of French [ty] and [tu] by AE L2 learners of French, Flege and Hillenbrand (1984) hypothesized that AE learners did not assimilate French [y] to any AE category, while they initially assimilated French [u] to AE [u]. However, this study did not include a perceptual assimilation task. In this context, AE [u] is highly fronted, which may account for the unexpected finding that inexperienced L2 speakers' productions of [ty] were more similar acoustically to native French speakers' [ty] productions (and identified more accurately by French listeners) than were their productions of [tu] to native French speakers' [tu] productions.

Gottfried (1984) examined discrimination of French vowel contrasts by naïve and experienced AE late L2 learners using a cross-speaker categorial ABX task. Results indicated that naïve listeners performed better on [u/y] and [y/ø] contrasts in #V# than in tVt context; the French [i/y] contrast was not tested. Gottfried (1984) also reported significant discrimination difficulties for naïve AE listeners on [i/e] and [e/ɛ] in both contexts. The poor discrimination of the front, rounded vowels [y/ø] reported by Gottfried (1984) differs from results reported by Best *et al.* (1996) on a single-speaker categorial discrimination test of CV French syllables [sy/sø], which most naïve AE listeners assimilated to two separate AE categories and discriminated with better than 95% accuracy. [See also Best *et al.* (2003) who reported very good discrimination of Norwegian front, in-rounded [sʉ] vs back [su], but less good discrimination of front, unrounded [si] vs out-rounded [sy].]

These discrepant results of perception of front, rounded vowels by naïve AE listeners indicate that performance can vary substantially as a function of stimulus, task, and listener factors, making comparisons across studies difficult. In the present study, the same listeners provided cross-language categorization and goodness ratings of NG and PF vowels produced by a representative speaker of each language, drawn from our earlier work on the acoustic structure of NG, PF, and AE vowels (Strange *et al.*, 2007). Thus, performance by the same naïve AE listeners could be compared directly across languages, with stimulus context and task demands held constant. This allowed the authors to ask the following questions:

(1) How do naïve AE listeners perceptually assimilate front, rounded NG [y:, ʏ, ø:, œ] and PF [y, ø] vowels in citation-form utterances? Specifically, the authors predicted the following.
   (a) Cross-language categorization patterns would reflect context-general spectral similarities of NG and PF front, rounded vowels to AE phonological categories, which include fronted allophones of back, rounded vowels. Despite their acoustic similarity to AE front vowels in this context, NG and PF front, rounded vowels (except for [œ]) would be categorized as more similar to back than to front AE vowels because high to mid, back AE vowels are allophonically fronted in alveolar contexts, whereas front AE vowels are rarely backed.
   (b) Listeners' goodness ratings would reflect their perception of context-specific phonetic differences between front and back, rounded NG and PF vowels; front, rounded vowels would be rated as poorer exemplars than back, rounded vowels of AE back vowels in a non-coronal context. Because PF [y] is more front (higher F2/F3 values) and spectrally closer to PF [i] than NG [y:] is to NG [i:], they also predicted that PF [y] would be considered a poorer exemplar than NG [y:] of AE [u:], or indeed might be judged as an exemplar of AE [i:] in this context by a majority of listeners.
   (c) Based on previous results, they predicted that AE listeners would show little sensitivity to duration differences in spectrally-similar NG [ʏ/ø:].

(2) How consistently are NG and PF mid-high to mid-low, front, unrounded and back, rounded vowels assimilated to their AE transcriptional counterparts? Specifically, they predicted the following.
   (a) NG and PF mid-high to mid-low vowels that are higher (F1 values lower relative to high vowels) in their respective vowel spaces than AE counterparts would not be consistently assimilated to their AE transcriptional counterparts. Strange *et al.* (2004, 2007) reported differences from AE in the relative heights of these vowels for both NG and PF, especially for front vowels [e, ɛ], which may account for previously reported perceptual difficulties.

(3) How are low vowels, NG [ɑ:, a] and PF [a], perceptually assimilated to AE vowels? Specifically, they predicted the following.
   (a) If categorized on the basis of spectral similarity, PF [a] would be assimilated more often to AE [æ:], while NG [ɑ:, a] would be judged as more similar to AE [ɑ:, ʌ], respectively.
   (b) Based on previous research, they predicted that duration differences between NG [ɑ:, a] would not contribute significantly to differences in perceived similarity for most AE listeners.

A. Method

1. Speakers and stimuli

Productions of one male speaker of NG and of PF were selected from the corpora of three speakers of each language analyzed in an earlier study of NG, PF, and AE vowel productions (Strange *et al.*, 2007). The 27-year-old NG speaker was from Fallingbostel in Northern Germany, while the 37-year-old PF speaker was from Les Mureaux, a Paris suburb. Both were proficient only in their native language and had been in the USA for only a short time. Details of recording procedures are available in the previous publication. For the present study, three tokens of each vowel category were selected and were verified as very good exemplars of each vowel category by a native NG and PF listener. For NG stimuli, vowels were produced in nonsense disyllables ([hVbə] spelled "Hieba, Hibba, Hehba …"), whereas PF vowels were produced by reading words spelled "hVb" but pronounced by PF speakers as /Vb/ with an audible voiced release of the final labial [Vb(ə)]. For a task familiarization procedure, 3 tokens of each of 11 AE vowels [i:, ɪ, eɪ, ɛ, æ:, ɑ:, ʌ, ɔ:, oʊ:, ʊ, u:] produced in citation-form [hVbə] disyllables by one of the AE male speakers (a 36-year-old native speaker of NYE dialect) from Strange *et al.* (2007) were also selected.

Figure 1 displays the mid-syllable spectral values (F1/F2 in barks) of the three tokens of each NG vowel (top plot) and each PF vowel (bottom plot) used in the perceptual assimilation tests. For comparison, these vowels are superimposed on ellipses depicting the range of values of 11 AE vowels produced by all three male speakers in citation disyllables from Strange *et al.* (2007). Since the point vowels [i, ɑ/a, u] were similar across the three languages, a comparison of other vowels was presumed to be meaningful with respect to dissimilarities in articulatory postures across the three languages, independent of vocal tract differences.

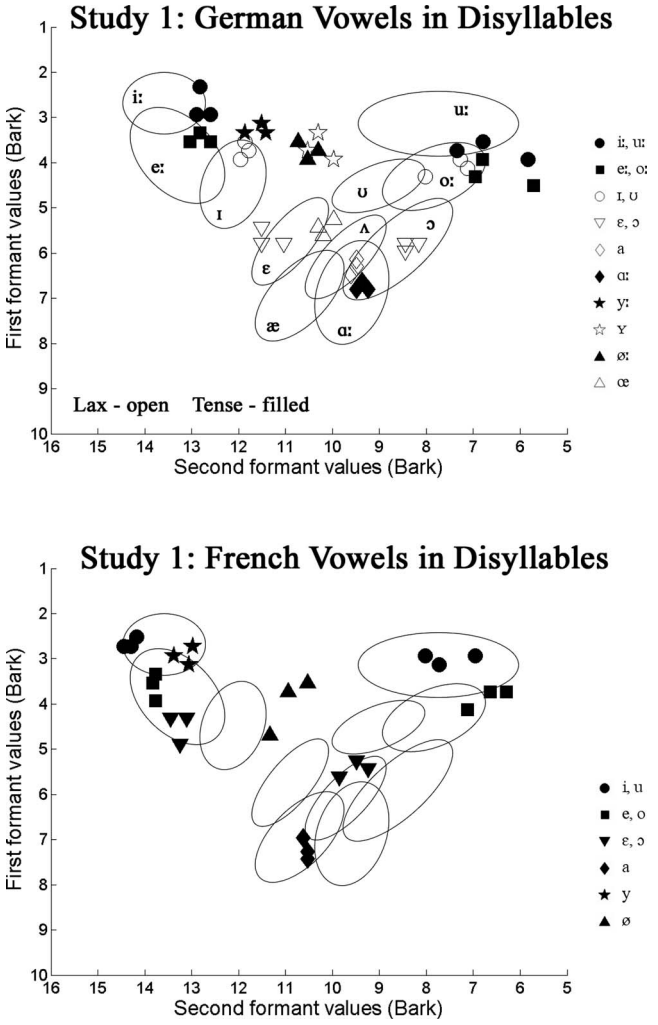


FIG. 1. NG (above) and PF (below) stimuli for study 1 (3 tokens of each vowel), superimposed on ellipses surrounding 11 AE vowels produced by 3 male speakers in citation-form hVb disyllables. For German vowels, short vowels are shown in open symbols, long vowel in closed symbols.

To quantify spectral similarities to AE vowels, the NG and PF vowel stimuli were submitted to linear discriminant analyses in which the AE vowel corpus (3 speakers × 4 tokens of each vowel) from Strange *et al.* (2007) served as the training set and the NG and PF stimuli served as the test sets. Classification of the NG and PF stimuli, relative to weighted centers of gravity for the 11 AE vowel categories, using F1/F2/F3 bark values as input parameters, can be summarized as follows (see Appendix, Table VI for classification results).

Front, rounded NG and PF vowels were almost all classified as closer to AE front than to back vowels. PF [y] was even more front (higher F2 values) and overlapped completely with AE [i:] in F1/F2 space (see Fig. 1). Second, NG and PF /e/ were located relatively high (low F1 values) and some tokens were classified as AE [i:, ɪ]. Third, PF [ɛ] was also very high (lower F1 values) relative to AE and NG [ɛ], with all tokens classified as AE [eɪ]. Finally, whereas the NG low vowels [ɑ:, a] were classified as back AE vowels [ɑ:, ʌ], PF [a] was consistently classified as more similar to front AE [æ:].

TABLE I. Perceptual assimilation of NG (A) and PF (B) back and front, rounded vowels to AE categories: Study 1—Citation materials. Modal AE V is the AE vowel chosen most often, summing over all 11 AE listeners. Percent is the overall percent of trials on which the modal response was chosen. (Rating) is the overall median goodness rating on modal response trials, summed over listeners. # of Ss is the number of AE listeners (out of 11) selecting a single response on at least 7 of 9 trials.

	Back Vs	Modal AE V	Categorization Percent	(Rating)	# of Ss	Front Vs	Modal AE V	Categorization Percent	(Rating)	# of Ss
(A)	uː	uː	94	(8)	10	yː	uː	77	(4)	8
	oː	oʊ	98	(8)	11	øː	ʊ	53	(2)	4[ʊ], 2[oʊ]
	ʊ	oʊ	65	(7)	4	ʏ	ʊ	62	(4)	5 [ʊ], 2 [uː]
	ɔ	oʊ	37	(4)	2	œ	ɛ	43	(5)	3 [ɛ], 1 [ʌ]
(B)	u	uː	84	(7)	9	y	uː	52	(2)	3 [uː], 1[iː]
	o	oʊ	69	(6)	7	ø	ʊ	48	(3)	3 [ʊ], 1 [oʊ]
	ɔ	ʌ	62	(6)	7 [ʌ], 1 [oʊ]					

2. Procedures and tests

All perceptual tests were performed on a personal computer located in a quiet testing room using a specialized program (PARADIGM ID by Bruno Tagliaferri) that controlled the task and recorded responses. Participants heard stimuli over STAX Professional SR Lambda earphones via a STAX Professional SRM-1/MK-2 amplifier with output set at a prearranged comfortable listening level. In these tests, a trial consisted of the following steps: (1) a stimulus was presented and response alternatives appeared on the screen (11 on-screen "buttons" labeled with hVd key words and IPA symbols), (2) the participant clicked on one of the buttons (categorization response) after which, (3) the same utterance was repeated and a nine-point Likert scale appeared (9 labeled "very American-like" and 1 labeled "very foreign sounding"), and (4) the participant clicked on the Likert scale to indicate the category goodness of the stimulus as an exemplar of the selected AE vowel category. During the NG and PF tests, no feedback was presented.

At the beginning of the experiment, each participant was given a task familiarization test that consisted of five blocks of trials using the AE stimuli. Correct/incorrect feedback was given for blocks 1–3 (11 vowels each). To pass criterion for inclusion in the study, participants could make no more than 2 errors total on block 4 or 5 (22 vowels each/2 tokens of each vowel), with no more than 1 error on any one vowel category. No feedback was given. If participants performed at 100% on block 4, they did not complete block 5.

The NG perceptual assimilation test consisted of 4 blocks of 42 stimuli—3 different tokens of each of 14 vowel categories, randomly ordered separately for each participant. The first block was used for familiarization with the nonnative stimuli and calibration of goodness ratings; the data were not included in the final scoring. Blocks 2–4 (nine judgments/vowel; three trials/token) served as the test data. The PF perceptual assimilation test had the same structure (4 blocks of 27 stimuli; 3 different tokens of each of 9 vowel categories). Again, the first block was used for familiarization and was not counted as test data. The order of languages was counterbalanced across participants with a pause between tests.

3. AE listeners

A total of 16 participants completed the AE familiarization task. Five participants were discontinued because of a failure to reach criterion; thus, 11 participants completed the NG and PF tests. None had any experience with French, German, or any other language with front, rounded vowels. Most had taken some foreign language classes in high school or college, but none could converse in any language but English by self-report. They were all current residents of the New York metropolitan area, had been raised in New York or New Jersey, and spoke a "standard" northeastern dialect in which [ɑː/ɔː] are differentiated. The authors can assume that these listeners' vowel spaces resembled those presented in Fig. 1 for AE speakers from the same dialect population. The participants ranged in age from 22 to 54 years of age and reported normal hearing.

B. Results

For the group analysis, overall categorization distributions (summed over 11 participants × 9 trials = 99 responses/vowel) were computed. For each NG and PF vowel, the modal AE categorization response, the overall consistency in the choice of that modal category as a percentage of total trials, and the overall median[1] goodness rating for trials on which the modal response alternative was selected were computed. For analyses of individual data, each listener's responses to each NG and PF vowel were designated as categorized (the same AE response chosen on at least 7 out of 9 trials)[2] or uncategorized. The number of listeners (Ss) with categorized responses for each NG and PF vowel was tallied.

1. Back and front, rounded vowels

Group categorization results for back (columns 1–4) and front (columns 6–9), rounded vowels are shown in Table I. In addition, the number of AE listeners (out of 11) who showed consistent categorization of each vowel is given (columns 5 and 10), with vowels indicated in brackets when AE vowels in addition to the group modal vowel were selected consistently by some individuals.

Regarding the back, rounded vowels, both group and individual data indicate that NG [uː, oː] were consistently categorized as most similar to their AE transcriptional coun-

TABLE II. Perceptual assimilation of NG (A) and PF (B) front, unrounded and low vowels to AE categories: Study 1—Citation materials. Modal AE V is the AE vowel chosen most often, summing over all AE listeners. Percent is the overall percent of trials on which the modal response was chosen. (Rating) is the overall median goodness rating on modal response trials, summed over listeners. # of Ss is the number of AE listeners (out of 11) selecting a single response on at least 7 of 9 trials.

	Front NG	Modal AE V	Categorization Percent	(Rating)	# of Ss	Low NG	Modal AE V	Categorization Percent	(Rating)	# of Ss
(A)	iː	iː	91	(7)	10	ɑː	ɑː	87	(7)	10
	eː	eɪ	85	(7)	10	a	ɑː	61	(7)	5 [ɑː] 1 [ʌ]
	ɪ	ɪ	87	(8)	9					
	ɛ	ɛ	95	(7)	10					
	Front PF	Modal AE V	Categorization Percent	(Rating)	# of Cons. Subjects	Low PF	Modal AE V	Categorization Percent	(Rating)	# of Ss
(B)	i	iː	96	(8)	11	a:	ɑ	80	(6)	8 [ɑː]
	e	eɪ	80	(7)	9					
	ɛ	ɛ	82	(6)	8 [ɛ] 1 [eɪ]					

terparts by almost all the listeners and were judged as very good exemplars of those categories. PF [u, o] were somewhat less consistently assimilated to their AE counterparts within and across listeners, with lower median goodness ratings. In contrast, NG [ʊ, ɔ] were not consistently assimilated as examples of their transcriptional counterparts in AE, although a few listeners categorized them both as AE [oʊ]. A majority of the listeners assimilated PF [ɔ] to AE [ʌ].

Turning to the front rounded vowels, the group data indicate somewhat less consistency within and across listeners in how these vowels were categorized, relative to the back vowels. However, except for NG [œ], the group modal response was an AE back vowel. Summing over all AE back vowel responses, NG [y:, ø:, ʏ] were perceptually assimilated to AE back vowels on 99%, 96%, and 89% of trials, respectively, whereas NG [œ] was assimilated to AE back vowels 52% of the time. PF [y] and [ø] were assimilated to AE back vowels 84% and 95% of the time, respectively. Thus, except for NG [y:], the front, rounded vowels could be considered uncategorizable back vowels for many AE listeners. Only one AE listener consistently heard PF [y] as most similar to AE [i:]. None of the listeners heard NG [y:] as most similar to a front AE vowel and eight listeners assimilated it consistently to AE [u:]. Comparing assimilation of NG [y:] and PF [y], significantly more subjects consistently categorized NG [y:] as AE [u:] (8 vs 3; significant by a Fisher's Exact test, $p=0.04$, one-tailed test). Perceptual assimilation patterns for NG [ø:, ʏ], which were spectrally very similar but differed in duration, gave only weak evidence that vowel duration was used by AE listeners in categorization decisions; both were assimilated primarily to short AE [ʊ].

An inspection of overall goodness ratings (columns 4 and 9) shows that, on average, NG and PF front, rounded vowels were considered poorer exemplars of AE back vowels than were NG and PF back vowels except for NG [œ, ɔ]. 10 of the 11 listeners rated NG [u:] as a better exemplar than NG [y:] of an AE back vowel ($p<0.02$ by a Sign test evaluated by the binomial expansion).[3] PF [u] was also rated as a better fit to AE back vowels than PF [y] for the three listeners who assimilated [y] to [u]; one listener who assimilated [y] to [i] rated it as a poorer exemplar than PF [i], and the remainder were inconsistent in their categorizations of one or both vowels, making goodness ratings difficult to interpret. In comparing ratings for NG [y:] with PF [y], 8 of the 9 listeners for whom a particular back AE vowel was the model response rated the NG vowel as a significantly better exemplar than PF [y] of that back vowel AE (Wilcoxon Signed-Ranks test, $N=9$, $T+=44$, $p<0.01$).

For NG [ø:, o:] and PF [ø, o], relative goodness ratings also reflected the judgment that front, rounded vowels were poorer exemplars of AE back vowels than were back, rounded vowels. All 11 participants rated NG [o:] as a better fit than NG [ø:] to some AE back vowel ($p<0.01$ by a Sign test); 10 of the 11 participants also rated PF [o:] as a better exemplar than PF [ø:] of some AE back vowel ($p<0.02$ by a Sign test). Because of the great variability within and across participants in the assimilation of NG [ʏ, œ] to various AE categories, an analysis of differences in goodness ratings was less interpretable.

2. Front, unrounded and low vowels

As Table II (columns 1–5) shows, both NG and PF front, unrounded vowels were perceptually assimilated primarily to their transcriptional counterparts in AE and considered very good exemplars of those categories. This reflects generally good consistency both within and across listeners in assimilation patterns. However, a few listeners were inconsistent in their perceptual assimilation of PF [e, ɛ] and one listener consistently assimilated PF [ɛ] to AE [eɪ]. The remaining eight listeners judged NG and PF [ɛ] to be equally good exemplars of AE [ɛ], despite large differences in F1 values (see Fig. 1). This can be contrasted with back NG [ʊ, ɔ] and PF [ɔ, o] (shown in Table I), which were not consistently assimilated to their AE counterparts.

As shown in Table II (columns 6–10), NG [ɑː] was assimilated as a relatively good match to AE [ɑ:] by all but one

listener. In contrast, NG and PF [a] were less consistently assimilated to AE vowels both within and across individual listeners. Again, most listeners' perceptual assimilation of the spectrally-similar NG [ɑ:/a] appeared not to reflect an influence of vowel duration on categorization patterns.

C. Discussion

These results generally replicated earlier research with respect to how front, rounded vowels in citation-form materials are perceptually assimilated by naïve AE listeners. They differed somewhat with respect to assimilation patterns for similar NG and PF vowels that vary phonetically from their AE transcriptional counterparts. In addition, patterns of assimilation of PF vowels, relative to NG vowels, could be evaluated for the same AE listeners.

Cross-language categorization patterns suggested that responses on NG [y:, ʏ, ø:] and PF [y, ø] were based primarily on context-independent relationships to AE phonological categories, i.e., their acoustic similarity to AE vowels produced in this context did not predict perceptual assimilation patterns. Rather, as predicted from previous research, these vowels were categorized as more similar to back than to front AE vowels. Only one participant assimilated PF [y] to AE [i], and none heard NG [y:] as more similar to an AE front vowel. In an earlier study with different NG speakers, but similar materials (Strange *et al.*, 2004), only 3 out of 12 AE listeners categorized NG [y:] as more similar to front than to back AE vowels. In the present study, NG [ʏ, ø:] and PF [ø] were also categorized as more similar to AE back vowels by all 11 listeners. In the earlier study, 2 out of 12 AE listeners assimilated these NG vowels to front AE categories.

A comparison of category goodness ratings indicated that almost all the listeners heard NG [y:, ø:] and PF [y, ø] as poorer exemplars of AE back vowel categories than NG [u:, o:] and PF [u, o]. This may be characterized as a category-goodness assimilation pattern for front vs back, rounded contrasts in the Perceptual Assimilation Model (PAM) framework (Best, 1995). Second, the finding of poorer within-listener consistency in categorization of the front, relative to the back, rounded NG and PF vowels also suggests that listeners detected phonetic differences between them. This might be interpreted as reflecting an uncategorized-categorized pattern according to PAM. Thus, in this study, listeners appeared to be able to access detailed phonetic information about the deviation of front, rounded NG and PF vowels from AE back vowels in this non-coronal context including, for most listeners, the perception that PF [y] was more deviant than NG [y] as an exemplar of any AE back vowel. The authors would expect then that in this context with these materials, discrimination of front/back rounded vowel contrasts would be significantly above chance for naïve AE listeners, although perhaps not as accurate as native listeners' performance.

NG [œ], which was acoustically similar to AE [ɛ], was nevertheless uncategorized for 8 of 11 AE listeners; these results replicate the results of Strange *et al.* (2004) for this vowel. With respect to the PAM taxonomy, NG [ɛ/œ] in this context was a category-goodness contrast for a minority of AE listeners, whereas for most, it was categorized-uncategorized. To our knowledge, this contrast has not been tested in studies of naïve AE listeners' perception of NG vowels.

The acoustic realization of NG [e:, o:] and PF [e, o] as somewhat higher on average than AE [eɪ, oʊ], as well as the fact that these NG and PF vowels are not diphthongized, led to the expectation that they would be perceived as more similar to higher AE vowels or as poor tokens of AE mid vowels. However, in an earlier study (Strange *et al.*, 2004), NG [o:] was consistently categorized as a relatively good exemplar of AE [oʊ], despite its acoustic dissimilarity, whereas 9 of the 12 listeners assimilated NG [e:] to AE [i:, ɪ]. The results of the present study did not replicate this finding. Here, NG [e:, o:] were both consistently categorized as their AE counterparts by most of the listeners. However, fewer listeners consistently categorized PF [e, o] as their AE counterparts, with more assimilation responses to higher AE vowels. These patterns of perceptual assimilation are only partially predictable from context-specific spectral similarity relationships (see the Appendix).

Listeners in the present study also categorized NG [ɛ] somewhat more consistently than PF [ɛ] to the AE counterpart; the NG results replicate the earlier finding for this vowel. In contrast, NG and PF [ɔ] were both poor perceptual matches to any AE vowel, even though NG [ɔ] was spectrally quite similar to its AE counterpart. This could be due to the fact that in NYE dialect, this vowel tends to be long and heavily diphthongized in some speakers' productions.[4] As in the earlier study, NG [ʊ] was not assimilated to its AE counterpart. However, although it was inconsistently categorized as a very poor exemplar of any AE vowel in the earlier study, here it was more consistently categorized as similar to AE [oʊ], as was predictable from the analysis of acoustic similarity. Finally, as in the earlier study, NG [ɪ] was consistently assimilated to its AE counterpart by most listeners.

On the basis of spectral and temporal similarities, differences in the assimilation of NG and PF low vowels were expected. However, assimilation patterns on NG [ɑ:, a] suggested that only a few listeners differentiated these vowels on the basis of their temporal and (small) spectral differences. For most listeners, this NG contrast constituted a single-category assimilation pattern according to the PAM (Best, 1995). Most listeners also categorized PF [a] as a good exemplar of AE [ɑ:], despite its short duration and greater spectral similarity to AE [æ:].

These data reveal patterns of perceptual similarity of NG and PF vowels to AE categories in citation-form utterances (non-coronal context). In the next study, the effects of contextual variation on assimilation patterns for vowels produced in sentence materials were explored. This allowed the authors to establish the extent to which listeners had access to context-specific phonetic information during online processing of continuous speech input.

III. STUDY 2: GERMAN AND FRENCH VOWELS IN SENTENCES

In an earlier acoustic study comparing NG, PF, and AE vowels in sentence materials, striking differences across lan-

guages in the contextual variation in coarticulated vowels were revealed (Strange *et al.*, 2007). For NG vowels in labial context, mid-syllable formant frequencies for all 14 vowels differed little from canonical values. In alveolar context, NG short vowels [a, ɔ, ʊ] showed some coarticulatory fronting and raising, but the remaining vowels changed little. In contrast, PF vowels in both contexts in sentence materials varied considerably from canonical values derived from citation-form utterances. In alveolar context especially, PF low and back vowels [a, ɔ, o, u] showed more coarticulatory raising and/or fronting than NG vowels, whereas front, unrounded vowels and [y] were slightly more back. Finally, AE vowels showed very small shifts from canonical values for all 11 vowels in labial context, but extreme fronting of high to mid back vowels [u:, ʊ, o:] and some raising and fronting of [ɛ, ʌ] in alveolar contexts.

Cross-language differences in coarticulatory patterns gave rise to notable differences in spectral similarity of NG and PF vowels to AE categories, as established by cross-language discriminant analyses of six speakers of each language (Strange *et al.*, 2007). NG and PF /y, ø/ were more similar to front AE vowels in labial context, whereas they were more similar to fronted allophones of back AE vowels in alveolar context, except for PF [y], which was still more similar to AE [i]. NG [ʏ, œ] were more similar to back rounded AE vowels in both contexts. Other NG and PF vowels also changed their spectral similarity to their AE transcriptional counterparts; NG [ɪ, e:, ɛ] and PF [ɛ] were somewhat better matches than in citation-form materials, while NG and PF [a] were spectrally more raised relative to their canonical targets and to AE low vowels.

Contextual variations in cross-language spectral and temporal similarity were predicted to affect perceptual assimilation patterns in the present study. It was also predicted that perceptual assimilation patterns might differ from those found in study 1 because here listeners were asked to judge cross-language similarity "on the fly" while listening to continuous speech utterances. A question of interest was the extent to which context-specific phonetic information was accessible to listeners as they made categorization responses and goodness judgments.

Using similar sentence materials, Strange *et al.* (2005) reported that naïve AE listeners did not differ systematically in overall categorization consistency or in median goodness ratings of NG front rounded vowels produced in labial, alveolar, and velar contexts; they were assimilated as fair exemplars of back AE vowels in all contexts. In that study, the context in which the vowels occurred varied randomly from trial to trial. In another study (Strange *et al.*, 2004) with NG vowels produced in [hVp] context in the same carrier sentence, all 12 AE listeners categorized NG front, rounded vowels (except for [œ]) as most similar to back AE categories, but judged them to be relatively poorer exemplars than NG back, rounded vowels. Thus, it is not clear whether it was the coarticulatory variability or the contextual uncertainty that led to the failure of listeners in Strange *et al.* (2005) to reflect context-specific differences in similarity of front and back, rounded NG vowels to AE vowels in their categorization and goodness ratings. In addition, whereas some similar NG vowels were assimilated differently in sentence and citation materials, there were few differences in assimilation patterns across labial, alveolar, and velar contexts in categorization consistency or judged goodness. It appears that listeners adopted a context-independent strategy for categorizing and rating the NG vowels when the immediate context varied from trial to trial and when the target syllables were embedded in sentence-length utterances.

Levy and Strange (2008) used a cross-speaker AXB discrimination task to examine naïve AE listeners on perceptual differentiation of front, rounded PF vowels. Stimuli were disyllables /raCVC/ in labial and alveolar contexts in phrase-length utterances, with contexts presented in separate blocks. The results supported the hypothesis that [y] was assimilated to a front AE vowel in labial context and to a back AE vowel in alveolar context by the majority of naïve listeners; [ø] was assimilated as a relatively poor exemplar of a back vowel in both contexts. However, no perceptual assimilation data were gathered on these listeners.

In follow-up experiments, Levy (2009) reported perceptual assimilation data for naïve and experienced L2 learners of French. The AE response alternatives included palatalized [^{j}u] (as in "hue") and /ɝ/ (as in "herd"), as well as the other 11 vowel categories used in previous research (Strange *et al.*, 2004, 2005) and the present study. Perceptual assimilation patterns for naïve listeners showed significant differences as a function of context: PF [y] was judged more often as similar to [^{j}u] in labial than in alveolar context, and there were more assimilations to AE [i] in labial context than in alveolar context. PF [ø][5] also differed with context for naïve listeners. The results suggest that both PF front and back, rounded vowels were perceived as similar to back AE [u] with the exception of PF /y/ in bilabial context.

In the present study, naïve AE listeners were tested on the complete (oral) vowel inventories of both NG and PF in a repeated measures design. Vowels produced in labial and alveolar contexts were presented in blocked format; order of context and of language was counterbalanced across listeners. Several questions were of interest.

(1) When listeners could anticipate the consonantal context, would assimilation of coarticulated NG and PF front, rounded vowels produced in labial and alveolar contexts reflect a context-specific mode of processing? If it did, we predicted the following.
   (a) Front, rounded NG and PF vowels (especially [y]) produced in labial context would be assimilated as poorer exemplars of AE back vowels than those produced in alveolar context.
   (b) PF [y] would be judged a poorer exemplar than NG [y:] of AE [u:] in labial context.

(2) Would judged goodness of NG and PF back, rounded vowels differ with context, given cross-language differences in their coarticulatory fronting? Specific predictions were the following.
   (a) NG [u:, o:] would be considered poorer exemplars of back AE categories in alveolar than in labial context because they were not fronted as much as their AE transcriptional counterparts.

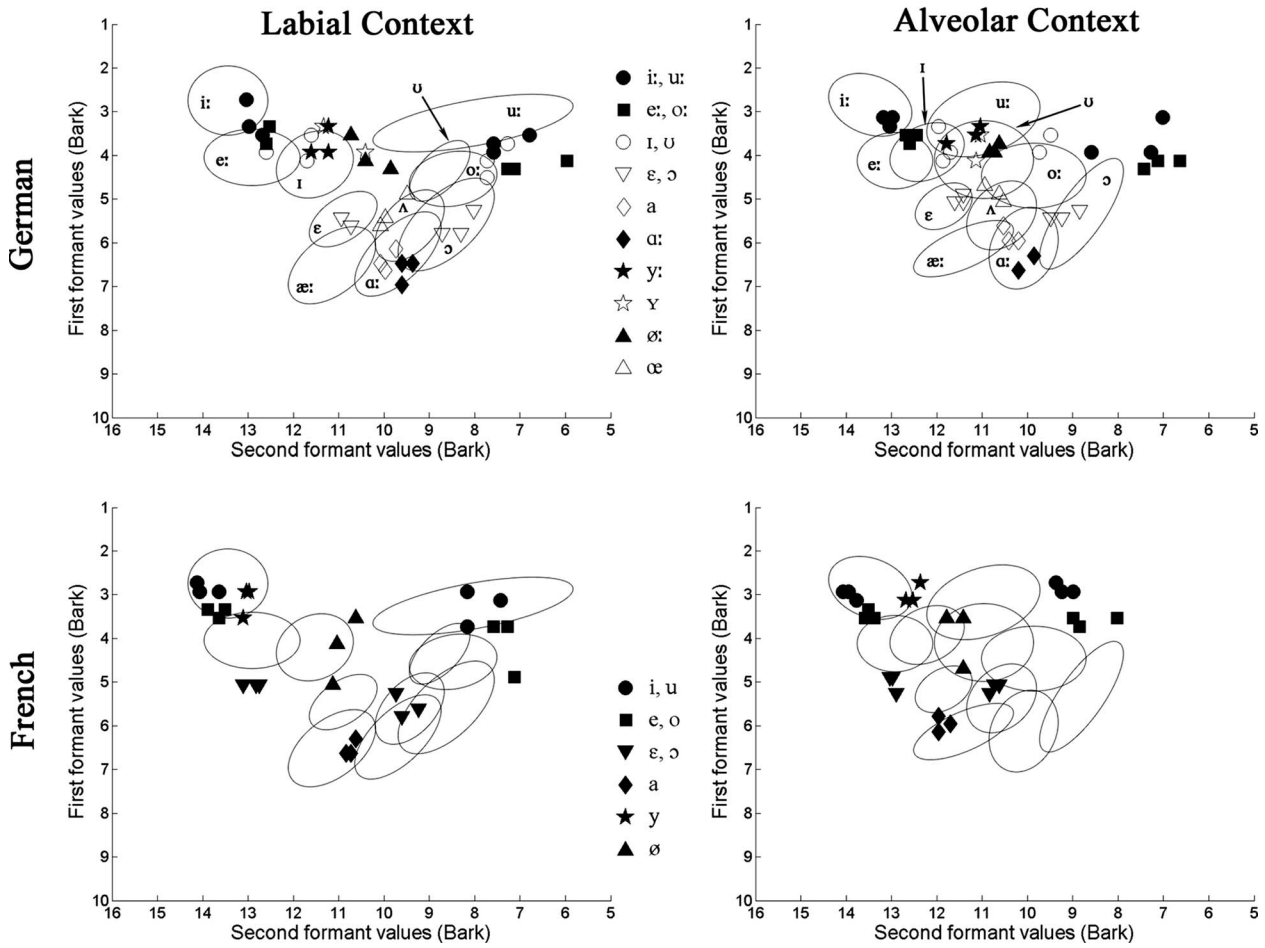


FIG. 2. NG (above) and PF (below) stimuli for study 2 (3 tokens of each vowel). Vowels were produced in labial context (left) and alveolar context (right) in nonce words embedded in sentences. Ellipses surround 11 AE vowels produced by 3 male speakers in the same sentence materials (labial context left, alveolar context right). For German vowels, short vowels are shown in open symbols, long vowels in closed symbols.

(b) PF back, rounded vowels would be considered better exemplars than NG back, rounded vowels of their AE counterparts in both contexts because they were more fronted and thus more similar to AE vowels.

(3) Would perceptual assimilation of low NG and PF vowels reflect relative differences in contextual raising across contexts? Specific predictions were the following.
(a) PF [a] would be assimilated more often than NG [a] to higher AE vowels, especially in labial context, where it was raised relative to canonical values.
(b) NG [a] would be assimilated more often to higher AE vowels in alveolar than in labial context because of its raising in alveolar context.

A. Method

1. Speakers and stimuli

The same two male native speakers of NG and PF as in study 1 produced the stimuli for study 2 (see Strange *et al.*, 2007). NG vowels were produced in the carrier sentence, "Ich habe fünf /gəCVCə/ gesagt;" PF vowels were produced in the carrier sentence, "J'ai dit neuf /ra/CVC/ à des amis." The target vowels occurred in the stressed or prominent syllable of the nonsense word; the consonantal contexts were bVp and dVt.[6] Recording procedures and stimulus selection were the same as in study 1; speakers produced the sentences fluently at a speaking rate that was appropriate for speaking to native listeners.

Figure 2 displays the stimuli in F1/F2 bark space, superimposed over the distributions of AE vowels produced by three males (Strange *et al.*, 2007) in labial and alveolar contexts in multisyllabic utterances embedded in a carrier sentence ("I said five gəCVCə this time"). To establish acoustic similarity patterns, context-specific linear discriminant analyses were performed in which the AE vowels produced in labial and alveolar contexts by the male AE speakers served as the training sets and the NG and PF stimulus materials served as the test sets. Two sets of analyses were conducted; one in which F1/F2/F3 bark values served as the input parameters for both training and test sets, and a second where vocalic duration was added as a fourth parameter. The Appendix presents the classification data for the first analyses of cross-language spectral similarities since the inclusion of duration had little effect on classification patterns, except for NG and PF [a] (more classifications as short [ʌ, ɛ], respectively).

There were clear differences in cross-language spectral similarity as a function of consonantal context as well as differences in acoustic similarity of the "same" NG and PF vowels to transcriptional counterparts in AE. Both NG and PF back, rounded vowels were, in general, more spectrally dissimilar from AE counterparts than in Study 1. In labial context, NG [o:, ɔ] were better spectral matches to AE counterparts than were PF [o, ɔ], whereas PF [u] was a better match than NG [u] to its AE counterpart. In alveolar context, the spectral dissimilarity of both NG and PF back, rounded vowels to AE counterparts was even greater because the AE high to mid back vowels in this context were produced as fronted allophones (see Fig. 2).

NG [y:, ʏ, ø:]) were classified primarily as front AE vowels in labial context, whereas they were classified as back AE vowels when produced in alveolar context due to the extreme fronting of back AE vowels in this context. In contrast, PF front, rounded vowels were spectrally more similar to AE front vowels in both contexts (i.e., they were more front than NG vowels).

Front, unrounded NG vowels [i:, e:, ɪ, ɛ] were quite similar spectrally (and temporally) to their AE counterparts in both contexts, whereas PF [e, ɛ] were higher relative to AE (and NG) counterparts. Finally, while NG low vowels [ɑ:, a] were both spectrally most similar to AE [ɑ:], PF [a] was spectrally more similar to AE [æ:] in both contexts. When vocalic duration was included as an input parameter, results indicated that NG and PF [a] and NG [ɔ] were more similar to AE [ʌ].

2. Procedures and tests

Testing procedures followed the same structure as in study 1, except that listeners heard two separate tests for each language: target vowels in labial context and in alveolar context. The orders of languages and contexts within languages were counterbalanced across participants. Prior to testing, listeners completed a familiarization test in which they heard AE vowels produced in sentences "I said five gaCVCa this time" with the vowel and consonantal context varying randomly. For this study, the key words were changed to "eek" [i:], "if" [ɪ], "ache" [$e^{ɪ}$], "heck" [ɛ], "as" [æ:], "ah" [ɑ:], "awe" [ɔ];, "uh" [ʌ], "hook" [ʊ], "oh" [o:], and "ooze"[u:] so that they contained no labial or alveolar stops. There were five familiarization blocks; feedback was given on blocks 1–4, but not on block 5. Listeners were required to make no more than 2 errors on block 5 (22 items) and no more than one error on any one vowel. Listeners then completed four tests of four blocks each for each context and each language. The first block of each test was for familiarization; data from the final three blocks (nine judgments/vowel/context) were retained for analysis. Subjects were tested in the same environment using the same equipment as in study 1.

3. Listeners

Participants were drawn from the same population of NYE speakers as in study 1. None had any experience with German or French or any other language with front, rounded vowels. A total of 28 participants was tested in the familiarization task. 11 failed to meet the criterion for further participation, and 1 who met the familiarization criterion failed to follow instructions in subsequent testing and was not included. Thus, 16 listeners completed all 4 experimental tests (4 male and 11 female); their average age was 31 years (range from 21 to 48 years). They all reported normal hearing.

B. Results

As in study 1, data on blocks 2–4 for all 16 participants on each NG and PF test were tallied, and categorization responses for each NG and PF vowel in each context were summed over the 16 listeners ($9 \times 16 = 144$ total trials for each vowel in each context and language). Overall modal responses and median goodness ratings on trials on which the modal response was selected were derived. In addition, the number of listeners who consistently (at least 7 out of 9 responses) categorized each NG and PF vowel as a particular AE vowel was tallied.

1. Back and front, rounded vowels

Given the differences across languages in the fronting of back, rounded vowels, it was expected that NG [u:, o:] might be heard as relatively poorer exemplars of their AE counterparts than in study 1. As Table III indicates, this proved true, especially in alveolar context; both group and individual data showed these vowels to be less consistently assimilated and with lower median goodness ratings than in study 1. NG [u:] was consistently assimilated to its AE counterpart by significantly fewer listeners in alveolar than in labial context [Fisher's Exact test, $p < 0.04$). In contrast, PF [u] was a better fit to its AE counterpart in both contexts relative to study 1, although goodness ratings were slightly lower. PF [o] was less consistently categorized overall than PF [u] in labial context and fewer listeners consistently assimilated PF [o] than PF [u] to their AE counterparts (Fisher's Exact test, $p < 0.01$, two-tailed test). Fewer individuals were consistent in assimilating PF [o] to its AE counterpart in labial than in alveolar context, but this difference was not statistically reliable.

NG [ʊ, ɔ] and PF [ɔ] showed very inconsistent patterns of assimilation both within and across listeners in both contexts. More listeners consistently categorized NG and PF [ɔ] as AE [ɔ:] in alveolar than in labial context (Fisher's Exact tests, $p < 0.06$; $p < 0.04$, respectively), but approximately half the listeners were inconsistent in their categorization of this vowel in both contexts. Finally, NG [ʊ] was categorized inconsistently by all but one listener in labial context and all but three listeners in alveolar context, with responses distributed over all back AE vowels for the other participants. In general then, these mid-high and mid-low back, rounded NG and PF vowels can be considered uncategorized back vowels in both consonantal contexts for about half the AE listeners. This reflects, in part, their spectral differences from AE vowel distributions (Fig. 2) as well as differences in length and diphthongization from NYE dialect.

TABLE III. Perceptual assimilation of NG and PF back, rounded vowels to AE categories: Study 2—Sentence materials. Modal AE V is the AE vowel chosen most often, summing over all AE listeners. Percent is the overall percent of trials on which the modal response was chosen. (Rating) in the overall median goodness rating on modal response trials, summed over listeners. # of Ss is the number of AE listeners (out of 16) selecting a single response on at least 7 of 9 trials.

	Labial context					Alveolar context				
	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(A)	uː	uː	86	(6)	15	uː	uː	67	(5)	9
	oː	oʊ	97	(7)	15	oː	oʊ	74	(7)	11 [oʊ], 3 [ɔː]
	ʊ	oʊ	47	(6.5)	1 [uː]	ʊ	ʊ	38	(5)	2 [ʊ], 1 [uː]
	ɔ	ʌ	30	(6)	2 [ɔː], 1 [oʊ], 1 [ʌ]	ɔ	ɔː	61	(6)	8 [ɔː]
	Labial context					Alveolar context				
	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(B)	u	uː	94	(6)	16	u	uː	92	(6)	14
	o	oʊ	60	(6)	7 [oʊ], 3 [uː]	o	oʊ	71	(6)	10 [oʊ], 1 [uː], 1 [ɔː]
	ɔ	ʌ	38	(7)	4 [ʌ], 2 [oʊ], 1 [ɔː]	ɔ	oʊ	47	(5)	5 [ɔː], 4 [oʊ]

NG front, rounded vowels in both labial and alveolar contexts (Table IV) were assimilated more often to back than to front AE categories, although the overall consistency with which they were assimilated to particular AE back vowels differed considerably across particular NG vowels. PF front, rounded vowels were also assimilated primarily to AE back vowels in both contexts, although overall consistency varied by context for [y] and the group modal AE vowel categories differed across contexts for [ø].

NG [y:] and PF [y] were assimilated primarily to AE [u:] in labial context. Whereas the group overall consistency and median goodness ratings suggested that PF [y] was heard as less similar than NG [y:], this was due primarily to two listeners who assimilated PF [y] consistently to AE [i:]. For the remaining listeners, there was no significant difference in number of listeners who consistently categorized each vowel as AE [u:](13 vs 11), nor in goodness ratings across the two languages (Wilcoxon Signed-Ranks test, $N=10$, $T+=38$, n.s.). In alveolar context, both NG [y:] and PF [y] were assimilated overwhelmingly to AE [u:]; only three participants were inconsistent in categorizing either NG [y:] or PF [y]. For the other 13 listeners, only 5 rated the NG [y:] as a better exemplar than the PF [y] of AE [u:] (Wilcoxon Signed-Ranks test, $N=13$, $T+=54.5$, n.s.). Thus, in both contexts, NG [y:] and PF [y] were assimilated as relatively good exemplars of AE [u:] by most naïve AE listeners. Indeed, these vowels

TABLE IV. Perceptual assimilation of NG and PF front, rounded vowels to AE categories: Study 2—Sentence materials. Modal AE V is the AE vowel chosen most often, summing over all AE listeners. Percent is the overall percent of trials on which the modal response was chosen. (Rating) is the overall median goodness rating on modal response trials, summed over listeners. # of Ss is the number of AE listeners (out of 16) selecting a single response on at least 7 of 9 trials.

	Labial context					Alveolar context				
	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(A)	yː	uː	83	(6)	13	yː	uː	92	(6)	15
	øː	uː	31	(5)	2 [uː], 2 [ʌ] 1 [ʊ]	øː	ʊ	30	(3)	2 [ʊ], 2 [uː] 1[ʌ], 1[oʊ]
	ʏ	uː	44	(5)	3	ʏ	uː	52	(6)	4 [uː], 2 [ʊ]
	œ	ʌ	75	(6)	9	œ	ʌ	69	(6)	8 [ʌ], 1 [ʊ] 1[oʊ]
	Labial context					Alveolar context				
	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(B)	y	uː	74	(4)	11 [uː], 2 [iː]	y	uː	94	(6)	14
	ø	ʌ	38	(4)	3 [ʌ], 1 [uː] 1 [ʊ]	ø	oʊ	44	(4)	3 [oʊ], 1 [uː]

TABLE V. Perceptual assimilation of NG and PF front, unrounded and low vowels to AE categories: Study 2—Sentence materials. Modal AE V is the AE vowel chosen most often, summing over all AE listeners. Percent is the overall percent of trials on which the modal response was chosen. (Rating) is the overall median goodness rating on modal response trials, summed over listeners. # of Ss is the number of AE listeners (out of 16) selecting a single response on at least 7 of 9 trials.

	Labial context					Alveolar context				
	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss	NG V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(A)	iː	iː	100	(8)	16	iː	iː	100	(7)	16
	eː	eɪ	71	(7)	10 [eɪ], 3 [iː]	eː	eɪ	87	(7)	14
	ɪ	ɪ	75	(7)	9	ɪ	ɪ	94	(7)	15
	ɛ	ɛ	99	(7)	16	ɛ	ɛ	94	(7)	14
	ɑː	ɑː	84	(7)	13	ɑː	ɑː	88	(7)	12
	a	ɑː	84	(7)	12	a	ʌ	74	(6)	8 [ʌ], 1 [ɑː]
	Labial context					Alveolar context				
	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss	PF V	Modal AE V	Categorization Percent	(Rating)	# of Ss
(B)	i	iː	97	(7)	16	i	iː	90	(7)	13
	e	eɪ	51	(6)	6 [eɪ], 1 [iː], 1 [ɪ]	e	eɪ	40	(6)	2 [ɪ], 1 [iː]
	ɛ	ɛ	79	(6)	12	ɛ	ɛ	85	(6)	13
	a	ɑː	64	(6)	7 [ɑː], 2 [ʌ], 1 [ɛ]	a	ɑː	49	(7)	7 [ɑː], 3 [ɛ], 2 [eɪ], 1 [æː]

were considered good or better exemplars of AE [uː] than were NG and PF /u/ except for PF [y] in labial context (see Table IV).

In contrast, NG [øː] and PF [ø] were not assimilated to any one AE vowel on a majority of trials in either context, reflecting both within- and across-listener inconsistencies. Thus, NG [øː] and PF [ø] can be considered uncategorizable back vowels for naïve AE listeners when produced in sentence materials. NG [ʏ] was also uncategorizable in both contexts for most listeners. The majority of listeners were consistent in their categorization of NG [œ] as AE [ʌ] in one or the other context, but only six listeners were consistent in both contexts. For the remaining listeners, this vowel was also uncategorizable in one or both contexts.

2. Front, unrounded, and low vowels

As seen in Table V, NG [iː] and PF [i] were very consistently categorized as their AE transcriptional counterpart in both contexts. NG and PF [ɛ] were only slightly less consistently assimilated to their AE transcriptional counterpart in both contexts. NG [eː, ɪ] and PF [e] showed different group and individual patterns of assimilation across languages and contexts. In labial context, NG [eː] was more consistently categorized than PF [e] as AE [eɪ] overall, but this was not reliable for individual subjects (Fisher's Exact test, n.s.). However, for the ten listeners whose modal response was AE [eɪ] for both NG [eː] and PF [e], goodness ratings were significantly higher for the NG vowel (Wilcoxon Signed-Ranks test, $N=8$, $T+=32.5$, $p<0.03$). In alveolar context, PF [e] was not consistently assimilated to AE [eɪ] by any listener, whereas 14 listeners consistently heard the NG [eː] as a relatively good exemplar of AE [eː] (Fisher's exact test, $p<0.001$). NG [ɪ] was consistently assimilated to AE [ɪ] by more listeners in alveolar than in labial context (Fisher's Exact test, $p<0.04$), reflecting their acoustic differences (see Appendix, Table VI).

For the three low vowels, NG [ɑː, a] and PF [a], the group modal assimilation responses were AE [ɑː], except for NG [a] in alveolar context, which was assimilated primarily to AE [ʌ]. However, analysis of individual listeners' response patterns revealed differences across contexts and languages. NG [a] was consistently categorized by significantly more listeners as AE [ɑː] in labial than in alveolar context (12 vs 1, Fisher's Exact test, $p<0.001$). In labial context, ten listeners assimilated both NG [ɑː, a] to the same AE category with identical or very similar goodness ratings, whereas in alveolar context, only one listener assimilated both vowels as equally good exemplars of AE [ɑː] (Fisher's Exact test, $p<0.01$). Eight listeners assimilated them consistently to AE [ɑː] and [ʌ], respectively. This reflects the fact that NG [a] is fronted and raised in alveolar context (Strange *et al.*, 2007). For other listeners, this contrast constituted a categorized-uncategorized contrast in both contexts.

For PF [a], the overall modal assimilation to AE [ɑː] was lower than for NG [a] in both contexts primarily due to across-listener variability. In labial context, 10 listeners were consistent in categorization responses, but selected different AE vowels; 13 listeners consistently categorized PF [a] as a particular AE vowel in alveolar context. However, only six listeners assimilated PF [a] in the same pattern across labial and alveolar contexts. Note that only one listener perceived PF [a] as most similar to AE [æː] in alveolar context.

C. Discussion

These results demonstrated differences across languages and contexts in naïve AE listeners' perceptual assimilation of

vowels in sentence materials. In some cases, these patterns of perceived similarity could not be predicted from context-specific, cross-language spectral similarity patterns (see the Appendix). In other cases, differences in perceptual similarity patterns were predicted from changes in acoustic similarities across contexts and languages due to cross-language differences in coarticulatory patterns (Strange *et al.*, 2007).

NG and PF front, rounded vowels were assimilated to back AE vowels even when they were spectrally more similar to front AE vowels. Almost all listeners assimilated NG [y:] as a relatively good exemplar of AE [u:] in both consonantal contexts. Two listeners consistently assimilated PF [y] to AE [i:] in labial context, but the remaining listeners assimilated PF [y] to AE [u:], with no significant differences in goodness ratings across NG [y:] and PF [y] in this context. In alveolar context, all listeners heard PF [y] as very similar to fronted allophones of AE [u:]. The remaining front, rounded vowels were, in general, not consistently categorized within or across listeners as any particular back AE vowel in both contexts.

The results for NG front, rounded vowels by and large replicate those reported in Strange *et al.* (2005), for which both NG speakers and the AE dialect group (mostly Midwestern-born living in Florida) differed from the present study. In general, when sentence materials are used, naïve AE listeners perceive front vs back, rounded NG pairs, especially [y:/u:], as very similar to each other and more similar to back than to front AE vowels in both coronal and noncoronal contexts. The results for the PF front, rounded vowels are less easily compared with those reported by Levy (2009) because the latter study included the AE palatalized [^{j}u] response category. However, as a group, naïve AE listeners in that study assimilated PF [y] in labial context to AE [i:] on a small proportion of trials and the proportion of [^{j}u] responses was greater for PF [y] in labial than in alveolar context. Thus, listeners heard PF [y] as less similar to AE [u] in labial than in alveolar context. Levy also reported less consistency in individuals' assimilation patterns for PF [ø] than for PF [y] in labial context. In alveolar context, the two PF vowels were assimilated with the same overall consistency. Thus, Levy (2009) concluded that naïve AE listeners' responses to PF front, rounded vowels produced in sentence materials showed some context-specific patterns of assimilation. In the present study, there was less evidence that listeners heard the extremely front PF [y] as a poorer exemplar of AE [u] in labial than in alveolar context.

Assimilation patterns for back, rounded NG and PF vowels showed that AE listeners were able to access some context-specific phonetic information about cross-language differences in coarticulatory variation for these similar NG and PF vowels. NG [u:] in alveolar context, which was acoustically farthest back (lowest F2 values), was judged a poorer exemplar of the fronted allophone of AE [u:] appropriate in that context. Group data appeared to show that NG [o:] and PF [o] were assimilated to AE [oʊ] somewhat differently across contexts, although the differences were not reliable across individual listeners' data. PF [o] was assimilated to other AE back vowels or uncaegorizable by many listeners.

Perceptual similarity patterns with respect to vowel height contrasts in front, unrounded and back, rounded vowels for these sentence materials differed from those reported in study 1 and showed marked effects due to consonantal context. NG [o:] was heard as very similar to AE [oʊ] in labial context; when coarticulated with alveolar consonants it was categorized as AE [ɔ:] or as in between [o:] and [ɔ:] by many listeners. In contrast, NG [e:] was perceptually more similar to its AE counterpart in alveolar than in labial context. The latter results were not predictable from the spectral similarity patterns derived from discriminant analysis. In general, PF [o, e, ɛ] were perceived as poorer exemplars than the same NG vowels of their AE counterparts in both contexts. This was predictable from their relatively higher locations (lower F1 values) in vowel space than for the NG vowels. The authors would predict therefore that the PF [e/ɛ] contrast would be more difficult to discriminate[7] (cf., Gottfried, 1984) than the same contrast in NG. They are not aware of any studies that examine the perception of this contrast by AE learners of German. Likewise, PF [o/ɔ] might be expected to be more difficult than the same contrast in NG [see Gottfried and Beddor (1988) for French data]. For both contrasts, if AE learners of German could be trained to attend to duration differences, they should be able to distinguish these vowels easily; in French, duration differences are more subtle or nonexistent (Strange *et al.*, 2007).

Patterns of perceptual assimilation of NG [ɑ:, a] and PF [a] showed that AE listeners varied their responses as a function of differences in the phonetic realization of these low vowels across languages and contexts. In labial context, NG [ɑ:, a] were both assimilated AE [ɑ:]; in alveolar context, short NG [a] was raised and fronted enough that most AE listeners perceived it as more similar to AE [ʌ]. AE listeners also assimilated PF [a] differently in labial and alveolar contexts; however, these differences could not be predicted readily from spectral or temporal similarity patterns. In general, PF [a] was a poor perceptual match to any AE low vowel. Because PF includes only one (oral) low vowel, AE learners of French would not be predicted to have difficulty discriminating this vowel from PF [ɛ, ɔ], which are quite high phonetically (low F1 values) relative to [a]. However, based on individual differences in perceptual assimilation patterns, the authors might expect considerable problems in accurate *production* of PF [a] in some contexts.

More generally, a significant finding of the present study was that perceptual assimilation of NG and PF vowels presented in coarticulated nonsense words embedded in carrier sentences often differed from the patterns of assimilation revealed in study 1, in which the vowels were produced and presented in citation-form utterances. These differences are discussed in Sec. IV.

IV. GENERAL DISCUSSION

The data from both studies are summarized here and implications for current models of non-native and L2 speech perception are discussed. In Sec. IV A, the authors characterize perceptual assimilation patterns for German and French front, rounded vowels, which would be considered

"new" vowels in Flege's SLM framework (Flege, 1995; Flege and Hillenbrand, 1984). From comparisons of perceptual similarity of front vs back, rounded vowels, the authors draw conclusions about how these contrasts differ across languages, speaking styles, and phonetic contexts with respect to Best's PAM taxonomy of assimilation patterns (Best, 1995; Best and Tyler, 2007). In Sec. IV B, they characterize perceptual assimilation patterns of other German and French vowels that have transcriptional counterparts in English. These might be considered similar vowels in the SLM framework. Again, differences in assimilation patterns as a function of language, stimulus materials, and phonetic context are characterized.

A. Front vs back, rounded vowels

In replication of earlier cross-language categorization studies of French (Levy, 2009) and German (Polka, 1995; Strange *et al.*, 2004, 2005), front, rounded vowels were generally perceived by naïve AE listeners as more similar to back than to front AE vowels in both citation and sentence materials. As expected, NG and PF back, rounded vowels were also assimilated to back AE categories in all contexts, although goodness ratings varied considerably across particular vowels, languages, citation vs sentence materials, and consonantal contexts.

In study 1, NG and PF back, rounded vowels were generally heard as much better exemplars of AE back vowels than were NG and PF front, rounded vowels. Second, the majority of AE listeners were less consistent in their categorization responses of NG [ʏ, ø:, œ] and PF [y, ø] to particular AE back vowels, suggesting that these vowels were heard as Uncategorizable back vowels in citation-form utterances. Thus, contrasts between front vs back, rounded vowels in NG and PF demonstrated category goodness or uncategorized-categorized patterns of perceptual assimilation for the majority of naïve AE listeners, predictive of intermediate levels of discrimination difficulty (Best, 1995). In study 2, listeners' perceptual assimilation patterns suggested that detailed, context-specific phonetic information was less accessible to listeners when they categorized and judged the goodness of these new vowels. NG [u:/y:] and PF [u/y] were judged as equally good exemplars of AE [u:] by the majority of listeners, especially when the vowels were surrounded by alveolar consonants, despite large differences in acoustic structure (i.e., differences in F2 > 3 barks). Thus, in sentence materials, these contrasts were assimilated in single category or category goodness patterns with very small goodness differences by most listeners. In contrast, NG [o:/ø:] and PF [o/ø] reflected, for the most part, categorized-uncategorized patterns, as they did in study 1.

The results of these two studies suggest that perceptual assimilation patterns derived from studies using citation materials (e.g., Polka, 1995; Best *et al.*, 1996; study 1 here) may significantly underestimate discrimination difficulties involving some NG and PF front, rounded vowels for beginning AE L2 learners in online speech processing situations. This may be especially true for the high vowels, where front vs back, rounded vowels constitute allophonic variations in AE [u:]. Furthermore, AE listeners appeared not to attend to duration differences in categorizing and rating NG [ø:/ʏ]. Context-specific spectral similarity relationships did not predict perceptual assimilation patterns for front, rounded vowels in either language. Thus, the authors conclude that, under stimulus and task conditions approaching continuous speech processing, attunement to language-general, acoustic-phonetic L1/L2 dissimilarities is not possible for most naïve listeners when judging vowels that do not occur as distinct phonological categories in their L1. Rather, naïve listeners must use L1 SPRs (Strange and Shafer, 2008) to categorize these non-native vowel segments. Even when the immediate phonetic context is not changing, listeners seem unable to make consistent judgments about the phonetic appropriateness of new vowels in those contexts. That is, they appear to use an L1 phonological mode of perception and fail to differentiate distinctive L2 vowels that are allophonic variants in L1, despite very large differences in their spectral structure.

B. Height contrasts and low vowels

Almost all listeners heard dissimilarities among NG [i:, e:, ɛ] and among PF [i, e, ɛ] in both studies 1 and 2, despite differences from AE counterparts in relative height of the mid and mid-low vowels in some or all contexts. Thus, predictions based on context-specific spectral similarity patterns that PF [ɛ] would be perceived as similar to AE [e^{ɪ}] or [ɪ] and that [e/i, e/ɛ] would yield single category assimilation patterns (cf. Gottfried, 1984) were not borne out for most AE listeners in this study. Similarly, the mid to mid-low, back vowel pairs in both languages were generally perceived as dissimilar by most AE listeners, except for PF [ɔ/o] in alveolar context (see Gottfried and Beddor, 1988). Again, these patterns were, in general, not predictable on the basis of spectral similarity relationships with AE vowels (see Figs. 1 and 2). Perceptual assimilation patterns in study 2 for these vowels were more similar to those in study 1 than were contrasts involving the front, rounded vowels. In general, pairs of similar vowels in both NG and PF that differed in height (as well as length and/or diphthongization) from their AE counterparts showed two-category or categorized-uncategorized patterns. This suggests that AE listeners were attuned to small height differences (F1 values), as would be expected given that AE contrasts five vowel heights. Thus, even though these vowels (except NG [ɛ]) differed from their AE counterparts in relative locations in vowel space, AE listeners were apparently able to differentiate them after they had a bit of practice with each speaker's complete vowel inventory (i.e., block 1 of the tests in which all three tokens of each vowel were presented once). That is, they appeared to be able to adjust their perceptual boundaries for similar vowels rather rapidly within the context of the experiment. Such rapid adjustment is useful in adapting to accented versions of English spoken by native speakers (Bradlow and Bent, 2008).

Finally, perceptual assimilation patterns for NG [ɑ:/a] and PF [a] indicate that these vowels were perceived as similar to AE [ɑ:] by more listeners in non-coronal contexts (study 1 and labial condition of study 2) than in coronal consonants (study 2, alveolar condition). The greater differ-

entiation in alveolar context reflects the fact that NG and PF [a] are raised (and more spectrally distant from NG [ɑ:]) in this context. In both studies 1 and 2, most AE listeners' performance indicated that the large duration differences between the NG vowels did not affect perceptual assimilation patterns. Thus, NG [ɑ:/a]) might be difficult for most AE L2 learners of German to discriminate especially in non-coronal contexts where they overlap almost completely in formant structure, unless the learners have been explicitly taught to attend to duration differences.

To conclude, similarities and differences in the patterns of perceptual assimilation of non-native vowels by naïve listeners across languages, prosodic contexts, and consonantal contexts lead to the following generalizations:

(1) Perceptual assimilation tests of non-native vowels that have no phonologically distinctive counterparts in the native language (new vowels) often show different patterns of perceived L1/L2 similarity from those predicted from context-specific comparisons of their spectral and temporal properties. Patterns of assimilation are better predicted on the basis of cross-language differences in the systematic allophonic characteristics of vowel categories across languages. In the case demonstrated here, the authors conclude that front, rounded vowels are assimilated to AE back vowels because AE back vowels include highly fronted allophones. That is, that portion of "vowel space" occupied by (contrastive) front, rounded vowels in NG and PF has been subsumed by phonologically back, rounded vowels in AE.
(2) Direct tests of perceptual similarity may be the best predictors of discrimination problems by L2 learners if details about systematic allophonic variations in native and non-native phonological categories are unknown. However, perceptual assimilation tests using citation-form utterances may not accurately predict beginning L2 listeners' discrimination difficulties when listening to continuous speech utterances. Thus, to be maximally generalizable, tests of cross-language similarity of vowels might better be performed using materials in which vowels are produced in multiple consonantal contexts in phrase-length utterances. In addition, individual listeners often show markedly different patterns of perceived similarity; thus, it would be valuable to examine individual L2 learners' perceptual similarity patterns in order to make better predictions about their learning difficulties and to structure individualized training materials for them.
(3) Variations in perceptual assimilation patterns across languages, contexts, and prosodic conditions allow the authors to infer that listeners' knowledge of native categories includes both language-specific phonetic detail related to systematic allophonic variation and context-independent similarity relationships, traditionally characterized by a phonemic level of linguistic analysis. In the automatic selective perception model (Strange and Shafer, 2008), it is hypothesized that there are two "modes" of perception—a phonetic mode and a phonological mode. Perceptual responses may reflect either or both modes, depending upon experimental variables such as stimulus complexity and task demands (cf., Werker and Curtin, 2005).
(4) In perceptual assimilation tests using phrase-length materials, naïve listeners may not have access to detailed phonetic information when categorizing new non-native vowels. However, they do appear to be attuned to small phonetic differences in the phonetic realization of similar non-native vowels. The authors infer that L1 SPRs for perceiving native phonological categories are subject to rapid and temporary readjustment, as when native listeners are listening to non-native speakers' utterances.

ACKNOWLEDGMENTS

The authors thank Robert Lehnhoff Jr., for testing subjects and analyzing data, James J. Jenkins and Kikuyo Ito for comments on a draft of the manuscript, and Natalia Martínez and Diana Posadas for help in checking the references. These data were reported in part in Strange (2007b) and at meetings of the Acoustical Society of America. This research was supported by a grant to the first author (NIH DC-00323); support while writing this manuscript was from NSF.

APPENDIX

See Table VI.

TABLE VI. Context-specific discriminant analyses of NG and PF vowels tested against AE categories (F1, F2, and F3 in barks).

	German vowels		
	hVbə	gəbVpə	gədVtə
iː	i 2, e 1	i 2, e 1	i 3
eː	e 3	e 3	e 2, ɪ 1
ɪ	ɪ 3	ɪ 2, e 1	ɪ 3
ɛ	ɛ 3	ɛ 3	ɛ 3
yː	ɪ 3	ɪ 3	u 3
øː	ɪ 3	ɪ 2, ʊ 1	u 2, ʊ 1
ʏ	ɪ 2, ʊ 1	ɪ 3	ʊ 1, u 1, ɪ 1
œ	ɛ 3	ʌ 1, o 1, ɛ 1	ʌ 2, ɛ 1
uː	u 2, o 1	o 3	ɔ 2, o1
oː	o 3	o 3	ɔ 3
ʊ	o 3	o 3	o 3
ɔ	ɔ 3	ɔ 2, ɑ 1	ɔ 2, o1
ɑː	ɑ 3	ɑ 3	ɑ 3
a	ʌ 2, ɑ 1	ɑ 3	ɑ 2, ʌ 1
	French vowels		
	Vb(ə)	rab Vp(ə)	radVt(ə)
i	i 3	i 3	i 3
e	i 2, e 1	i 3	i 3
ɛ	e 3	ɪ 2, e 1	e 2, ɛ 1
y	e 3	i 2, e 1	i 3
ø	ɪ 3	ɪ 2, ɛ 1	ɪ 2, ɛ 1
u	u 3	u 3	o 3
o	o 3	u 2, ɔ 1	ɔ 2, o 1
ɔ	ʌ 2, ʊ 1	ʌ 2, ʊ 1	ʌ 3
a	æ 3	æ 3	æ 3

[1]Medians were used as the measure of central tendency because the authors assumed that goodness ratings constituted only an ordinal level of measurement (see Strange, 2007a).

[2]This minimal criterion requires that each token of a vowel category must be categorized as the same AE vowel at least once (3 tokens × 3 repetitions). In many cases, all three vowels were categorized as the same AE vowel on all repetitions.

[3]Nonparametric statistics (Sign tests and Wilcoxon Signed-Ranks tests) are the appropriate statistics for repeated measures comparisons of ordinal data.

[4]In Strange *et al.* (2004), most AE listeners were not from the Northeast, and some did not differentiate [ɑː/ɔː] in their own speech.

[5]Levy and Strange (2008) and Levy (2009) used [œ] to represent the mid, front, rounded PF vowel represented here as [ø]. PF [œ] and [ø] are typically considered allophones in French.

[6]None of the NG and PF nonsense utterances were lexical items in English. Thus, lexical effects that could have confounded earlier results using CVC utterances were minimized here.

[7]The /e, ɛ/ contrast is phonemically contrastive in French for many lexical pairs and, in some dialects, signifies a grammatical marker, differentiating the conditional and future tenses (first person, singular). Failure of AE listeners to discriminate these French vowels may thus lead to difficulty in perceiving the correct tense of a verb.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., Faber, A., and Levitt, A. (**1996**). "Assimilation of non-native vowel contrasts to the American English vowel system," J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., Hallé, P. A., Bohn, O.-S., and Faber, A. (**2003**). "Cross-language perception of non-native vowels: Phonological and phonetic effects of listeners' native languages," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Sale, D. Rescasens, and J. Romero (Causal Productions, Barcelona, Spain), pp. 2889–2892.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 13–34.

Bradlow, A. R., and Bent, T. (**2008**). "Perceptual adaptation to non-native speech," Cognition **106**, 707–729.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 233–277.

Flege, J. E., and Hillenbrand, J. (**1984**). "Limits on phonetic accuracy in foreign language speech production," J. Acoust. Soc. Am. **76**, 708–721.

Flege, J. E., Munro, M. J., and Fox, R. A. (**1994**). "Auditory and categorical effects on cross-language vowel perception," J. Acoust. Soc. Am. **95**, 3623–3641.

Gottfried, T. L. (**1984**). "Effects of consonant context on the perception of French vowels," J. Phonetics **12**, 91–114.

Gottfried, T. L., and Beddor, P. S. (**1988**). "Perception of temporal and spectral information in French vowels," Lang Speech **31**, 57–75.

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

Levy, E. S. (**2009**). "Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French," J. Acoust. Soc. Am. **125**, 1138–1152.

Levy, E. S., and Strange, W. (**2008**). "Perception of French vowels by American English adults with and without French language experience," J. Phonetics **36**, 141–157.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Rochet, B. L. (**1995**). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Timonium, MD), pp. 379–410.

Strange, W. (**2006**). "Second-language speech perception: The modification of automatic selective perceptual routines," J. Acoust. Soc. Am. **120**, 3137.

Strange, W. (**2007a**). "Selective perception, perceptual modes, and automaticity in first- and second-language processing," J. Acoust. Soc. Am. **122**, 2970.

Strange, W. (**2007b**). "Cross-language phonetic similarity of vowels: Theoretical and methodological issues," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, edited by O.-S. Bohn and M. J. Munro (John Benjamins, Amsterdam), pp. 35–55.

Strange, W. (**2009**). "Automatic selective perception (ASP) of first language and second language speech: A working mode," J. Acoust. Soc. Am. **125**, 2769.

Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (**2005**). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.

Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (**2004**). "Acoustic and perceptual similarity of North German and American English vowels," J. Acoust. Soc. Am. **115**, 1791–1807.

Strange, W., and Shafer, V. L. (**2008**). "Speech perception in second language learners: The re-education of selective perception," in *Phonology and Second Language Acquisition*, edited by J. G. Hansen Edwards and M. L. Zampini (John Benjamins, Amsterdam), pp. 153–191.

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (**2007**). "Acoustic variability within and across German, French and American English vowels: Phonetic context effects," J. Acoust. Soc. Am. **122**, 1111–1129.

Werker, J. F., and Curtin, S. (**2005**). "PRIMIR: A developmental framework of infant speech processing," Lang. Learn. Dev. **1**, 197–234.

Immediate and long-term effects of hearing loss on the speech perception of children

Andrea Pittman,[a] Kendell Vincent, and Leah Carter
Arizona State University, P.O. Box 870102, Tempe, Arizona 85287-0102

(Received 8 August 2008; revised 15 June 2009; accepted 16 June 2009)

The purpose of the present study was to examine the immediate and long-term effects of hearing loss on the speech perception of children. Hearing loss was simulated in normally-hearing children and their performance was compared to that of children with hearing loss (long-term effects) as well as to their own performance in quiet (immediate effects). Eleven children with normal hearing (7–10 years) were matched to five children with mild to moderate sensorineural hearing loss (8–10 years). Frequency-shaped broadband noise was used to elevate the hearing thresholds of the children with normal hearing to those of their matched hearing-impaired peer. Meaningful and nonsense sentences were presented at five levels and quantified using an audibility index (AI). Comparison of the AI functions calculated for each group and listening condition revealed a significant, immediate effect of elevated hearing thresholds in the children with normal hearing but no long-term effects of hearing loss. The results of this study suggest that hearing loss affects speech perception adversely and that amplification does not fully compensate for those effects. However, the data suggest that over the long term children may develop compensatory strategies to reduce the effects of hearing loss. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3177265]

PACS number(s): 43.71.Ky, 43.71.Es, 43.71.Ft, 43.71.Gv [RSN] Pages: 1477–1485

I. INTRODUCTION

Children's ability to perceive speech begins early in life and matures throughout adolescence (Elliott, 1979; Holt and Carney, 2007; Stelmachowicz *et al.*, 2000). The maturation of speech perception proceeds over a long period of time suggesting that it is an advanced skill. Research showing the earlier maturation of auditory skills that form the foundation for speech perception (e.g., intensity discrimination, temporal resolution, and localization) supports this notion (Buss *et al.*, 2009; Allen and Wightman, 1994; Stuart, 2005; Wightman *et al.*, 1989; Van Deun *et al.*, 2009). Additionally, the results of other studies suggest that auditory skills are reorganized periodically throughout development to form the mature, efficient, and robust perceptual processes needed to perceive speech (Boothroyd, 1997; Gershkoff-Stowe and Smith, 1997; Nittrouer and Miller, 1997).

Less is known about the development of speech perception in the presence of hearing loss (HL). Fundamental to an understanding of speech perception in children with HL is how they differ from adults with HL. Adults generally acquire HL later in life after communication skills are well established. Children do not have the same advantage. Instead, children with HL struggle to *learn* to communicate while adults struggle to *continue* to communicate. This places children at risk for delayed or impaired speech perception development. Jerger (2007) summarized well our current state of knowledge regarding the development of speech perception in children with HL and concluded that, in general, basic auditory skills develop more normally than the advanced skills necessary for linguistic processing. Also, children with more severe hearing impairment demonstrate greater perceptual difficulties than children with milder degrees of HL. Later, Jerger *et al.* (2009) provided additional evidence that children with HL process auditory stimuli differently than children with normal hearing (NH). They examined the development of phonological processing and found that the presence of HL predisposes young children to represent speech more in visual rather than in auditory forms compared to children with NH. Although these visual representations eventually take on more appropriate auditory forms, there is evidence that auditory representations may be irrevocably affected by HL in childhood and that those effects may be carried into adulthood (Pittman, 2008).

To overcome the effects of elevated hearing thresholds, amplification is routinely provided to children with HL (e.g., hearing aids) so that they may receive a signal of sufficient audibility. With amplification, few differences between groups are observed when listening in quiet (Stelmachowicz *et al.*, 2000) suggesting that speech perception is comparable at suprathreshold levels. However, significant effects of HL re-emerge when speech is presented in noise (i.e., background competitor) suggesting that elevated hearing thresholds alone do not fully account for the effects of HL. Hicks and Tharpe (2002) examined word recognition in children with NH compared to children with HL who wore their personal hearing aids. The stimuli were presented in quiet and at several signal-to-noise ratios. They found small but significant differences in performance between the groups (mean score ~85% HL and 92% NH) in the most difficult (signal-to-noise ratio +10 dB) as well as greater variability in performance (standard deviation ~10% HL and 5% NH). Although amplification was provided, the poorer performance of the children with HL suggests that factors in addition to elevated thresholds affected their performance.

[a]Author to whom correspondence should be addressed. Electronic mail: andrea.pittman@asu.edu

One likely factor may be the varying degrees of audibility provided by children's personal hearing aids. Scollie (2008) carefully examined speech perception in noise as a function of audibility for children with HL compared to children with NH. The purpose of her study was, in part, to determine whether or not the performance of children with HL could be predicted on the basis of the audibility of the stimuli. Transfer functions were derived for the children with NH that accounted for 94.6% of the variance in the data. Unfortunately, the same transfer function accounted for only 87% of the variance in the data for the children with HL. The results indicate that speech perception in children with HL cannot be predicted solely on the basis of age and stimulus audibility.

The results of these studies suggest that the full effect of HL is not captured by estimates of hearing threshold. This would be consistent with studies demonstrating the suprathreshold psychophysical deficits that accompany mild to moderate HL (e.g., poor frequency selectivity, poor temporal resolution, and loudness recruitment) (Fabry and Van Tasell, 1986; Humes *et al.*, 1987; Needleman and Crandell, 1995; Dubno and Schaefer, 1992, 1995; Dubno *et al.*, 2000). Suprathreshold deficits would also explain the difficulty that hearing-impaired children (and adults) experience in noise.

Another possibility is that the presence of HL may delay or impair the development of speech perception. That is, HL, which can occur at any time during childhood, may interact adversely with the development of speech perception. It can be argued that any examination of speech perception in children with HL would be confounded by the interaction between HL and the development of speech perception. However, it is likely that the effects of suprathreshold deficits and developmental factors co-occur.

The purpose of the present study was to examine the effects of elevated hearing thresholds on the speech perception of children with NH. That is, HL was simulated in children with NH so that the immediate and long-term effects of HL could be determined. Immediate effects were examined by comparing speech perception under conditions of NH and simulated HL. Long-term effects were examined by comparing the performance of children with simulated HL to that of children with actual HL. To simulate HL, hearing thresholds were elevated using a noise masker. This form of simulated HL has been used widely to examine the effects of HL in adults (Fabry and Van Tasell, 1986; Humes *et al.*, 1987; Needleman and Crandell, 1995; Dubno and Schaefer, 1992, 1995; Dubno *et al.*, 2000). It is considered to be the most valid approach to simulating sensorineural HL because the effects of masking are localized at the level of the cochlea and approximate well the frequency selectivity and loudness recruitment experienced in the impaired ear (Dubno and Schaefer, 1992; Humes *et al.*, 1988). Dubno and co-workers used noise masking successfully to examine the effects of HL in adults (Dubno and Ahlstrom, 1995; Dubno and Schaefer, 1992, 1995; Dubno *et al.*, 2000). To date, no studies of simulated sensorineural HL have been conducted in children.

A two-step approach was used. First, the speech perception of children with NH was examined in quiet and in a condition of simulated HL. The stimuli were frequency shaped relative to the children's quiet and masked thresholds to provide equivalent sensation in both listening conditions. Performance in each condition was compared to determine whether or not speech perception was significantly poorer when the hearing thresholds were elevated even though similar audibility was provided. Second, the speech perception of children with HL was measured with the same frequency-shaped stimuli presented to the children with NH in the simulated HL condition. The results were compared to determine whether or not the speech perception of children with HL is effected by factors in addition to elevated hearing thresholds. It was hypothesized that (1) speech perception would be poorer in the presence of elevated hearing thresholds and (2) children with HL would demonstrate poorer speech perception than children with simulated HL.

Finally, for both the quiet and masked listening conditions, performance measures for meaningful and nonsense sentences were obtained to capture the effects of familiar and unfamiliar communication contexts. Research has shown that the psychometric functions relating speech intelligibility to perception in young children and adults differ for materials having high- and low-predictability (Dubno *et al.*, 2000; Dirks *et al.*, 1986). That is, children require higher presentations levels for speech materials that are less familiar to them. Also, the linguistic development of children with mild to moderate HL is, on average, 2 years behind that of children with NH (Pittman *et al.*, 2005) and the delay is greater for children with more severe losses (Blamey *et al.*, 2001). It was anticipated that the effects of HL would be more apparent for nonsense sentences than for meaningful sentences. If so, the hypothesis that long-term HL negatively impacts the development of speech perception would be supported.

II. METHOD

A. Participants

Participants were 5 children with HL between the ages of 8 and 10 years with mild to moderate sensorineural HLs and 11 children with NH between the ages of 7 and 10 years. The HLs were known to be sensorineural by history, and thresholds were confirmed on the day of testing. Prior to testing, otoscopy and tympanometry were performed to confirm normal middle-ear function. Vocabulary age (VA) was determined using the Peabody Picture Vocabulary Test (PPVT), Form IIIB (Dunn and Dunn, 2006). Children with NH were matched to each child with HL based on their chronological age (CA). With the exception of two children with NH, each child was within 1 year of the CA of the child with HL. Differences in VA ranged from 1 month to 5 years, 2 months. Table I lists the gender, CA, standardized PPVT score, VA, and hearing thresholds of each child. The first three children with HL listed in the table had flat HL configurations whereas the remaining two children had sloping, high-frequency HLs.

Pure-tone thresholds [in decibel sound pressure level (SPL)] were obtained at octave frequencies from 0.25 to 8 kHz in the right ear only. The children were instructed to push a button (computer mouse, secured to a table) when

TABLE I. ID, group, gender, CA, standardized PPVT score, VA, and hearing thresholds for the children with HL and masked thresholds for the children with NH. Average minimum audibility levels for normal-hearing young adults (NHAs) are also provided.

ID	Group	Gender	CA (y:m)	PPVT Std.	VA (y:m)	Hearing thresholds (kHz) 0.25	0.5	1	2	4	8 (dB SPL)	rms error (dB)
HL1	HL	M	10:3	80	7:8	61	59	54	54	64	41	
NH1A	NH	M	7:4	107	8:0	62	58	53	58	70	38	4.3
NH1B	NH	M	10:3	103	10:10	66	56	53	50	58	36	2.4
NH1C	NH	F	10:4	113	12:6	64	61	55	55	62	37	3.3
HL2	HL	F	10:12	101	11:0	63	64	70	76	50	56	
NH2A	NH	M	10:7	94	9:8	62	68	75	71	63	56	6.3
NH2B	NH	F	10:7	106	11:6	66	62	68	79	52	55	2.3
HL3	HL	F	8:11	82	6:10	68	64	71	61	58	58	
NH3A	NH	M	8:10	121	12:00	68	64	72	61	59	59	0.7
HL4	HL	F	8:2	115	9:11	30	20	21	51	52	79	
NH4A	NH	F	7:11	107	8:7	35	29	28	51	48	83	5.6
NH4B	NH	M	8:4	108	9:4	29	27	26	54	60	79	4.9
NH4C	NH	M	10:7	100	10:2	24	22	23	48	55	79	3.2
HL5	HL	M	8:11	114	10:10	49	49	46	73	74	84	
NH5A	NH	F	8:4	116	10:6	44	49	53	70	71	86	4.0
NH5B	NH	M	9:6	108	10:9	47	50	52	59	75	91	6.9
NHAs (minimum audibility levels)						25	17	11	12	11	18	

they heard a beep. The signal duration was 1000 ms, including 20 ms rise-fall ramps. Threshold estimation was accomplished by a single interval, adaptive procedure using a stepping rule that approximated the 70.7% point on a psychometric function (Levitt, 1971). The initial step size was 20 dB until the first reversal, followed by a step size of 4 dB for 4 reversals and then 2 dB for 5 reversals. Initial signal levels were above the hearing thresholds of both the children with NH and with HL. No feedback was provided for correct or incorrect responses. The arithmetic average of the levels corresponding to the final five reversals was taken as the threshold estimate. Thresholds were repeated and averaged.

Elevated thresholds were simulated in the children with NH using a broadband noise that was generated and filtered digitally using ADOBE AUDITION (v1.5). These children will be referred to as the noise-masked NH children. The initial octave-band levels of the noise were calculated using critical ratio predictions (Hawkins and Stevens, 1950). The spectrum level of the noise was then adjusted manually (usually in 1–3 dB steps) until the masked thresholds were within ±4 dB of the target thresholds. At least two thresholds were obtained at each frequency and averaged.

Figure 1 displays the pure-tone thresholds for the five children with HL (open circles) and for the noise-masked NH children (filled circles). The thresholds of all five children with HL are shown in the lower right panel. The shaded area represents the range of thresholds obtained from 11 young adults with NH using the same equipment and procedures. The data in Fig. 1 indicate that the children with HL had mild to moderate sloping or flat HLs that were approximately 20–70 dB above normal. Also, the thresholds of the noise-masked NH children were in good agreement with the target thresholds of the children with HL. To evaluate the match

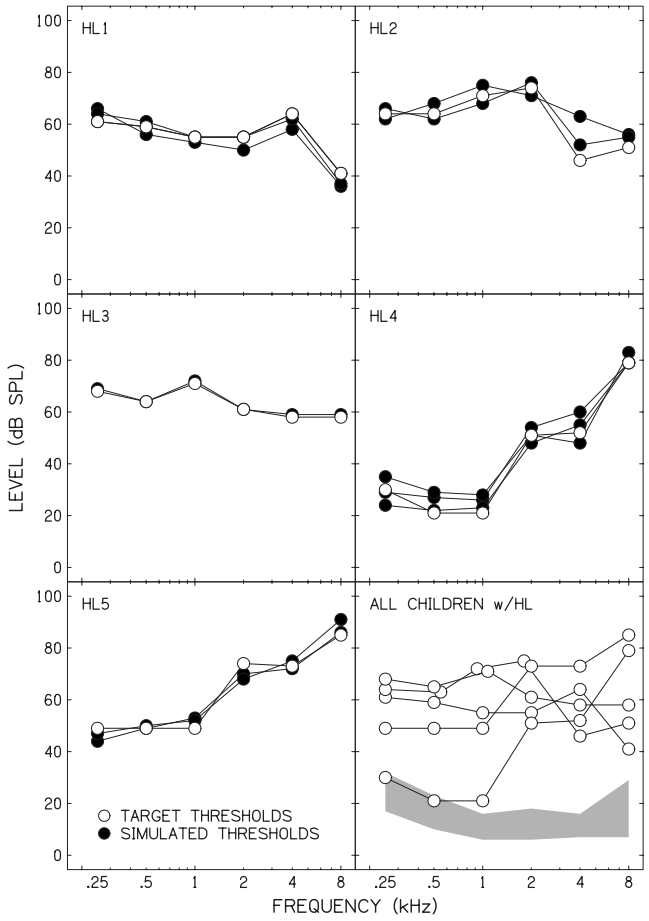


FIG. 1. Puretone thresholds for the five children with HL (open circles) and for the noise-masked NH children (filled circles). The thresholds of all five children with HL are displayed in the lower right panel. The shaded area represents the range of minimum audibility in NH adults for the equipment and procedures used.

between the target and obtained thresholds, the rms error was calculated for each noise-masked NH child. The individual rms errors varied between 0.7 and 6.9 dB indicating a close match between the target thresholds and those achieved through masking. These results are comparable to the rms error of four adults with similar HL configurations (2.2–6.1 dB) reported by Humes *et al.* (1987). The children's performance indicates that they were able to respond as reliably as adults in an abbreviated masking procedure. The thresholds obtained for the noise-masked NH children and the children with HL are provided in Table I, as well as the reference levels for NH adults.

B. Stimuli

The stimuli were two lists of meaningful sentences and two lists of nonsense sentences. Each sentence was comprised of 4 words and each list contained 30 sentences. The same words were used to construct both types of sentences. The meaningful sentences were grammatically and semantically correct (e.g., "Tough guys sound mean") whereas the nonsense sentences were grammatically correct, but semantically anomalous (e.g., "Blocks can't run sharp"). The sentences were generated originally for a study by Boothroyd and Nittrouer (1988) and then supplemented by Stelmachowicz *et al.* (2000). The sentences were rerecorded for this study using a female talker having a standard American dialect. The stimuli were digitally recorded at a sampling rate of 22.05 kHz using a microphone with a flat frequency response to 10 kHz (AKG, C535 EB). Individual sentences were extracted from the original recording, equated for rms level, and saved in separate files using ADOBE AUDITION (v1.5).

The sentences were first frequency shaped to accommodate the elevated thresholds of each child with HL and his/her noise-masked NH counterpart. DSL V5.0A fitting parameters provided approximate targets for average conversational speech (Scollie *et al.*, 2005). The targets were derived using age-appropriate real-ear-to-coupler differences and a speech weighted input level of conversational speech (65 dB SPL) and expressed in dB SPL relative to measures in a 2-cm^3 coupler. Because the stimuli were presented via supra-aural rather than insert earphones, target levels were approximated my measuring the 1/3-octave band levels developed in a 6-cm^3 coupler. The highest presentation level was 5 dB above the target levels with the 4 remaining presentation levels decreasing in 5 dB steps. These presentation levels were chosen to provide a range of sensation levels that would result in scores above floor and below ceiling values.

Figure 2 shows the hearing thresholds (filled circles) and the equivalent internal noise levels (solid line) of a child with HL in the upper panel. The middle and lower panels show the masked and quiet thresholds, respectively, for a child with NH. The solid line in the middle panel indicates the level of the masking noise used to elevate threshold. The dashed lines in each panel show the long-term average of the frequency-shaped sentences in 1/3-octave bands. Note that the sensation level of the stimuli was similar across frequency for both the child with HL and the noise-masked NH child (upper panels). Likewise, sensation level was similar

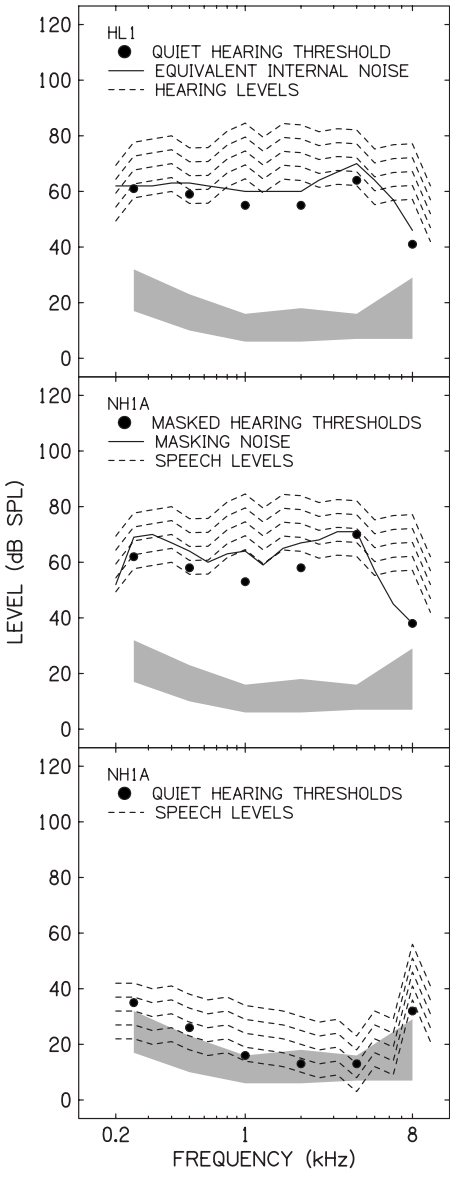


FIG. 2. Quiet pure-tone thresholds (filled circles) are shown for a child with HL in the upper panel. Masked and quiet pure-tone thresholds for a child with NH are shown in the middle and lower panels, respectively. The long-term average of the frequency-shaped sentences is shown in 1/3-octave bands at each of five presentation levels (dashed lines). The solid line represents the long-term average of the masking noise for the child with NH and the equivalent internal noise level for the child with HL. The shaded area represents the range of minimum audibility for a group of NH young adults.

for the child with NH in the masked and quiet listening conditions (lower panels). Sensation level was carefully controlled to reduce variability in performance that may result from variations in amplitude across frequency.

Sensation levels were calculated separately for the quiet and noise listening conditions. For the quiet condition, an estimate of internal noise level was calculated for each band by subtracting the normal critical ratio (Hawkins and Stevens, 1950) from the hearing level in SPL and then adding the bandwidth (in decibel) within each 1/3-octave band to that value (Sherbecoe and Studebaker, 2002; Studebaker *et al.*, 1993). This procedure produced internal noise levels that were within 7 dB of the hearing thresholds. For the noise

condition, the broadband noise used to elevate the thresholds as well as the speech stimuli were recorded in a 6-cm^3 coupler. rms amplitude was calculated in 1/3-octave bands. The ratio of speech-to-noise in each band was then determined and used to calculate the audibility of the speech stimuli.

To calculate audibility, the frequency-shaped sentences were concatenated into a single file and recorded in a 6-cm^3 coupler. Rms amplitude was measured in each of 18 1/3-octave frequency bands between 0.2 and 8 kHz using a 40 ms Hanning window with 50% overlap. To approximate the peaks of speech, 15 dB was added to the rms level in each 1/3-octave band. An audibility index (AI) was then calculated using the following formula:

$$\mathrm{AI} = \frac{1}{30}\sum_{i=1}^{18}[(\mathrm{SNR}_i + 15)W_i]\mathrm{LDF}_i,$$

where i is the number of the 1/3-octave band and SNR is the speech-to-noise ratio for the ith band. The result was multiplied by the importance value assigned to each band (W_i) and summed. As in previous studies involving these materials (Stelmachowicz *et al.*, 2000), the importance function for short passages was used (American National Standards Institute, 1997). Finally, a level distortion factor (LDF_i) was included in the calculation to accommodate for distortion that may occur at elevated presentation levels. The standard speech levels (Ui) provided in the ANSI standard were used.

C. Procedure

All testing was conducted in a sound-treated booth meeting ANSI standards for ambient noise (ANSI, 1999). All stimuli were processed using custom laboratory software designed for use with children and controlled with a standard desktop PC. The stimuli were presented monaurally under earphones having a flat frequency response through 10 kHz (Sennheiser, 25D). The laboratory equipment was calibrated prior to data collection by adjusting the overall output of the transducer for a 1 kHz pure tone in a 6-cm^3 hard-walled coupler to 125 dB SPL and then documenting the voltage at the earphone. Calibration was confirmed prior to testing each child.

For the sentence perception task, each child was instructed as follows: "You will hear a woman say a sentence. Some of the sentences will be normal and some will be silly. Listen to each sentence and repeat as much of it as you can. It's ok to guess. If you don't know, just say so." 6 sentences (24 words) were presented at each of the 5 presentation levels and 2 listening conditions (quiet and masked). The presentation levels proceeded from highest to lowest for all children. The order in which the sentence types were presented (meaningful and nonsense) was counterbalanced across children. Although the timing of the experiment was controlled by the laboratory software, the child's responses were self-paced. Specifically, after a sentence was presented, the child was allowed to respond at his/her own pace. The maximum response window was 15 s. As soon as the child responds the examiner entered the response on a computer monitor, which then prompted the presentation of the next sentence after a 1000 ms delay. The examiner entered the number of words (0–4) that the child was able to repeat correctly. The order of the words provided by the child or the addition of extra words was not considered during scoring. A second examiner was positioned outside the sound-treated room to administer the experiment and to record the child's responses in written form. The scores from the two examiners were compared and averaged. The mean difference between examiners for scores across all listening conditions (24 words) was 0.6 words with a range of 0–7 words. The median difference between examiners was 0 words.

Although the monitor microphone on the audiometer was used by the second examiner to hear the child's responses, a professional recording microphone was placed in front of the child with instructions to speak into it so that the examiner outside the booth could hear what was said. Because children typically respond well to microphones, there was little or no issue with the clarity of their speech. In the event the child's response was unclear, either examiner could request that the child repeat his/her answer. No feedback regarding the accuracy of the responses was provided.

III. RESULTS

Speech perception was examined as a function of stimulus audibility. Transfer functions were generated for each group, listening condition, and sentence type. That is, proportion correct (PC) was estimated using the following formula:

$$\mathrm{PC} = 1 - 10^{(-\mathrm{AI}+k)/S},$$

where k and S are constants determined in a least-squares fit procedure for each data set. The constant k determines the vertical position of the function along the ordinate whereas the constant S is responsible for the shape of the function. Differences between the transfer functions derived for each group and listening condition were determined using a procedure described in Stelmachowicz *et al.*, 2000. In this procedure, functions that are significantly different from one another are best described by two separate functions rather than a single function for the combined data. This analysis captured the variability in performance within and across groups. The procedure requires the calculation of residuals derived from the actual data points and the performance predicted by the transfer function (see Stelmachowicz *et al.*, 2000 for a step-by-step procedure).

Figure 3 shows the performance for the children with NH as a function of AI for the quiet (open symbols) and noise-masked (filled symbols) conditions. The meaningful and nonsense sentences are shown in the upper and lower panels, respectively. The correlations between predicted and observed scores were 0.62 and 0.58 in quiet for the meaningful and nonsense sentences, respectively. Correlations for the noise-masked condition were 0.71 and 0.65 for the meaningful and nonsense sentences, respectively. The predictive accuracy of the transfer functions improved with masking noise, more for the meaningful sentences than for the nonsense sentences. These correlations are similar to those reported in Stelmachowicz *et al.*, 2000 for their NH 8- to 10-year-old children listening in quiet. However, they reported

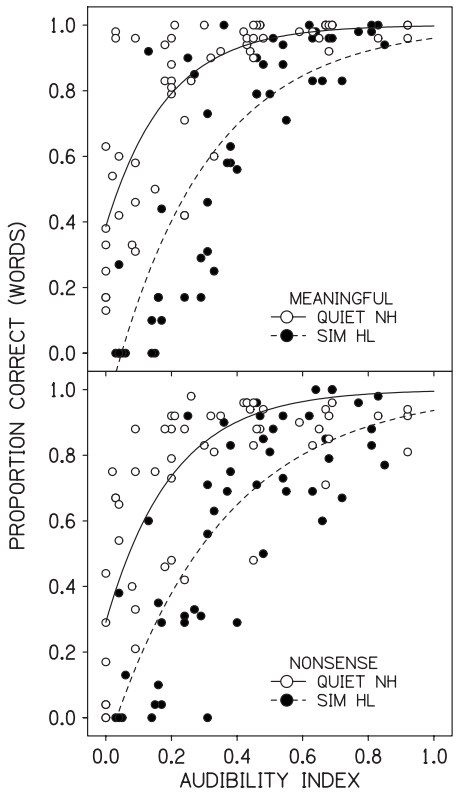


FIG. 3. Speech perception (in PC) as a function of audibility (AI) for the meaningful sentences in the upper panel and for the nonsense sentences in the lower panel. The open symbols are the data for the children with NH listening in quiet (quiet NH). The filled symbols are for the same children in the simulated HL condition (Sim HL). The dashed and solid lines are AI transfer functions for each listening condition.

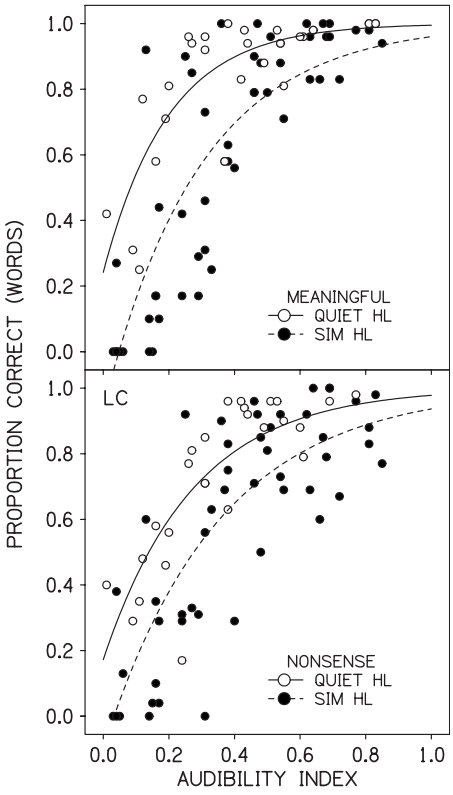


FIG. 4. Same convention as in Fig. 3 but the open symbols are the data for the children with HL (quiet HL) and the filled symbols are for the children with NH in the simulated HL condition (Sim HL). The dashed and solid lines are AI transfer functions for each group.

higher correlations for the nonsense sentences for all but their youngest age group (5 year olds). Figure 4 shows the performance of the children with HL (open symbols) and the noise-masked NH children (filled symbols). The meaningful and nonsense sentences are shown in the upper and lower panels, respectively. Correlations between predicted and observed scores for the children with HL were similar to those of the noise-masked NH children for the meaningful and nonsense sentences (0.65 and 0.67, respectively).

For comparison, the transfer functions for each group and condition are shown together in Fig. 5. The immediate effects of HL on the perception of speech were determined by comparing the performance of the children with NH in quiet (quiet NH) to that of the children with simulated hearing loss (Sim HL). No significant difference was found for the meaningful sentences [$F(81,79)=1.2598$, $p=0.1522$]; however, a significant difference was observed for the nonsense sentences [$F(92,90)=2.0753$, $p=0.0003$]. That means that similar variability in performance was observed in both listening conditions for the meaningful sentences whereas performance for the nonsense sentences was better described by two separate transfer functions. These results indicate that the immediate effects of HL are more apparent for unfamiliar materials than for familiar ones. Because the development of speech perception is a process of decoding unfamiliar utterances, these results suggest that HL may significantly impede that development.

The long-term effects of HL on the perception of speech were determined by comparing the performance of the children with HL (quiet HL) to that of the noise-masked NH children (Sim HL). No significant difference was observed for the meaningful [$F(59,57)=1.3756$, $p=0.1144$] or nonsense sentences [$F(64,62)=1.1832$, $p=0.2537$]. It is interesting to note that the performance of the children with HL fell in between that of the NH children in the quiet and simulated HL conditions. In a final analysis, the performance of the children with HL (quiet HL) was compared to that of the children with NH in the quiet listening condition (quiet NH) for the nonsense sentences only. The nonsense sentences were chosen because a significant difference was observed between listening conditions for the children with NH. No significant difference was observed between the children with HL and the children with NH [$F(75,64)=0.7368$, $p=0.8845$]. These results suggest that children with long-standing HL may develop compensatory strategies that reduce the deleterious effects of HL on speech perception.

IV. DISCUSSION

Recall that the purpose of the present study was to examine the immediate and long-term effects of HL on the speech perception of children. To examine the immediate effects of HL, perception of meaningful and nonsense sentences was examined in NH children under conditions of simulated HL. It was hypothesized that their performance would be poorer in the simulated HL condition compared to quiet despite equivalent audibility in both conditions. Sig-

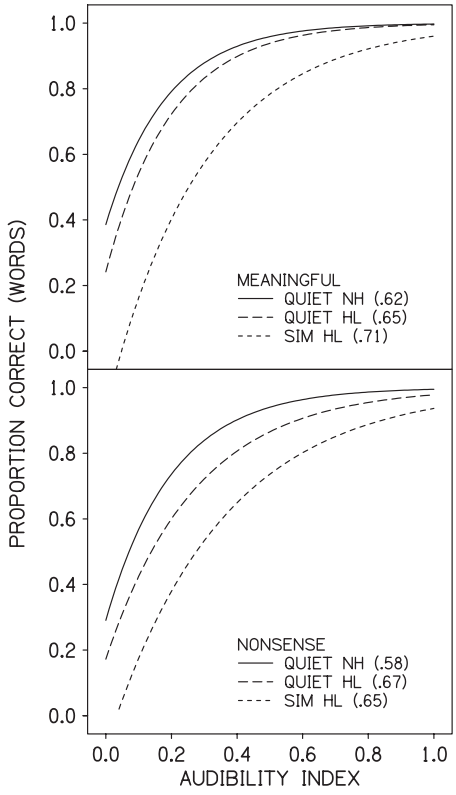


FIG. 5. Transfer functions showing the relation between performance and audibility for the meaningful sentences in the upper panel and for the nonsense sentences in the lower panel. Functions for the NH children (quiet NH—solid line), hearing-impaired children (quiet HL—long dashed line), and noise-masked NH children (Sim HL—short dashed line) are shown. Correlation coefficients (r^2) are provided for each transfer function.

nificant differences in performance were observed for the nonsense sentences but not for the meaningful sentences. These results suggest that when contextual information was provided, the presence of HL was not a determining factor. However, when the same words were presented in nonsense sentences, perception was no longer as simple, revealing the potential effects of HL.

Second, the long-term effects of HL were examined by comparing the performance of the children with actual HL to that of the children with simulated HL. It was hypothesized that the performance of the children with actual HL would be poorer than that of the children with simulated HL due to the long-standing effects of hearing impairment on the development of speech perception. For example, the presence of HL may have made certain acoustic elements of the speech signal inaudible over the long term, slowing their ability to decode the auditory signal and fill in the missing information. This would be particularly evident for speech in unpredictable contexts like the nonsense sentences used in this study. However, the results revealed no significant differences between the groups for either type of sentence suggesting that long-term HL did not further reduce their ability to perceive the sentences.

It is important to note, however, that children are often required to perceive more complex materials in more difficult listening conditions than those in the present study. For example, the speech materials used were short sentences composed of simple words. In a learning environment such as the classroom, children are expected to perceive new and more complex phrases every day. Likewise, the masking noise used for this study was chosen to simulate HL and to facilitate the presentation of equivalent sensation level in each condition. In the classroom, children must perceive speech in a variety of background competitors. Children's performance in more difficult listening situations may better reveal any effects of long-standing HL. The most striking example of this is a study by Crandell (1993) in which he examined sentence recognition in elementary school children with minimal HLs compared to children with NH. The sentences were presented in a multi-talker noise over a range of signal-to-noise ratios. Not only did the multi-talker noise mask the perception of the sentences, the phonological content of the babble offered contextual interference as well; similar to the kind of interference children encounter in the classroom. The children with HL showed significantly greater declines in performance with increases in signal-to-noise ratio compared to the children with NH. Moreover, the poorer performance of the children with minimal HL was substantial for school-age children. That is, although the children with NH could tolerate some levels of background noise (70% recognition in −6 dB SNR), the children with minimal HLs were nearly incapable of communicating in the same level of noise (38% recognition in −6 dB SNR).

Procedurally, the results of this study suggest that the immediate effects of HL may be simulated well in children using masking noise. This technique may prove useful for certain aspects of pediatric hearing research. Difficulty recruiting sufficient numbers of children with similar HLs is a common problem. Typically, a compromise is made in which one or more recruitment criteria are relaxed. For example, children over a wide range of ages may be recruited to examine a particular configuration of HL. For other studies, the degree and configuration of loss may be allowed to vary in order to examine children within a narrow age range. Although statistical procedures are available to accommodate for variability in age or hearing level, even larger numbers of children are required to meet the assumptions of the test statistic. As a result, research in children with HL is often underpowered and only applicable to a portion of the entire population. The results of this study suggest that this procedure may be useful for examining other aspects of auditory development and hearing impairment in larger numbers of children with NH.

Another important issue to consider is the use of an AI to compare the performance of the groups and listening conditions. It should be noted that the application of an AI in this context differs from that of the speech intelligibility index procedure provided in the American National Standards Institute (1997) standard. The speech intelligibility index was developed to reconcile stimulus audibility with performance so that speech perception may be predicted on the basis of audibility alone and eliminate the need for empirical testing. This approach is particularly attractive to clinicians and scientists who work with populations that are difficult to test (e.g., infants and young children). Many revisions to the speech intelligibility index have been proposed that adjust or

change audibility values to more precisely coincide with performance. These revisions are in the form of factors (e.g., proficiency, desensitization, age, and HL) that adjust the audibility values, and therefore the transfer function, to better fit the performance data. It could be argued that the differences observed in the present study were due to the manner in which the AI was calculated. Although it may be possible to revise the AI calculation so that the transfer functions for each group and condition converge, that was not the purpose of the study. Instead, the same audibility procedure was applied in all conditions to determine the effects of elevated hearing thresholds on speech perception. The only factor accounted for in the calculation of audibility was for level distortion, which was appropriate given that the presentation levels were higher than most children with NH experience during speech perception. Factors associated with masking and near-threshold testing also may be applicable, however, the results are unlikely to change. That is, the long-term effects of HL on speech perception are minimal and the immediate effects of HL are more apparent for linguistically challenging materials.

Related to the calculation of audibility is the considerable variability observed in the performance of each group and in each listening condition. This kind of variability is commonly observed in this population (Stelmachowicz *et al.*, 2000; Scollie, 2008) and limits interpretation of the results. One possible cause for this variation may have been the manner in which the quiet and masked thresholds were matched. Specifically, thresholds were matched at octave intervals only. It is possible that the children with HL had inter-octave thresholds that varied somewhat in level resulting in more or less audibility than estimated. Because the same problem may occur if thresholds were obtained at more frequencies (e.g., 18 1/3-octave intervals), not to mention the impracticality of such an approach with children, a better solution may be to present the same frequency-shaped masking noise to both the children with NH and the children with HL. Dubno *et al.* (2000) used this procedure to elevate the thresholds of adults with NH and with HL to a specified level above target while preserving the configuration of HL. A similar approach in children may reduce undetected differences in hearing thresholds and equate AI across groups more precisely. This, in turn, may reduce some of the variability in performance at all levels of audibility.

V. SUMMARY

An experiment was conducted to determine the immediate and long-term effects of HL on the development of speech perception. The results suggest that childhood HL has the potential to delay or impair the development of speech perception as evidenced by the poorer performance of children in the simulated HL condition. However, children appear to develop compensatory strategies that are sufficient to overcome some of the deleterious effects of HL.

ACKNOWLEDGMENTS

The authors would like to thank Christina Sergi for her help with data collection, Chad Rotolo for developing and supporting the customized software used, Susan Scollie and Terry Wiley for their editorial comments, two anonymous reviewers for providing thoughtful and professional critiques during the review process, and the children and their families for taking the time to help us learn a little more about HL in children.

Allen, P., and Wightman, F. (**1994**). "Psychometric functions for children's detection of tones in noise," J. Speech Hear. Res. **37**, 205–215.

American National Standards Institute (**1997**). Methods for calculation of the speech intelligibility index (ANSI S3.5-1997), American National Standards Institute, New York.

American National Standards Institute (**1999**). Maximum permissible ambient noise levels for audiometric test rooms (ANSI S3.1 1-1999), American National Standards Institute (ANSI), New York.

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Wright, M., Psarros, C., Rattigan, K., and Tooher, R. (**2001**). "Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing," J. Speech Lang. Hear. Res. **68**, 264–285.

Boothroyd, A. (**1997**). "Auditory development of the hearing child," Scand. Audiol. Suppl. **46**, 9–16.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2009**). "Psychometric functions for pure tone intensity discrimination: Slope differences in school-aged children and adults," J. Acoust. Soc. Am. **125**, 1050–1058.

Crandell, C. C. (**1993**). "Speech recognition in noise by children with minimal degrees of sensorineural hearing loss," Ear Hear. **14**, 210–216.

Dirks, D. D., Bell, T. S., Rossman, R. N., and Kincaid, G. E. (**1986**). "Articulation index predictions of contextually dependent words," J. Acoust. Soc. Am. **80**, 82–92.

Dubno, J. R., and Ahlstrom, J. B. (**1995**). "Masked thresholds and consonant recognition in low-pass maskers for hearing-impaired and normal-hearing listeners," J. Acoust. Soc. Am. **97**, 2430–2441.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2000**). "Use of context by young and aged adults with normal hearing," J. Acoust. Soc. Am. **107**, 538–546.

Dubno, J. R., and Schaefer, A. B. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110–2121.

Dubno, J. R., and Schaefer, A. B. (**1995**). "Frequency selectivity and consonant recognition for hearing-impaired and normal-hearing listeners with equivalent masked thresholds," J. Acoust. Soc. Am. **97**, 1165–1174.

Dunn, L. M., and Dunn, L. M. (**2006**). *Peabody Picture Vocabulary Test III* (American Guidance Services, Inc., Circle Pines, MN).

Elliott, L. L. (**1979**). "Performance of children aged 9 to 17 years on a test of speech intelligibility in noise using sentence material with controlled word predictability," J. Acoust. Soc. Am. **66**, 651–653.

Fabry, D. A., and Van Tasell, D. J. (**1986**). "Masked and filtered simulation of hearing loss: Effects on consonant recognition," J. Speech Hear. Res. **29**, 170–178.

Gershkoff-Stowe, L., and Smith, L. B. (**1997**). "A curvilinear trend in naming errors as a function of early vocabulary growth," Cognit Psychol. **34**, 37–71.

Hawkins, J. E., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Hicks, C. B., and Tharpe, A. M. (**2002**). "Listening effort and fatigue in school-age children with and without hearing loss," J. Speech Lang. Hear. Res. **45**, 573–584.

Holt, R. F., and Carney, A. E. (**2007**). "Developmental effects of multiple looks in speech sound discrimination," J. Speech Lang. Hear. Res. **50**, 1404–1424.

Humes, L. E., Dirks, D. D., Bell, T. S., and Kincaid, G. E. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765–773.

Humes, L. E., Espinoza-Varas, B., and Watson, C. S. (**1988**). "Modeling sensorineural hearing loss. I. Model and retrospective evaluation," J. Acoust. Soc. Am. **83**, 188–202.

Jerger, S. (**2007**). "Current state of knowledge: Perceptual processing by children with hearing impairment," Ear Hear. **28**, 754–765.

Jerger, S., Tye-Murray, N., and Abdi, H. (**2009**). "Role of visual speech in

phonological processing by children with hearing loss," J. Speech Lang. Hear. Res. **52**, 412–434.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Needleman, A. R., and Crandell, C. C. (**1995**). "Speech recognition in noise by hearing-impaired and noise-masked normal-hearing listeners," J. Am. Acad. Audiol **6**, 414–424.
Nittrouer, S., and Miller, M. E. (**1997**). "Developmental weighting shifts for noise components of fricative-vowel syllables," J. Acoust. Soc. Am. **102**, 572–580.
Pittman, A. L., Lewis, D. E., Hoover, B. M., and Stelmachowicz, P. G. (**2005**). "Rapid word-learning in normal-hearing and hearing-impaired children: Effects of age, receptive vocabulary, and high-frequency amplification," Ear Hear. **26**, 619–629.
Pittman, A. (**2008**). "Perceptual coherence in listeners having longstanding childhood hearing losses, listeners with adult-onset hearing losses, and listeners with normal hearing," J. Acoust. Soc. Am. **123**, 441–449.
Scollie, S., Seewald, R., Cornelisse, L., Moodie, S., Bagatto, M., Laurnagaray, D., Beaulac, S., and Pumford, J. (**2005**). "The desired sensation level multistage input/output algorithm," Trends Amplif. **9**, 159–197.
Scollie, S. D. (**2008**). "Children's speech recognition scores: The speech intelligibility index and proficiency factors for age and hearing level," Ear Hear. **29**, 543–556.
Sherbecoe, R. L., and Studebaker, G. A. (**2002**). "Audibility-index functions for the connected speech test," Ear Hear. **23**, 385–398.
Stelmachowicz, P. G., Hoover, B. M., Lewis, D. E., Kortekaas, R. W., and Pittman, A. L. (**2000**). "The relation between stimulus context, speech audibility, and perception for normal-hearing and hearing-impaired children," J. Speech Lang. Hear. Res. **43**, 902–914.
Stuart, A. (**2005**). "Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise," Ear Hear. **26**, 78–88.
Studebaker, G. A., Gilmore, C., and Sherbecoe, R. L. (**1993**). "Performance-intensity functions at absolute and masked thresholds," J. Acoust. Soc. Am. **93**, 3418–3421.
Van Deun, L., van Wieringen, A., Van den Bogaert, T., Scherf, F., Offeciers, F. E., Van de Heyning, P. H., Desloovere, C., Dhooge, I. J., Deggouj, N., De Raeve, L., and Wouters, J. (**2009**). "Sound localization, sound lateralization, and binaural masking level differences in young children with normal hearing," Ear Hear. **30**, 178–190.
Wightman, F., Allen, P., Dolan, T., Kistler, D., and Jamieson, D. (**1989**). "Temporal resolution in children," Child Dev. **60**, 611–624.

An algorithm that improves speech intelligibility in noise for normal-hearing listeners

Gibak Kim, Yang Lu, Yi Hu, and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75080

(Received 30 October 2008; revised 27 March 2009; accepted 1 July 2009)

Traditional noise-suppression algorithms have been shown to improve speech quality, but not speech intelligibility. Motivated by prior intelligibility studies of speech synthesized using the ideal binary mask, an algorithm is proposed that decomposes the input signal into time-frequency (T-F) units and makes binary decisions, based on a Bayesian classifier, as to whether each T-F unit is dominated by the target or the masker. Speech corrupted at low signal-to-noise ratio (SNR) levels (−5 and 0 dB) using different types of maskers is synthesized by this algorithm and presented to normal-hearing listeners for identification. Results indicated substantial improvements in intelligibility (over 60% points in −5 dB babble) over that attained by human listeners with unprocessed stimuli. The findings from this study suggest that algorithms that can estimate reliably the SNR in each T-F unit can improve speech intelligibility.
 [DOI: 10.1121/1.3184603]

PACS number(s): 43.72.Ar, 43.72.Dv [MSS] Pages: 1486–1494

I. INTRODUCTION

Dramatic advances have been made in automatic speech recognition (ASR) technology (Rabiner, 2003). Despite these advances, human listener's word error rates are often more than an order of magnitude lower than those of state-of-the art recognizers in both quiet and degraded environments (Lippmann, 1997, Sroka and Braida, 2005; Scharenborg, 2007). Large advances have also been made on the development of algorithms that suppress noise without introducing much distortion to the speech signal (Loizou, 2007). These algorithms, however, have been shown to improve primarily the subjective quality of speech rather than speech intelligibility (Hu and Loizou, 2007a, 2007b). Speech quality is highly subjective in nature and can be easily improved, at least to some degree, by suppressing the background noise. In contrast, intelligibility is related to the underlying message or content of the spoken words and can be improved only by suppressing the background noise without distorting the underlying target speech signal. Designing such algorithms has been extremely challenging, partly because of inaccurate and often unreliable estimates of the background noise (masker) signal from the corrupted signal (often acquired using a single microphone). Algorithms that would improve intelligibility of speech in noisy environments would be extremely useful not only in cellphone applications but also in hearing aids/cochlear implant devices. The development of such algorithms has remained elusive for several decades (Lim, 1978; Hu and Loizou, 2007a), and perhaps this was due to the fact that algorithms were sought that would work for all types of maskers and for all signal-to-noise ratio (SNR) levels, clearly an ambitious goal. In some ASR applications (e.g., voice dictation) and hearing aid applications (e.g., Zakis *et al.*, 2007), however, the algorithm can be speaker and/or masker dependent. Such an approach was taken in this study.

The approach that is being pursued in the present study was motivated by intelligibility studies of speech synthesized using the ideal binary mask (IdBM) (Brungart *et al.*, 2006; Li and Loizou, 2008b, 2008a). The IdBM is a technique explored in computational auditory scene analysis (CASA) that retains the time-frequency (T-F) regions of the target signal that are stronger than the interfering noise (masker), and removes the regions that are weaker than the interfering noise. Previous studies have shown that multiplying the IdBM with the noise-masked signal can yield large gains in intelligibility, even at extremely low (−5, −10 dB) SNR levels (Brungart *et al.*, 2006; Li and Loizou, 2008b). In these studies, prior knowledge of the true IdBM was assumed. In practice, however, the binary mask needs to be estimated from the corrupted signal. Motivated by the successful application of the IdBM technique for improvement of speech intelligibility, we focused on developing a classifier that would identify T-F units as either target-dominated or masker-dominated.[1] This is a conceptually and computationally simpler task than attempting to mimic the human auditory scene analysis using grouping and segmentation principles (Hu and Wang, 2004, 2008; Wang and Brown, 2006), such as common periodicity across frequency, common offsets and onsets, and common amplitude and frequency modulations. Such techniques would require the reliable detection of F0 and onset/offset segments in noise, a formidable task. The challenge faced in the present work is in the design of an accurate classifier capable of operating at negative SNR levels, wherein performance of normal-hearing (NH) listeners is known to degrade. While many techniques have been proposed to estimate the IdBM (Wang and Brown, 2006; Hu and Loizou, 2008), none of the techniques were evaluated with human listeners at extremely low (negative) SNR levels. Most of the proposed algorithms have been

[a)]Author to whom correspondence should be addressed. Electronic mail: loizou@utdallas.edu

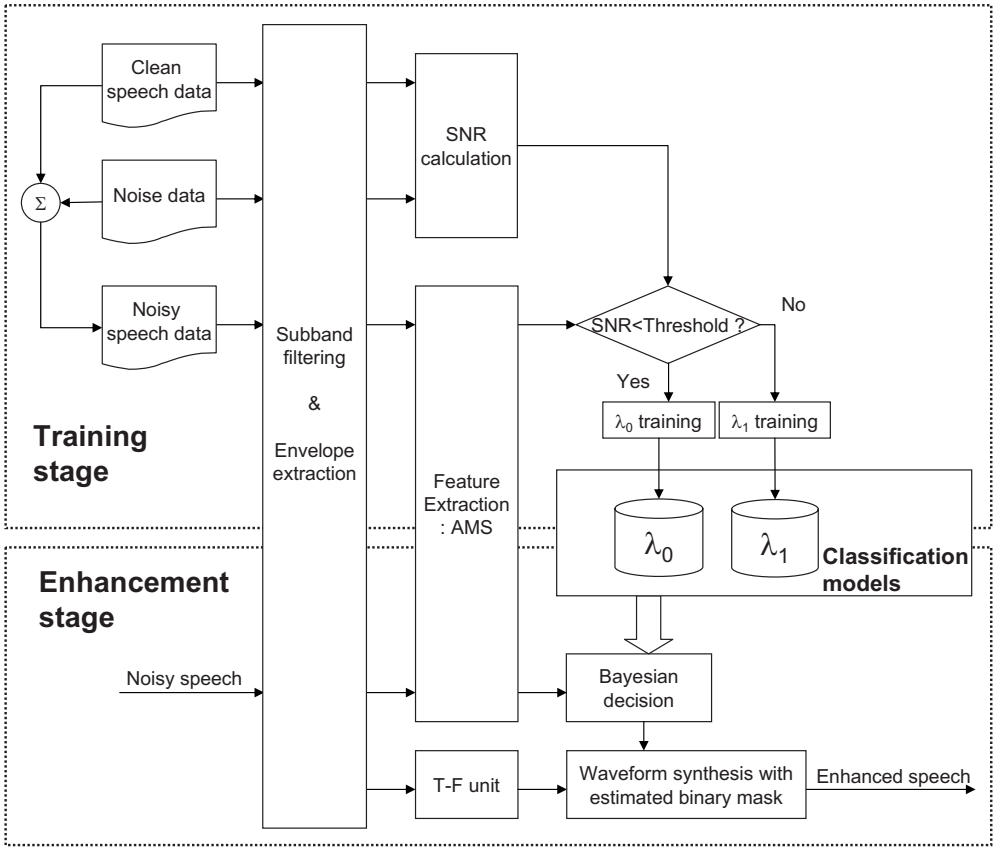


FIG. 1. Block diagram of the training and enhancement stages of the proposed algorithm.

evaluated using objective measures (Hu and Wang, 2004, 2008) and in terms of ASR error rates (Seltzer *et al.*, 2004) rather than in terms of speech intelligibility scores.

The goal of this study is to evaluate the intelligibility of speech synthesized via an algorithm that decomposes the input signal into T-F regions, with the use of a crude auditory-like filterbank, and uses a simple binary Bayesian classifier to retain target-dominated spectro-temporal regions while removing masker-dominated spectro-temporal regions. Amplitude modulation spectrograms (AMSs) (Kollmeier and Koch, 1994) were used as features for training Gaussian mixture models (GMMs) to be used as classifiers. Unlike most noise-suppression algorithms (Loizou, 2007), the proposed algorithm requires no speech/noise detection nor the estimation of noise statistics. Speech corrupted at low SNR levels by different types of maskers is synthesized using this algorithm and presented to human listeners for identification. The present work tests the hypothesis that algorithms that make use of knowledge of when the target is stronger than the masker (at each T-F unit) *can* improve speech intelligibility in noisy conditions.

II. PROPOSED NOISE-SUPPRESSION ALGORITHM

Figure 1 shows the block diagram of the proposed algorithm, consisting of a training stage (top panel) and an intelligibility enhancement stage (bottom panel). In the training stage, features are extracted, typically from a large speech corpus, and then used to train two GMMs representing two feature classes: target speech dominating the masker and masker dominating target speech. AMS are used in this work as features, as they are neurophysiologically and psychoacoustically motivated (Kollmeier and Koch, 1994; Langner and Schreiner, 1988). In the enhancement stage, a Bayesian classifier is used to classify the T-F units of the noise-masked signal into two classes: target-dominated and masker-dominated. Individual T-F units of the noise-masked signal are retained if classified as target-dominated or eliminated if classified as masker-dominated, and subsequently used to reconstruct the enhanced speech waveform.

A. Feature extraction

The noisy speech signal is first bandpass filtered into 25 channels according to a mel-frequency spacing (shown in the subband filtering block in Fig. 1). The envelopes in each band are computed by full-wave rectification and then decimated by a factor of 3 (shown in the envelope extraction block in Fig. 1). The decimated envelope signals are subsequently segmented into overlapping segments of 128 samples (32 ms) with an overlap of 64 samples. Each segment is Hanning windowed and following zero-padding, a 256-point fast Fourier transform (FFT) is computed. The FFT computes the modulation spectrum in each channel, with a frequency resolution of 15.6 Hz. Within each band, the FFT magnitudes are multiplied by 15 triangular-shaped windows spaced uniformly across the 15.6–400 Hz range and summed up to produce 15 modulation spectrum amplitudes. The 15 modulation amplitudes represent the AMS feature vector (Tchorz and Kollmeier, 2003), which we denote

by $\mathbf{a}(\tau,k)$, where τ indicates the time index and k indicates the subband. In addition to the AMS feature vector, we also include delta features to capture feature variations across time and frequency. The overall feature vector is given by

$$\mathbf{A}(\tau,k)=[\mathbf{a}(\tau,k),\Delta\mathbf{a}_T(\tau,k),\Delta\mathbf{a}_K(\tau,k)], \tag{1}$$

where

$$\Delta\mathbf{a}_T(1,k)=\mathbf{a}(2,k)-\mathbf{a}(1,k), \quad \tau=1,$$

$$\Delta\mathbf{a}_T(\tau,k)=\mathbf{a}(\tau,k)-\mathbf{a}(\tau-1,k), \quad \tau=2,\ldots,T, \tag{2}$$

$$\Delta\mathbf{a}_K(\tau,1)=\mathbf{a}(\tau,2)-\mathbf{a}(\tau,1), \quad k=1,$$

$$\Delta\mathbf{a}_K(\tau,k)=\mathbf{a}(\tau,k)-\mathbf{a}(\tau,k-1), \quad k=2,\ldots,K, \tag{3}$$

where $\Delta\mathbf{a}_T(\tau,k)$ and $\Delta\mathbf{a}_K(\tau,k)$ denote the delta feature vectors computed across time and frequency, respectively, and T is the total number of segments. The number of subbands, K, was set to 25 in this work, and the total dimension of the feature vector $\mathbf{A}(\tau,k)$ was 45 $(=3\times15)$.

B. Training stage

A two-class Bayesian classifier was used to estimate the binary mask for each T-F unit. The distribution of the feature vectors of each class was represented with a GMM. The two classes, denoted as λ_0 for mask 0 (masker-dominated T-F units) and λ_1 for mask 1 (target-dominated T-F units), were further subdivided into two smaller classes, i.e., $\lambda_0=\{\lambda_0^0,\lambda_0^1\}$, $\lambda_1=\{\lambda_1^0,\lambda_1^1\}$. This sub-class division yielded faster convergence in GMM training and better classification. In the training stage, the noisy speech spectrum, $Y(\tau,k)$, at time slot τ and k-th subband, was classified into one of four subclasses as follows:

$$Y(\tau,k)\in\begin{cases}\lambda_0^0 & \text{if } \xi(\tau,k)<T_{\mathrm{SNR0}}\\ \lambda_0^1 & \text{if } T_{\mathrm{SNR0}}\le\xi(\tau,k)<T_{\mathrm{SNR}}\\ \lambda_1^0 & \text{if } T_{\mathrm{SNR}}\le\xi(\tau,k)<T_{\mathrm{SNR1}}\\ \lambda_1^1 & \text{if } T_{\mathrm{SNR1}}\le\xi(\tau,k),\end{cases} \tag{4}$$

where $\xi(\tau,k)$ is the local (true) SNR computed as the ratio of envelope energies of the (clean) target speech and masker signals, and T_{SNR0}, T_{SNR1}, and T_{SNR} are thresholds. The T_{SNR0} was chosen in the training stage so as to have equal amount of training data in the λ_0^0 and λ_0^1 classes. Classification performance was not found to be sensitive to this threshold value. The SNR threshold, T_{SNR}, was set to −8 dB for the first 15 frequency bands (spanning 68–2186 Hz) and to −16 dB for the higher frequency bands. This was done to account for the non-uniform masking of speech by the maskers across the spectrum. We utilized 256-mixture Gaussian models for modeling the distributions of the feature vectors in each class. The initial Gaussian model parameters (mixture weights, mean vectors, and covariance matrices) were obtained by running 15 iterations of the k-means clustering algorithm. Full covariance matrices were used for each mixture. If a particular covariance matrix was found to be singular during training, the corresponding mixture weight was set to zero. The final GMM parameters were obtained using the expectation-maximization training algorithm (Dempster *et al.*, 1977). The *a priori* probability for each sub-class $[P(\lambda_0^0),P(\lambda_0^1),P(\lambda_1^0),P(\lambda_1^1)]$ was calculated by counting the number of feature vectors belonging to the corresponding class and dividing that by the total number of feature vectors.

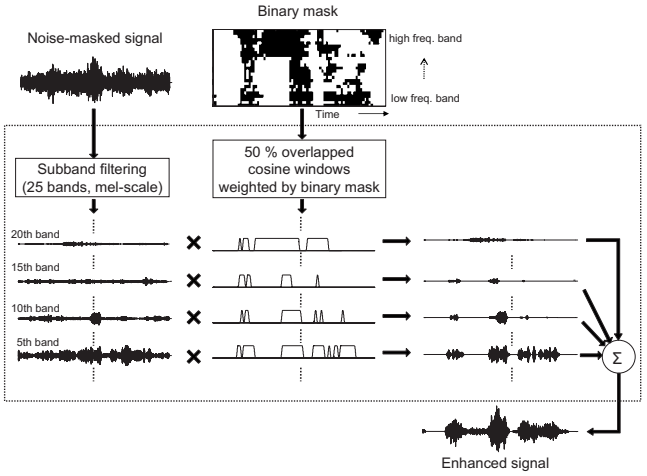


FIG. 2. Block diagram of the waveform synthesis stage of the proposed algorithm.

C. Enhancement stage

In the enhancement stage, the binary masks of each T-F unit are first estimated using a Bayesian classifier. Each T-F unit of noisy speech signal is subsequently retained or eliminated by the estimated binary mask and synthesized to produce the enhanced speech waveforms.

1. Bayesian classification

The T-F units are classified as λ_0 or λ_1 by comparing two *a posteriori* probabilities, $P(\lambda_0|\mathbf{A}_Y(\tau,k))$ and $P(\lambda_1|\mathbf{A}_Y(\tau,k))$. This comparison produces an estimate of the binary mask, $G(\tau,k)$, as follows:

$$G(\tau,k)=\begin{cases}0 & \text{if } P(\lambda_0|\mathbf{A}_Y(\tau,k))>P(\lambda_1|\mathbf{A}_Y(\tau,k))\\ 1 & \text{otherwise},\end{cases} \tag{5}$$

where $P(\lambda_0|\mathbf{A}_Y(\tau,k))$ is computed using Bayes' theorem as follows:

$$P(\lambda_0|\mathbf{A}_Y(\tau,k))=\frac{P(\lambda_0,\mathbf{A}_Y(\tau,k))}{P(\mathbf{A}_Y(\tau,k))}$$
$$=\frac{P(\lambda_0^0)P(\mathbf{A}_Y(\tau,k)|\lambda_0^0)+P(\lambda_0^1)P(\mathbf{A}_Y(\tau,k)|\lambda_0^1)}{P(\mathbf{A}_Y(\tau,k))}. \tag{6}$$

The *a posteriori* probability $P(\lambda_1|\mathbf{A}_Y(\tau,k))$ is computed similarly.

2. Waveform synthesis

Figure 2 shows the block diagram of the waveform synthesis stage. The corrupted speech signal is first filtered into 25 bands (same bands used in the feature-extraction stage). To remove across-channel differences, the output of each filter is time reversed, passed through the filter, and reversed

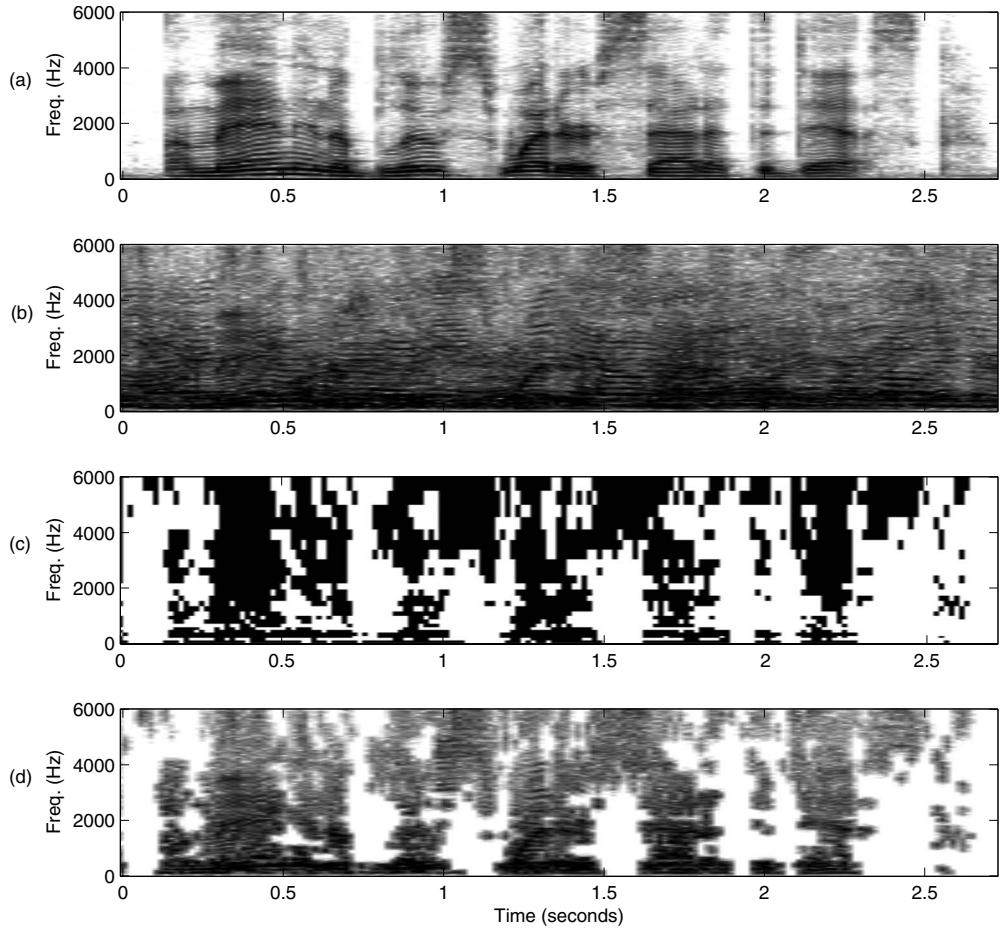


FIG. 3. (a) Wide-band spectrogram of an IEEE sentence in quiet. (b) Spectrogram of corrupted sentence by multitalker babble at −5 dB SNR. (c) Binary mask estimated using Eq. (5), with black pixels indicating target-dominated T-F units and white pixels indicating masker-dominated T-F units. (d) Synthesized signal obtained by multiplying the binary mask shown in panel (c) with the corrupted signal shown in panel (b).

again (Wang and Brown, 2006). The filtered waveforms are windowed with a raised cosine every 32 ms with 50% overlap between segments, and then weighted by the estimated binary mask [Eq. (5)]. Finally, the estimated target signal is reconstructed by summing the weighted responses of the 25 filters. Figure 3 shows an example spectrogram of a synthesized signal using the proposed algorithm. In this example, the clean speech signal [Fig. 3(a)] is mixed with multitalker babble at −5 dB SNR [Fig. 3(b)]. The estimated binary mask [as per Eq. (5)] and synthesized waveform are shown in Figs. 3(c) and 3(d), respectively.

III. LISTENING EXPERIMENTS

A. Stimuli

Sentences taken from the IEEE database (IEEE, 1969) were used as test material. The sentences in the IEEE database are phonetically balanced with relatively low word-context predictability. The sentences were produced by one male and one female speaker in a sound-proof booth using Tucker-Davis Technologies (TDT) recording equipment. The sentences were originally recorded at a sampling rate of 25 kHz and downsampled to 12 kHz. Three types of noise (20-talker babble, factory, speech-shaped noise) were used as maskers. The (steady) speech-shaped noise was stationary having the same long-term spectrum as the sentences in the IEEE corpus. The factory noise was taken from the NOISEX database (Varga and Steeneken, 1993), and the babble (20 talkers with equal number of female and male talkers) was taken from the Auditec CD (St. Louis, MO).

A total of 390 IEEE sentences were used to train the GMM models. These sentences were corrupted by three types of noise at −5, 0, and 5 dB SNR. The maskers were randomly cut from the noise recordings and mixed with the target sentences at the prescribed SNRs. Each corrupted sentence had thus a different segment of the masker, and this was done to evaluate the robustness of the Bayesian classifier in terms of generalizing to different segments of the masker having possibly different temporal/spectral characteristics. Three different training sets were prepared to train three GMM models and three test sets were used for the evaluation of the GMM models. Two types of GMM models were trained: (1) a single-noise GMM model (denoted as sGMM) trained only on a single type of noise (tested with the same type of noise) and (2) a multi-noise GMM model (denoted as mGMM) trained on all three types of noise (tested with one of the three types of noise). The latter models (mGMM) were used to assess the performance and robustness of a single GMM model in multiple noisy environments. As we were limited by the total number of sentences available in the IEEE corpus, we used different sets of training sentences

randomly assigned to the various conditions. This was necessary to avoid testing NH listeners with the same sentences used in the training stage. More specifically, three sets of training data were created with each set having 390 sentences and two training sets having an overlap of 150 sentences. There was no overlap between the training and testing sets in any condition.

B. Procedure

A total of 17 NH listeners (all native speakers of English) were recruited for the listening tests. All subjects were paid for their participation. The listeners were randomly assigned to the conditions involving processed IEEE sentences produced by the male and female speakers (eight listened to the IEEE sentences produced by the male speaker and nine listened to the IEEE sentences produced by the female speaker). Subjects participated in a total of 24 conditions [=2 SNR levels (−5 dB, 0 dB) × 4 processing conditions × 3 types of maskers]. The processing conditions included speech processed using (1) sGMM models, (2) mGMM models, (3) the idBM, and (4) the unprocessed (noise-masked) stimuli. The IdBM condition was included as a control condition to assess the performance of the proposed algorithms relative to the ideal condition in which we have *a priori* knowledge of the local SNR and IdBM. The IdBM was obtained by comparing the local (true) SNR against a predefined threshold. The SNR at each T-F unit was computed as the ratio of the envelope energies of the (clean) target speech and masker signals in each unit. The IdBM takes a value of 1 if the local SNR is greater than the threshold and takes the value of 0 otherwise.

The experiments were performed in a sound-proof room (Acoustic Systems, Inc.) using a PC connected to a Tucker-Davis system 3. Stimuli were played to the listeners monaurally through Sennheiser HD 250 Linear II circumaural headphones at a comfortable listening level. Prior to the sentence test, each subject listened to a set of noise-corrupted sentences to be familiarized with the testing procedure. During the test, subjects were asked to write down the words they heard. The whole listening test lasted for about 2–3 h, which was split into two sessions each lasting 1–1.5 h. 5 min breaks were given to the subjects every 30 min. Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists were repeated across conditions. The sentence lists were counterbalanced across subjects. Sentences were presented to the listeners in blocks, and 20 sentences were presented in each block for each condition. The order of the conditions was randomized across subjects.

IV. RESULTS

The mean performance, computed in terms of percentage of words identified correctly by the NH listeners, are plotted in Fig. 4 for sentences produced by male (top panel) and female speakers (bottom panel). A substantial improvement in intelligibility was obtained with the proposed algorithm using both sGMM and mGMM models, compared to that attained by human listeners with unprocessed (corrupted) speech. The improvement (over 60% points in some

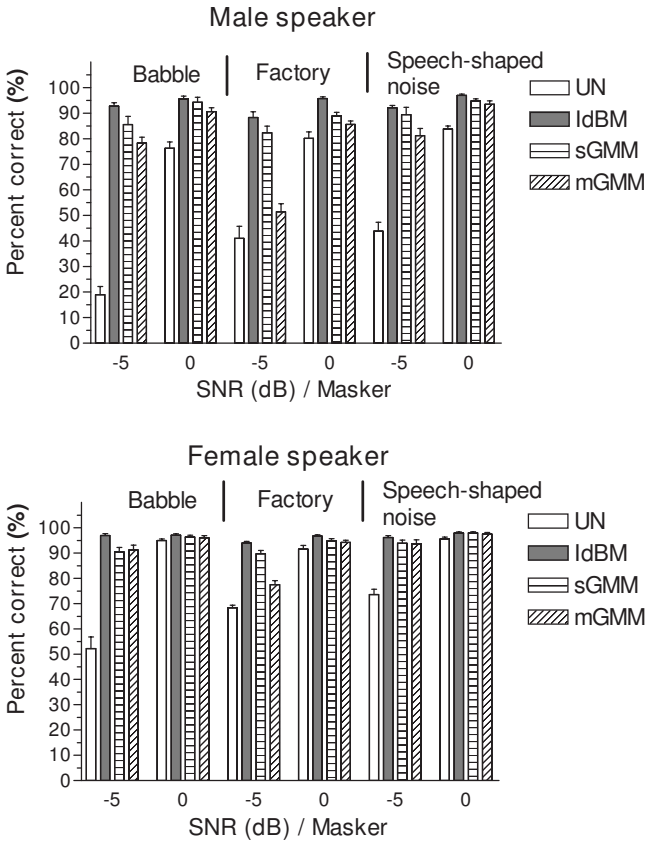


FIG. 4. Mean speech recognition scores obtained by 17 NH listeners for corrupted (unprocessed) sentences (denoted as UN), sentences processed using the sGMM (single-noise trained GMMs) and mGMM models (multiple-noise trained GMMs), and sentences processed using the IdBM in the various SNR/masker conditions. Error bars indicate standard errors of the mean.

cases) was more evident at −5 dB SNR levels for all three maskers tested. Performance at 0 dB SNR in the female-speaker conditions (bottom panel of Fig. 4) was limited in most cases by ceiling effects. Analysis of variance (with repeated measures) indicated significant effect of masker type [$F(2,14)=41.2$, $p<0.0005$], significant effect of SNR level [$F(1,7)=583.3$, $p<0.0005$], and significant effect of processing algorithm [$F(3,21)=314.1$, $p<0.0005$]. All interactions were found significant ($p<0.05$). *Post-hoc* analysis (Schéffe), corrected for multiple comparisons, was done to assess significant differences between conditions. For the male-speaker data (top panel of Fig. 4), performance at −5 dB SNR with mGMM models was significantly ($p<0.0005$) higher than that attained by the listeners in all baseline masker conditions (unprocessed sentences) except the factory condition. Performance at −5 and 0 dB SNR with sGMM models was significantly ($p<0.005$) higher than that attained in all baseline masker conditions. For the female-speaker data, performance at −5 dB SNR with sGMM and mGMM models was significantly ($p<0.0005$) higher than performance obtained with unprocessed speech in all masker conditions. There was no significant ($p>0.05$) difference in scores between the various algorithms in the 0 dB SNR masker conditions, as performance was limited by ceiling effects. Consistent with prior studies (Brungart *et al.*, 2006; Li and Loizou, 2008b), highest performance was obtained

TABLE I. Hit (HIT) and false alarm (FA) rates obtained using the sGMM and mGMM models for the male-speaker and female-speaker data in the various masker conditions.

			Babble		Factory		Speech-shaped	
Speaker	Model	Performance	−5 dB	0 dB	−5 dB	0 dB	−5 dB	0 dB
Male	sGMM	HIT	86.96%	82.49%	80.17%	75.18%	88.30%	84.99%
		FA	14.54%	10.43%	15.27%	11.44%	12.20%	9.48%
		HIT-FA	72.42%	72.06%	64.90%	63.74%	76.10%	75.51%
	mGMM	HIT	78.24%	75.94%	75.36%	75.79%	76.90%	76.61%
		FA	18.83%	17.69%	24.12%	22.91%	12.75%	13.24%
		HIT-FA	59.41%	58.25%	51.24%	52.88%	64.15%	63.37%
Female	sGMM	HIT	89.95%	86.21%	83.28%	79.52%	89.28%	85.82%
		FA	15.23%	11.36%	17.26%	12.12%	12.65%	10.26%
		HIT-FA	74.72%	74.85%	66.02%	67.40%	76.63%	75.56%
	mGMM	HIT	82.28%	79.46%	82.58%	81.93%	81.76%	80.34%
		FA	18.03%	16.63%	25.84%	21.60%	13.49%	13.86%
		HIT-FA	64.25%	62.83%	56.74%	60.33%	68.27%	66.48%

with the IdBM. Performance with sGMM models was comparable to that obtained with the IdBM in nearly all conditions. Note that with the exception of one condition (factory noise at −5 dB SNR), performance with mGMM models did not differ significantly ($p>0.05$) from that obtained with sGMM models, an outcome demonstrating the potential of the proposed approach in training a single GMM model that would be effective in multiple listening environments.

To quantify the accuracy of the binary Bayesian classifier, we computed the average hit (HIT) and false alarm (FA) rates for three test sets not included in the training. Each test set comprised of 60 sentences, for a total of 180 sentences corresponding to 893,950 T-F units (35,758 frames × 25 frequency bands) for the male-speaker sentences and 811,750 T-F units (32 470 × 25 frequency bands) for the female-speaker sentences. HIT and FA rates were computed by comparing the estimated binary mask against the (oracle) IdBM. Table I shows the results obtained using sGMM and mGMM models in the various masker conditions. High hit rates (lowest with factory noise at 0 dB, male speaker; 75.18%) and low false-alarm rates (highest with factory noise at −5 dB, female speaker; 17.26%) were obtained with sGMM models. The hit rate obtained with mGMM models was about 10% lower than that of sGMM models for the male speaker. The difference was much smaller for the female speaker (about 5%). As demonstrated in Li and Loizou, (2008b), low false alarm rates (<20% assuming high hit rates) are required to achieve high levels of speech intelligibility.

To predict the intelligibility of speech synthesized using the estimated binary masks (based on the Bayesian classifier), we propose a simple metric based on the difference between the hit rate and false alarm rates, i.e., HIT-FA. This metric bears resemblance to the sensitivity index, d', used in psychoacoustics (Macmillan and Creelman, 2005). The index d' is derived assuming a Gaussian distribution of responses. No such assumptions are made with the use of the HIT-FA difference metric. A modestly high correlation ($r=0.80$) was obtained between this simple difference metric and speech intelligibility scores based on data from the same three test sets used in the listening experiments (see Fig. 5). More generally, the difference metric, $d_\alpha=\alpha\cdot H-(1-\alpha)\cdot FA$, can be used to obtain higher correlation by optimizing the value of α for different speech materials. A value of $\alpha=0.3$ yielded a maximum correlation of $r=0.84$ for our test materials (IEEE sentences), suggesting that more weight needs to be placed on FA values, an outcome consistent with intelligibility studies (Li and Loizou, 2008b). Table II shows the performance (in terms of HIT and FA rates) of two conventional noise reduction algorithms, the Wiener algorithm (Scalart and Filho, 1996) and the MMSE algorithm (Ephraim and Malah, 1984). The binary mask was estimated by comparing the SNR in each frequency bin against the same threshold T_{SNR} used in the proposed algorithm (see Sec II B). The SNR was estimated from the corrupted signal using the decision-directed approach (Ephraim and Malah, 1984). As can be seen, the hit rates obtained by the GMM binary classifiers (Table I) are substantially higher than those obtained with conventional noise reduction algorithms. This outcome might explain, at least, partially why current noise reduction algorithms, even the most sophisticated ones, do not improve speech intelligibility (Hu and Loizou, 2008).

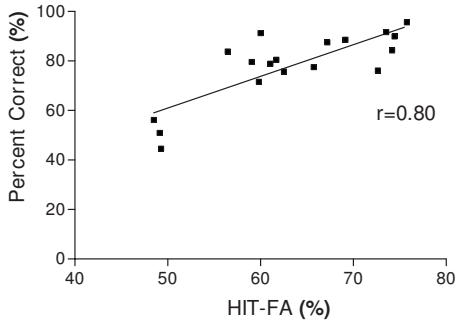


FIG. 5. Scatter plot showing the correlation between human listener's recognition scores, obtained for the male-speaker data at −5 dB in the three masker conditions, and a metric based on the difference between the resulting hit-rate (HIT) and false alarm (FA) values of the binary Bayesian classifier.

TABLE II. Binary mask accuracy obtained by two conventional noise reduction algorithms for the male-speaker data at −5 dB SNR.

	Babble		Factory		Speech-shaped	
	Wiener	MMSE	Wiener	MMSE	Wiener	MMSE
HIT	54.60%	68.44%	53.14%	57.52%	58.59%	58.89%
FA	55.62%	66.94%	54.48%	60.38%	50.44%	52.24%
HIT-FA	−1.02%	1.50%	−1.34%	−2.86%	8.15%	6.65%

To assess the robustness of the binary classifier in terms of handling speakers not included in the training, we performed a cross-gender test wherein we used the male-trained models to classify AMS features extracted from the female-speaker data, and vice versa (see Table III). The performance obtained with the cross-gender models was comparable to that obtained with the same-speaker models (Table I) showing differences ranging from 2.25% (factory noise, female speaker) up to 9.77% (speech-shaped noise, male speaker).

Finally, to quantify the gain in classification accuracy obtained with the proposed delta features [Eqs. (2) and (3)], we compared the hit and false alarm rates obtained with and without the use of delta features (see Table IV). The same test set used in the evaluation of the cross-gender models (Table III) was used in the evaluation of the delta features. As can be seen, delta features improved the hit rate considerably (by as much as 20% in some cases), without increasing the false alarm rate.

V. DISCUSSION AND CONCLUSIONS

Large gains in intelligibility were achieved with the proposed algorithm (Fig. 4). The intelligibility of speech processed by the proposed algorithm was substantially higher than that achieved by human listeners listening to unprocessed (corrupted) speech, particularly at extremely low SNR levels (−5 dB). We attribute this to the accurate classification of T-F units into target- and masker-dominated T-F units, and subsequently reliable estimation of the binary mask. As demonstrated by several intelligibility studies with NH listeners (Brungart *et al.*, 2006; Li and Loizou, 2008b) access to reliable estimates of the binary mask can yield substantial gains in intelligibility. The accurate classification of T-F units into target- and masker-dominated T-F units was accomplished with the use of neurophysiologically-motivated features (AMS) and carefully designed Bayesian classifiers (GMMs). Unlike the mel-frequency cepstrum coefficients (Davis and Mermelstein, 1980) features commonly used in ASR, the AMS features capture information about amplitude and frequency modulations, known to be critically important for speech recognition (Zeng *et al.*, 2005). Furthermore, the proposed delta features [Eqs. (2) and (3)] are designed to capture to some extent temporal and spectral correlations. Unlike the delta features commonly used in ASR (Furui, 1986), the simplified delta features proposed in Eq. (2) use only past information and are therefore amenable to real-time implementation with low latency.

GMMs are known to accurately represent a large class of feature distributions, and as classifiers, GMMs have been used successfully in several applications and, in particular speaker recognition (e.g., Reynolds and Rose, 1995). Other classifiers (e.g., neural networks, and support vector machines) could alternatively be used (Tchorz and Kollmeier, 2003). Our attempt, however, to use neural networks[2] as classifiers was not very successful as poorer performance was observed, particularly when different segments (randomly cut) of the masker were mixed with each test sentence (as done in the present study).

There exist a number of differences in our approach that distinguishes it from previous attempts to estimate the binary mask. First, their approach is simple as it is based on the design of an accurate (binary) Bayesian classifier. Others (Wang and Brown, 2006) focused on developing sophisticated grouping and segmentation algorithms that were motivated largely by existing knowledge in auditory scene analysis (Bregman, 1990). Second, the resolution of the auditory filters used in the present work is crude compared to that used by humans. A total of 128 Gammatone filters have been used by others (Brungart *et al.*, 2006; Hu and Wang, 2004)

TABLE III. Classification of male-speaker data using the female-speaker model (sGMM) and classification of the female-speaker data using the male-speaker model (sGMM) at −5 dB SNR.

		Male-speaker model			Female-speaker model		
		Babble	Factory	Speech-shaped	Babble	Factory	Speech-shaped
Male-speaker data	HIT	87.82%	82.88%	89.04%	79.82%	75.89%	78.68%
	FA	16.06%	16.97%	12.20%	15.34%	16.41%	11.61%
	HIT-FA	71.76%	65.91%	76.84%	64.48%	59.48%	67.07%
Female-speaker data	HIT	88.22%	82.68%	88.62%	89.52%	82.42%	88.81%
	FA	18.78%	17.34%	16.11%	13.72%	14.83%	10.83%
	HIT-FA	69.44%	65.34%	72.51%	75.80%	67.59%	77.98%

TABLE IV. Performance comparison, in terms of hits and false alarm rates, between the AMS feature vectors and AMS+Delta feature vectors for the male-speaker data at −5 dB SNR.

	Babble		Factory		Speech-shaped	
	AMS only	AMS+Delta	AMS only	AMS+Delta	AMS only	AMS+Delta
HIT	79.46%	87.82%	60.58%	82.88%	76.12%	89.04%
FA	18.19%	16.06%	19.32%	16.97%	15.84%	12.20%
HIT-FA	61.27%	71.76%	41.26%	65.91%	60.28%	76.84%

for modeling the auditory periphery. A smaller number (25) of channels was used in this work for two reasons: (a) to keep the feature dimensionality small and (b) to make it appropriate for hearing aid and cochlear implant applications, wherein the signal is typically processed through a small number of channels. Previous work in our laboratory (Li and Loizou, 2008a) demonstrated that spectral resolution has a significant effect on the intelligibility of IdBM speech, but the use of 25 channels seemed to be sufficient for accurate speech recognition. Third, our approach required limited amount of training data. Fewer than 400 sentences ($\sim$20 min) were used for training compared to thousands of sentences ($\sim$1–2 h) used by others (Seltzer *et al.*, 2004). Finally, the GMM training used in this work does not require access to a labeled speech corpus, while the methods proposed by others required the use of accurate F0 values or voiced/unvoiced segmentation (Hu and Wang, 2004, 2008; Seltzer *et al.*, 2004).

The proposed algorithm can be used not only for robust ASR (e.g., Cooke *et al.*, 2001) or cellphone applications but also for hearing aids or cochlear implant devices. Modern hearing aids use sound classification algorithms (e.g., Nordqvist and Leijon, 2004) to identify different listening situations and adjust accordingly hearing aid processing parameters. All advantages cited above make the proposed approach suitable for trainable hearing aids (Zakis *et al.*, 2007) and cochlear implant devices. As these devices are powered by a digital signal processor chip, the training can take place at the command of the user whenever in a new listening environment. Following the training stage, the user can initiate the proposed algorithm to enhance speech intelligibility in extremely noisy environments (e.g., restaurants). As shown in Sec. III, a single GMM trained on multiple types of noise (mGMM) can yield high performance; however, a user might encounter a new type of noise not included in the training set. In such circumstances, either new training needs to be initiated or perhaps adaptation techniques can be used to adapt the parameters of existing GMM models to the new data (Reynolds *et al.*, 2000).

Humans outperform ASR and CASA systems on various recognition tasks and are far better at dealing with accents, noisy environments, and differences in speaking style/rate (Lippmann, 1997; Scharenborg, 2007). Neither ASR nor CASA algorithms, however, have yet reached the level of performance obtained by human listeners, despite the level of sophistication built in these algorithms (Lippmann, 1997). The present study demonstrated that if the goal of CASA is to design algorithms that would perform as well or better than humans, it is not always necessary to mimic all aspects of the human auditory processing. Knowledge of when the target is stronger than the masker in each T-F unit is all that is required to achieve high levels of speech understanding (Li and Loizou, 2008b). This reduces the problem to that of designing an accurate binary classifier [see Eq. (5)]. Computers can generally be trained to classify accurately not only binary datasets (as in the present work) but also complex data patterns. The humans' ability, however, to detect the target signal in the presence of a masker within a critical band is limited by simultaneous (and temporal) masking and is dependent on several factors including the masker frequency (in relation to the target's), the masker level and the type of masker (e.g., tonal or noise-like) (Moore, 2003). The present study demonstrated that computer algorithms can be designed to overcome these shortcomings and subsequently improve speech intelligibility in noisy conditions.

ACKNOWLEDGMENTS

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

[1]A T-F unit is said to be target-dominated if its local SNR is greater than 0 dB and is said to be masker-dominated otherwise. These definitions can be extended by using a threshold other than 0 dB. In this paper, we define a target-dominated unit as a T-F unit for which the SNR is greater than a predefined threshold even if the power of the target signal is smaller than that of the masker (this occurs when the chosen threshold is $<$0 dB).

[2]A standard feed-forward neural network was trained with the same AMS feature vectors using the back-propagation algorithm. The network consisted of an input layer of 375 neurons (15×25), a hidden layer with 225 neurons, and an output layer with 25 output neurons, one for each channel. The output neuron activities indicated the respective SNR in each channel. The predicted SNR values from the output layer were compared against a SNR threshold of −8 dB to estimate the binary mask.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).
Brungart, D., Chang, P., Simpson, B., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.
Cooke, M., Green, P., Josifovski, L., and Vizinho, A. (**2001**). "Robust automatic speech recognition with missing and unreliable acoustic data," Speech Commun. **34**, 267–285.
Davis, S. B., and Mermelstein, P. (**1980**). "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**, 357–336.
Dempster, A. P., Laird, N. M., and Rubin, D. B. (**1977**). "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1–38.
Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**, 1109–1121.

Furui, S. (**1986**). "Speaker independent isolated word recognition using dynamic features of speech spectrum," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-34**, 52–59.
Hu, G., and Wang, D. L. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.
Hu, G., and Wang, D. L. (**2008**). "Segregation of unvoiced speech from nonspeech interference," J. Acoust. Soc. Am. **124**, 1306–1319.
Hu, Y., and Loizou, P. C. (**2007a**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777–1786.
Hu, Y., and Loizou, P. C. (**2007b**). "Subjective evaluation and comparison of speech enhancement algorithms," Speech Commun. **49**, 588–601.
Hu, Y., and Loizou, P. C. (**2008**). "Techniques for estimating the ideal binary mask," in The 11th International Workshop on Acoustic Echo and Noise Control, Seattle, WA
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.
Kollmeier, B., and Koch, R. (**1994**). "Speech enhancement based on physiological and psychoacoustical models of modulation perception and binaural interaction," J. Acoust. Soc. Am. **95**, 1593–1602.
Langner, G., and Schreiner, C. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**, 1799–1822.
Li, N., and Loizou, P. C. (**2008a**). "Effect of spectral resolution on the intelligibility of ideal binary masked speech," J. Acoust. Soc. Am. **123**, EL59–EL64.
Li, N., and Loizou, P. C. (**2008b**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.
Lim, J. S. (**1978**). "Evaluation of a correlation subtraction method for enhancing speech degraded by additive white noise," IEEE Trans. Acoust., Speech, Signal Process. **26**, 471–472.
Lippmann, R. P. (**1997**). "Speech recognition by machines and humans," Speech Commun. **22**, 1–15.
Loizou, P. C. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).
Macmillan, N., and Creelman, D. (**2005**). *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, New York).
Moore, B. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, London).
Nordqvist, P., and Leijon, A. (**2004**). "An efficient robust sound classification algorithm for hearing aids," J. Acoust. Soc. Am. **115**, 3033–3041.
Rabiner, L. (**2003**). "The power of speech," Science **301**, 1494–1495.
Reynolds, D., and Rose, R. (**1995**). "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Trans. Speech Audio Process. **3**, 72–83.
Reynolds, D., Quatieri, T., and Dunn, R. (**2000**). "Speaker verification using adapted Gaussian mixture models," Digit. Signal Process. **10**, 19–41.
Scalart, P., and Filho, J. (**1996**). "Speech enhancement based on a priori signal to noise estimation," in *Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing*, pp. 629–632.
Scharenborg, O. (**2007**). "Reaching over the gap: A review of efforts to link human and automatic speech recognition research," Speech Commun. **49**, 336–347.
Seltzer, M., Raj, B., and Stern, R. (**2004**). "A Bayesian classifier for spectrographic mask estimation for missing feature speech recognition," Speech Commun. **43**, 379–393.
Sroka, J. J., and Braida, L. D. (**2005**). "Human and machine consonant recognition," Speech Commun. **45**, 401–423.
Tchorz, J., and Kollmeier, B. (**2003**). "SNR estimation based on amplitude modulation analysis with applications to noise suppression," IEEE Trans. Speech Audio Process. **11**, 184–192.
Varga, A., and Steeneken, H. J. M. (**1993**). "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," Speech Commun. **12**, 247–251.
Wang, D. L., and Brown, G. J. (**2006**). *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications* (Wiley, Hoboken, NJ).
Zakis, J. A., Dillon, H., and McDermott, H. J. (**2007**). "The design and evaluation of a hearing aid with trainable amplification parameters," Ear Hear. **28**, 812–830.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.

Speech production modifications produced in the presence of low-pass and high-pass filtered noise

Youyi Lu[a)]
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, United Kingdom

Martin Cooke
Ikerbasque (Basque Science Foundation) and Language and Speech Laboratory, Facultad de Letras, Universidad del País Vasco, 01006 Vitoria, Spain

(Received 12 September 2008; revised 18 June 2009; accepted 18 June 2009)

In the presence of noise, do speakers actively shift their spectral energy distribution to regions least affected by the noise? The current study measured speech level, fundamental frequency, first formant frequency, and spectral center of gravity for read speech produced in the presence of low- and high-pass filtered noise. In both filtering conditions, these acoustic parameters increased relative to speech produced in quiet, a response which creates a release from masking for listeners in the low-pass condition but which actually increases masking in the high-pass noise condition. These results suggest that, at least for read speech, speakers do not adopt production strategies in noise which optimize listeners' information reception but that instead the observed shifts could be a passive response which creates a fortuitous masking release in the low-pass noise. Independent variation in parameters such as F0, F1 and spectral center of gravity may be severely constrained by the increase in vocal effort which accompanies Lombard speech.
 [DOI: 10.1121/1.3179668]

PACS number(s): 43.72.Ar, 43.72.Dv, 43.66.Dc [DOS] Pages: 1495–1499

I. INTRODUCTION

Speakers change the way they speak in the presence of noise (Lombard, 1911), causing, amongst others, an increase in speech level and fundamental frequency (F0), a flattening of spectral tilt (i.e., more energy at higher frequencies), and a tendency for an upward shift of F1 frequency. While the scale of changes in acoustic parameters observed in "Lombard" speech appears to be related to the relative level of the masker (Summers *et al.*, 1988; Tartter *et al.*, 1993), noise maskers with differing spectral shapes and temporal fluctuations have led to consistent changes in speech level, F0, and spectral tilt. For example, different studies employed white noise (Pisoni *et al.*, 1985; Summers *et al.*, 1988; Junqua, 1993), pink noise (Bond *et al.*, 1989; Hansen, 1996), traffic noise (Letowski *et al.*, 1993), multitalker babble (Garnier, 2007), and competing talkers (Lu and Cooke, 2008). One interpretation of the consistency with which various types of noise provoke speech production modifications is that the spectro-temporal properties of the noise may play little or no role in the Lombard effect. Under this view, speakers cannot, or do not, engage in active strategies which take into account the effect of noise at the ears of listeners.

However, other studies have raised the possibility that Lombard speech has an active component. Junqua *et al.* (1998) studied the influence of noise spectral tilt on Lombard speech, with a constant masker level of 85 dB sound pressure level (SPL). Speech level and F0 increased relative to a quiet background when talkers spoke with noise in the background in all conditions of spectral tilt, supporting the notion of a passive Lombard component. On the other hand, the size of the increase in speech level varied with noise spectral tilt. Mokbel (1992) recorded speech in the presence of white noise which was presented either low- or high-pass filtered or without filtering, at a fixed level. An increase in speech energy in frequency regions where the noise energy was most concentrated was observed, suggesting a dependency of the Lombard effect on the noise frequency distribution. However, Mokbel's study involved only one single speaker and did not report detailed changes in acoustic parameters, so it is difficult to appreciate the precise pattern as well as the reliability of the results, given that significant speaker-dependency of speech produced in noise has been observed (Summers *et al.*, 1988; Junqua 1993). However, the studies of Junqua *et al.* (1998) and Mokbel (1992) raise the intriguing possibility that the Lombard effect may have an active component which depends on the spectral characteristics of the background noise. In other words, talkers might use information gained by listening-while-talking to affect purposeful modifications to their speech, perhaps with the goal of improving intelligibility at the ears of the interlocutor.

Lu and Cooke (2008) investigated the effect of *N*-talker babble noise on speech production for *N* ranging from 1 (a single competing talker) to "infinity" (speech-shaped stationary noise), and taking in various multitalker babble conditions for intermediate values of *N*. Consistent with other Lombard studies, an overall shift in the center of gravity (CoG) of energy from lower to higher frequencies was observed at all values of *N*. Further, listeners found Lombard

a)Author to whom correspondence should be addressed. Electronic mail: y.lu@dcs.shef.ac.uk

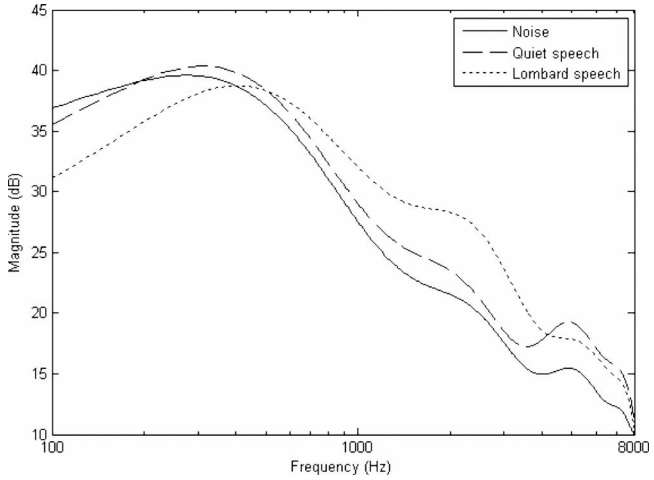


FIG. 1. Long-term average spectra of speech-shaped noise, and speech produced in quiet and noise in Lu and Cooke (2008). Note that the signals have normalized rms energy. A clear Lombard effect of energy shift to higher frequencies relative to quiet speech is visible.

speech substantially more intelligible than speech produced in quiet when both were presented in speech-shaped noise at the same signal-to-noise ratio. Since the long term spectrum of the noise was speech-shaped (for all N), an upward shift in CoG causes a degree of release from energetic masking (Fig. 1). Thus, the improvement in intelligibility could be fortuitous, since noise-induced speech changes may coincidentally be in the right direction to be advantageous for the speech-shaped noise maskers. An alternative possibility is that the observed shifts were caused by speakers making an active attempt to place spectral information in locations where it was less likely to be masked. The purpose of the current study was to distinguish these two possibilities.

In the present study, changes in speech production were measured in conditions of low-pass, high-pass, and full-band speech-shaped noise, relative to quiet. If speakers adopt an optimal strategy in order to minimize the effect of noise on listeners, they would be expected to shift their spectral CoG downwards for high-pass filtered noise condition compared to quiet, and in the opposite direction for low-pass noise condition. For each of the high- and low-pass conditions, two noise bandwidths were used to investigate the effect of varying the size of the noise-free part of the spectrum. Again, a "listener-optimal" speaking strategy should lead to greater changes for the smaller noise-free regions because the shift in speech spectral energy would need to be larger to reach the clean parts of the spectrum.

II. SPEECH CORPUS COLLECTION

A. Speech material

Speakers produced sentences defined by the Grid structure used in previous collections of normal (Cooke *et al.*, 2006) and Lombard speech (Lu and Cooke, 2008). Grid specifies simple six-word sentences such as "bin green at K 4 now" or "place red by E 7 please." While Grid sentences are not representative of natural tasks, they control for differences in speaking style and syntax, and the existence of many keyword repetitions allows for cross-condition comparisons of acoustic properties. Talkers produced an identical set of 30 Grid sentences in each of the conditions (see Sec. II B). To introduce some variation and remove any sentence dependency effect, each talker used a different sentence set.

B. Noise backgrounds

Speech was collected in quiet and in the presence of five noise backgrounds, one full-band, two high-pass filtered, and two low-pass filtered. The full-band noise had a spectrum equal to the long-term spectrum of utterances drawn from the 16 female and 18 male talkers of the Grid corpus (Cooke *et al.*, 2006), shown in Fig. 1. Low- and high-pass noise were derived from full-band noise using Chebyshev filter implementations with 0 dB pass-band gain and 60 dB stop-band attenuation, with frequency responses illustrated in Fig. 2. To investigate the effect of the size of the stop-band on speech production in noise, narrow- and wide-band versions of both high- and low-pass noise were generated using cutoff frequencies of 1 and 2 kHz. Note that in the low-pass conditions, the 1 kHz cutoff results in a narrow-band noise while in the high-pass condition the same cutoff leads to a wide-

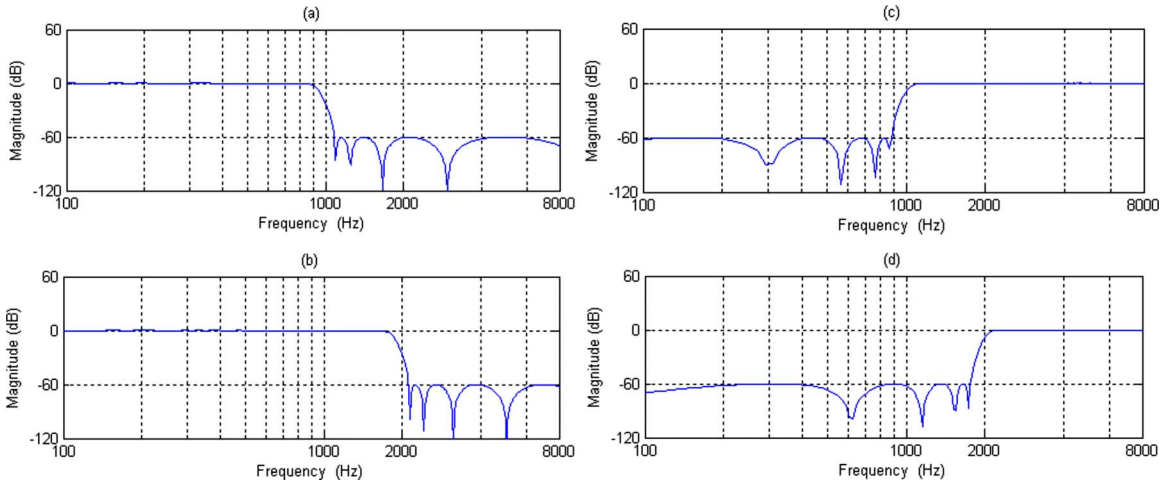


FIG. 2. (Color online) Frequency responses of the low- and high-pass digital filters. Panel (a) and (b) represent the low-pass filters with cutoff frequencies of 1 and 2 kHz respectively. Panel (c) and (d) represent the high-pass filters with cutoff frequencies of 1 and 2 kHz, respectively.

band noise, and vice versa for the 2 kHz cutoff. All maskers were normalized to 89 dB SPL prior to presentation, as measured with a Bruel & Kjaer (B & K) type 2603 sound level meter and B & K type 4153 artificial ear.

C. Talkers

Eight native speakers of British English (four males and four females) drawn from staff and students in the Department of Computer Science at the University of Sheffield participated in the corpus collection. All received a hearing test using a calibrated software audiometer which was used to test each ear separately at the six frequencies: 250, 500, 1000, 2000, 4000, and 8000 Hz. All participants had normal hearing (better than 20 dB hearing level in the range of 250–8000 Hz). Ages ranged from 24 to 48 years (mean: 29.8 years). Ethics permission was obtained following the University of Sheffield Ethics Procedure. Talkers were paid for their participation.

D. Procedure

Corpus collection sessions took place in an industrial acoustics company single-walled acoustically-isolated booth. Speech material was collected using a B & K type 4190 $\frac{1}{2}$ in. microphone coupled with a preamplifier (B & K type 2669) placed 30 cm in front of the talker. The signal was further processed by a conditioning amplifier (B & K Nexus model 2690) prior to digitisation at 25 kHz with a Tucker-Davis Technologies (TDT) RP2.1 system. Simultaneously, maskers were presented diotically over Sennheiser HD 250 Linear II headphones using the TDT system. Talkers wore the headphones throughout, including for the quiet condition. In order to compensate for sound attenuation introduced by the closed ear headphones, the talkers' own voice was fed back via the TDT system and mixed with the noise signal prior to presentation over the headphones. At the beginning of the recording session, each talker was asked to speak freely into the microphone while wearing the headphones. The level of voice feedback was manually adjusted until the talker felt that the overall loudness level matched that when not wearing headphones. Voice feedback level was then held constant for all the recording conditions and talkers were unable to adjust the level.

Sentence collection and masker presentation was under computer control. Talkers were asked to read out sentences presented on a computer screen and had 3 s to produce each sentence. They were allowed to repeat the sentence if they felt it necessary, with the final repetition used for further analysis. In practice, talkers made only a few repetitions in any single condition with maximum of 4 out of 30 sentences and a mean of less than 2. Across-talker means of repetition in the six conditions were not statistically different [$F(1,7)=0.86$, $p=0.44$]. Maskers were gated with the 3 s recording time. Condition and sentence orders within each condition were randomized. Talkers recorded all the six conditions (i.e., five noise conditions plus quiet) in one session of approximately 20 min.

E. Postprocessing

In order to identify and remove leading and trailing silent intervals of the collected sentences, a set of speaker-independent phoneme-level hidden Markov models was built from speech material in the Grid corpus using the HTK toolkit (Young *et al.*, 1999). These models were used to produce phoneme-level transcriptions of the collected utterances via forced alignment using the HVITE tool in HTK. The leading and trailing silent intervals identified via the alignment process were removed. Transcriptions of the leading and trailing silent intervals for all the utterances were manually inspected and found to be accurate within approximately 15 ms relative to human judgements.

III. ACOUSTIC MEASUREMENTS AND STATISTICAL ANALYSIS

Four acoustic properties were estimated for each utterance. Root mean square (rms) energy, mean fundamental frequency (F0), spectral CoG, and mean first formant (F1) frequency were computed via PRAAT 4.3.24 (Boersma and Weenink, 2005). F0 estimates were provided at 10 ms intervals using an autocorrelation-based method (Boersma, 1993) implemented in the PRAAT program. Spectral CoG was computed on the spectrum of an entire utterance by averaging the frequency spectrum weighted by its power magnitude. Mean F1 frequency was obtained by averaging F1 values estimated for voiced frames using the BURG algorithm (Burg, 1975) implemented in PRAAT. These parameters were selected since reliable changes in these properties have been reported in earlier Lombard studies, and, apart from rms energy, all these properties cue the location of spectral information, which allows the pattern of shifts in spectral energy distribution to be determined.

Across-talker means in quiet, speech-shaped noise and filtered noise conditions for each of the acoustic parameters are shown in Fig. 3. For all parameters and in both low- and high-pass conditions, noise resulted in increases in all parameters. In the low-pass case, little difference between the two filtered and full-band noises is visible, while for high-pass noise, filtered noise tended to result in smaller increases than in the full-band condition. While some variability among the individual talkers was present, similar patterns in each of the acoustic parameters and across backgrounds were observed.

Due to the likelihood of moderate correlations between acoustic parameters such as speech level and both F0 and F1 frequency (Alku *et al.*, 2002; Garnier 2007), multivariate analysis of variance (MANOVA) was used to examine the effect of noise background. Separate MANOVAs were computed for the low- and high-pass cases, with rms energy, F0, F1, and CoG as dependent variables. Initially, MANOVAs with one within-subject factor representing four types of background (quiet, narrow, wide, full) and one between-subject factor (gender) revealed that while gender differences were observed for F0 and F1, the pattern of results was the same for the male and female talkers since no significant interaction was found between gender and background type

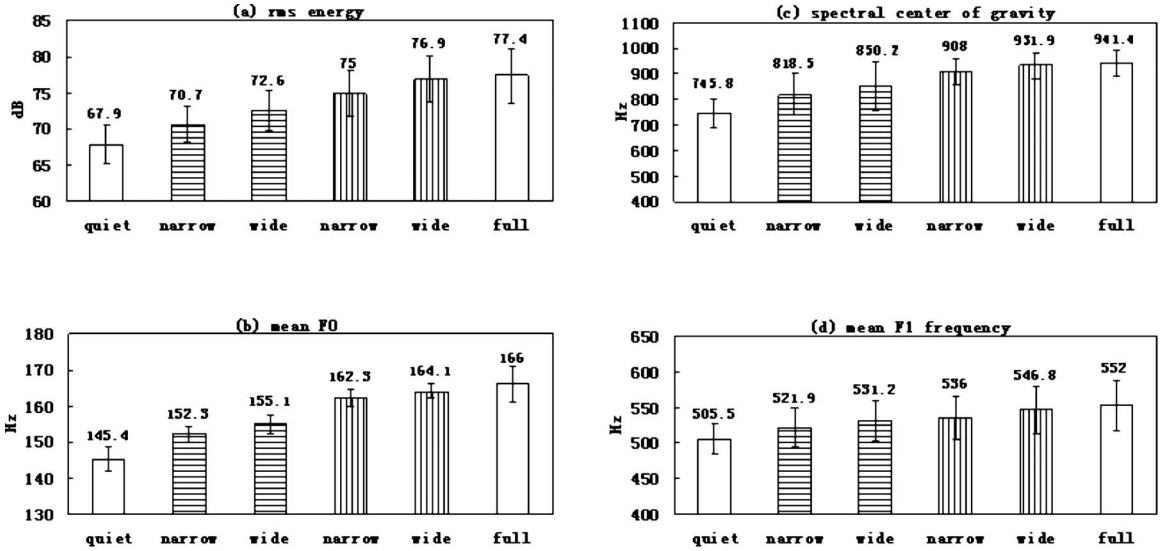


FIG. 3. Acoustic parameter values for quiet, two high-pass noise conditions (shaded bars with horizontal lines) with cutoff frequencies at 2 kHz ("narrow" bandwidth) and 1 kHz ("wide" bandwidth), two low-pass noise conditions (shaded bars with vertical lines) with cutoff frequencies at 1 kHz (narrow bandwidth) and 2 kHz (wide bandwidth), and speech-shaped noise condition ("full" bandwidth). Values shown are means over talkers and error bars indicate 95% confidence intervals.

($p > 0.05$). In order to increase statistical power with the limited number of speakers used in the current study, data for male and female talkers were combined.

For the low-pass case, there was a significant multivariate effect of differences between the four backgrounds {quiet, two low-pass noise, speech-shaped noise} [$F(12,47.9) = 9.37$, $p < 0.001$, $\eta^2 = 0.66$], as well as for the four parameters individually [$F(1.23,8.62) = 49.15$, $p < 0.001$, $\eta^2 = 0.88$ for rms energy; $F(1.38,9.65) = 27.66$, $p < 0.001$, $\eta^2 = 0.80$ for mean F0; $F(1.24,8.67) = 21.87$, $p < 0.01$, $\eta^2 = 0.76$ for CoG; $F(2.05,14.37) = 97.64$, $p < 0.001$, $\eta^2 = 0.93$ for mean F1 frequency]. *Post hoc* pairwise comparisons (here and elsewhere by paired *t*-tests with Bonferroni-adjustment) showed that the quiet condition was significantly different from the rest ($p < 0.01$) for all four parameters. None of the differences between the three noise conditions was statistically significant.

As expected, given the difference between the quiet and full-band conditions, for the high-pass case, the multivariate effect of background type {quiet, two high-pass noise, speech-shaped noise} was also significant [$F(12,47.9) = 5.99$, $p < 0.001$, $\eta^2 = 0.55$]. Of more interest is the confirmation by *post hoc* pairwise comparisons that the high-pass conditions resulted in significant increases in all parameters relative to quiet ($p < 0.05$), and, unlike in the low-pass case, increases were significantly smaller than the full-band condition ($p < 0.05$) apart from the wide-band/full-band comparison for F1 ($p = 0.06$). The tendency, visible in Fig. 3, for the wide-band high-pass noise to provoke larger parameter excursions than the narrow-band high-pass condition was not statistically significant except in the case of rms energy ($p < 0.05$).

IV. DISCUSSION

The current study extends to both low- and high-pass filtered noise backgrounds the finding that talkers modify their productions when exposed to full-band noise. The low-pass conditions resulted in increases in F0 and F1 frequencies, and spectral CoG. While these results are consistent with the hypothesis that speakers were actively avoiding the presence of noise whose spectrum was concentrated at low frequencies, two findings suggest otherwise. First, the full-band and low-pass filtered noise provoked statistically-identical increases in these parameters. One might expect to see a larger amount of shift in the low-pass condition to take advantage of the noise-free part of the spectrum relative to the full-band case. Second, there was no difference between the narrow- and wide-band low-pass conditions, where an active strategy would predict larger increases in the presence of wide-band low-pass noise in order to place spectral energy in the noise-free region.

High-pass filtering conditions also led to clear increases in F0, F1 and spectral CoG, suggesting that speakers are unable to adopt the speaking strategy of adapting speech production to place information-bearing elements of speech in regions devoid of noise. Further, speakers reacted similarly to the wide- and narrow-band conditions, where optimality would suggest that a smaller noise-free spectral region would lead to differential shifts in acoustic parameters. The absence of the "optimal" response to high-pass noise may be attributed to articulatory side-effects of an increase in vocal effort, which was observed in all noise backgrounds. For example, the rise in subglottal pressure needed to increase vocal output leads to an increase in F0 (Schulman, 1985; Gramming *et al.* 1988), and the wider jaw opening in order to increase sound amplitude induces an increase in F1 frequency (Stevens, 2000; Huber and Chandrasekaran, 2006). Thus, the scope for active control of F0 and F1 frequencies might be limited by the stronger desire to increase output level in response to noise.

One surprising aspect of the current study is the fact that noise bandlimited to the region below 1 kHz produced an

equivalent Lombard effect as full-band noise. This might result from the upward spread of masking into higher frequencies produced by the 1 kHz low-pass noise, a phenomenon first reported by Egan and Hake (1950). In addition, since all noises employed were presented at the same level, the little difference of Lombard effect between the low-pass filtered and full-band noise conditions appears to support the studies cited in the Introduction which argued that noise level is the dominant component of the Lombard effect. However, the high-pass filtered noise conditions led to a significantly smaller increase in parameters such as rms energy (2.8 and 4.7 dB compared to 7.1 and 9 dB in the low-pass conditions, a difference which probably also accounts for the lower scale of increases in other acoustic parameters given the articulatory constraints discussed above), suggesting that noise level is not the only factor in the Lombard effect. It is possible that the difference in response to high- and low-pass noise reflects the relative importance that these frequency regions have in speech perception or in own-voice monitoring. F0 information is more clearly masked in the low-pass conditions, for instance.

Overall, these findings do not support the idea of an active response to noise. However, there are several aspects of the current task which may have limited the scope or motivation on the part of talkers to exploit noise-free spectral regions. First, noise was gated on and off to coincide with the 3 s recording period. It is possible that speakers were not exposed to noise for long enough to learn about the potential benefit of re-allocating spectral energy. Second, the task for talkers did not involve communication of information, so the notion that talkers were motivated to make things easier for a listener is suspected. Further studies involving communicative tasks and continuous noise backgrounds may lead to different results. Finally, the observed change in speech level produced by noise may act to mask the effect of noise on other parameters. Experiments designed to inhibit the change in vocal effort (e.g., Pick *et al.*, 1989) may provide a more sensitive measure of differential response to the spectral content of the background.

V. CONCLUSION

An effective speaking strategy for the maintenance of intelligibility in noise would be to place information in those spectral regions least affected by the noise. However, the current study found little evidence that speakers were able to modify their speech productions in this way to take advantage of noise-free regions. In the presence of high-pass noise, speech parameters such as F0 and F1 frequencies, and spectral CoG did not shift downwards but instead increased relative to speaking in quiet conditions. One explanation for this result is that the increase in vocal effort caused by noise limited the scope for variability of other speech parameters such as fundamental frequency. However, there remains the possibility that under more realistic communicative conditions, speakers may adopt active strategies to reduce the effect of noise for listeners.

Alku, P., Vintturi, J., and Vilkman, E. (**2002**). "Measuring the effect of fundamental frequency raising as a strategy for increasing vocal intensity in soft, normal and loud phonation," Speech Commun. **38**, 321–334.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a samples sound," Proc. Inst. Phonetic Sci. **17**, 97–110.

Boersma, P., and Weenink, D. (**2005**). "Praat: Doing phonetics by computer (version 4.3.14) (computer program)," from http://www.praat.org. (Last viewed May, 2005).

Bond, Z., Moore, T., and Gable, B. (**1989**). "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912.

Burg, J. P. (**1975**). "Maximum entropy spectrum analysis," Ph.D. thesis, Stanford University, Palo Atto, CA.

Cooke, M. P., Barker, J., Cunningham, S., and Shao, X. (**2006**). "An audio-visual corpus for speech perception and automatic speech recognition," J. Acoust. Soc. Am. **120**, 2421–2424.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., and Perkins, W. (**1988**). "Relationship between changes in voice pitch and loudness," J. Voice **2**, 118–126.

Garnier, M. (**2007**). "Communiquer en environnement bruyant: de l'adaptation jusqu'au forçage vocal [Communication in noisy environments: From adaptation to vocal straining]," These de Doctorat de l'Universite Paris 6.

Hansen, J. H. L. (**1996**). "Analysis and compensation of speech under stress and noise for environmental robustness in speech recognition," Speech Commun. **20**, 151–170.

Huber, J. E., and Chandrasekaran, B. (**2006**). "Effects of increasing sound pressure level on lip and jaw movement parameters and consistency in young adults," J. Speech Lang. Hear. Res. **49**, 1368–1379.

Junqua, J. C. (**1993**). "The Lombard reflex and its role on human listeners and automatic speech recognizers," J. Acoust. Soc. Am. **93**, 510–524.

Junqua, J. C., Fincke, S., and Field, K. (**1998**). "Influence of the speaking style and the noise spectral tilt on the Lombard reflex and automatic speech recognition," in International Conference Spoken Language Proceedings, pp. 467–470.

Letowski, T., Frank, T., and Caravella, J. (**1993**). "Acoustical properties of speech produced in noise presented through supra-aural earphones," Ear Hear. **14**, 332–338.

Lombard, E. (**1911**). "Le signe de l'elevation de la voix (The sign of the rise in the voice)," Ann. Maladiers Oreille, Larynx, Nez, Pharynx **37**, 101–119.

Lu, Y., and Cooke, M. P. (**2008**). "Speech production modifications produced by competing talkers, babble and stationary noise," J. Acoust. Soc. Am. **124**, 3261–3275.

Mokbel, C. (**1992**). "Reconnaissance de la parole dans le bruit: Bruitage/debruitage [Voice recognition in noisy environments: Sound/denoising]," Ph.D. thesis, Ecole Nationale Superieure des Telecommunications, Paris.

Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.

Pisoni, D. B., Bernacki, R. H., Nusbaum, H. C., and Yuchtman, M. (**1985**). "Some acoustic-phonetic correlates of speech produced in noise," in International Conference on Acoustics Speech and Signal Processing, pp. 1581–1584.

Schulman, R. (**1985**). "Dynamic and perceptual constraints of loud speech," J. Acoust. Soc. Am. **178**, S37.

Stevens, K. N. (**2000**). *Acoustic Phonetics (Current Studies in Linguistics)* (MIT, Cambridge, MA).

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. (**1988**). "Effects of noise on speech production: Acoustic and perceptual analysis," J. Acoust. Soc. Am. **84**, 917–928.

Tartter, V. C., Gomes, H., and Litwin, E. (**1993**). "Some acoustic effects of listening to noise on speech production," J. Acoust. Soc. Am. **94**, 2437–2440.

Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V., and Woodland, P. (**1999**). *The HTK Book 2.2*, (Entropic, Cambridge).

Characteristics of speaking style and implications for speech recognition

Takahiro Shinozaki[a)]
Department of Computer Science, Tokyo Institute of Technology, Tokyo 152-8552, Japan

Mari Ostendorf and Les Atlas
Department of Electrical Engineering, University of Washington, Seattle, Washington 98195-2500

(Received 25 August 2008; revised 5 April 2009; accepted 30 June 2009)

Differences in speaking style are associated with more or less spectral variability, as well as different modulation characteristics. The greater variation in some styles (e.g., spontaneous speech and infant-directed speech) poses challenges for recognition but possibly also opportunities for learning more robust models, as evidenced by prior work and motivated by child language acquisition studies. In order to investigate this possibility, this work proposes a new method for characterizing speaking style (the modulation spectrum), examines spontaneous, read, adult-directed, and infant-directed styles in this space, and conducts pilot experiments in style detection and sampling for improved speech recognizer training. Speaking style classification is improved by using the modulation spectrum in combination with standard pitch and energy variation. Speech recognition experiments on a small vocabulary conversational speech recognition task show that sampling methods for training with a small amount of data benefit from the new features.
 [DOI: 10.1121/1.3183593]

PACS number(s): 43.72.Ar, 43.72.Ne [DOS] Pages: 1500–1510

I. INTRODUCTION

It is well known that spoken language varies with different situations, including the formality or informality of the situation, the familiarity of speakers with their conversational partners and relative seniority, whether or not the listener is a language learner, the noise level of the environment, etc. Both the word choices and the speaking style can vary, where by speaking style the authors mean the quality of articulation as well as prosodic characteristics, including intonation, timing, and energy patterns associated with emphasis and phrasing. Both have an effect on speech recognition performance, but in this paper the authors will focus on acoustic characteristics of speaking style, particularly in terms of situational context.

Speaking style reflects, in part, the speaker's effort to be understood. For example, a news announcer or someone giving a speech will tend to articulate more clearly than someone engaged in casual conversation, and people hyperarticulate in a situation where they think they have been misunderstood. Hyperarticulation also appears in language learning contexts, both in speech of second-language teachers and in adults talking to children. Studies on infant-directed speech suggest that its prosodic features attract and hold infant attention and that the phonetic cues are exaggerated and more acoustically distinct,[1,2] though another study finds that the vowel space is expanded only for pitch-accented words.[3] Further, researchers have found a correlation between the clarity of mothers' speech and infants' discrimination capabilities.[4] Overemphasized phonetic contrasts also appear to be useful in second-language learning.[5]

In terms of recognition accuracy, human listeners are relatively insensitive to the change in the speaking styles as they do not experience special difficulties in listening whether it is read or conversational speech. In fact, the word error rate (WER) by human listeners for the switchboard conversational telephone speech (CTS) corpus was 4%,[6] which was not very different from the error rates of 2.6% for the read utterances in the Wall Street Journal corpus.[7] On the other hand, the recognition performance of automatic speech recognition (ASR) systems is significantly affected by the difference in the speaking styles, and the error rates often become one or more orders of magnitude higher than that of human.[8]

When a recognizer that has not been trained with hyperarticulated speech has to recognize it, performance degrades. However, even in matched train/test conditions, style impacts performance. ASR systems designed for conversational speech typically perform much worse than similar systems trained and tested on news recordings, even though the conversational speech task is "simpler" in the sense of language models having lower perplexity. In a study by researchers at SRI,[9] the language model is factored out by collecting spontaneous conversational speech and then having the same speakers come back and read their transcripts. The read speech had a lower recognition error, even though the words spoken were the same. These findings were confirmed in a subsequent study on pronunciation modeling using the same data.[10] Studies using the same recognition technology on different genres show that broadcast news tends to be easier to recognize than conversational speech genres (talk shows, telephone conversations, and meetings) and that even within

[a)]This work was carried out when the author was at the Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500.

news broadcasts, professional announcers tend to be associated with lower WERs. Another contrast in speaking styles is adult-directed vs infant-directed speech. Analogous to the above results for spontaneous speech, Kirchhoff and Schimmel found that infant-directed speech has a higher recognition error rate than adult-directed speech in matched training conditions.[3] In these studies, both conversational and infant-directed speech are shown to have more variability than their counterpart in terms of the spectral realization of phonemes. In addition, they are more dynamic prosodically in terms of fundamental frequency (F0) and speaking rate variation, which may be helpful to human listeners.

While variability is problematic for recognition, it can be useful for robust training, i.e., for cases where the ASR system may need to recognize speech in a style that it was not trained on. In an experiment of using Japanese Spontaneous speech corpus[11] and Japanese newspaper article sentences corpus,[12] the authors observed that training on the spontaneous speech and testing on the read speech gave similar performances to the matched training condition for the read speech, but the reverse led to significant degradation in performance. More precisely, the experiments were performed using spontaneous and read speech models trained, respectively, from 52 h of gender balanced training data from the corpora and using standard test sets associated with the corpora. The WER for the read speech by the spontaneous speech model was 9.5%, and this was similar to 8.7% by the read speech model. On the other hand, the WER for the spontaneous speech by the read speech model was 38.2%, which was significantly higher than 25.0% by the spontaneous speech model. Similarly, the mismatched train/test condition for adult- and infant-directed speech has greater degradation in performance relative to the matched condition for the case using adult-directed training compared to using infant-directed training.

One of the possible hypotheses for this is that the more careful types of speech lead to models with tighter variances, which are less able to handle cases in the overlap regions associated with less careful speech. Variability in training is leveraged even for matched training conditions in the sense that it has been proposed to put a greater weight on "difficult cases," either through sampling[13] or boosting.[14] However, many studies of human language acquisition suggest that infant-directed speech might be useful in providing better prototypes for different speech sounds, assuming that children are focusing on the emphasized examples. These suggest very different methods for sampling speech in learning: bringing in hyperarticulated examples as outliers later in training vs initializing with exaggerated examples.

The prior work thus suggests two possible reasons for recognizing speech style: detecting different styles in order to adjust the recognition models and selecting or weighting speech for training. As a first step in exploring these ideas, this paper proposes a new method for characterizing speaking style, examines spontaneous, read, adult-directed, and infant-directed styles in this space, and conducts pilot experiments in style detection and sampling for improved ASR training. In particular, the authors propose the use of the modulation spectrum to characterize the acoustics of speaking style, with the idea of representing the greater variation that they anecdotally perceive in spontaneous (vs read) and infant-directed (vs adult-directed) speech.

In the sections that follow, the authors begin by motivating the modulation approach to speech analysis and introduce the basic mathematical framework in Sec. II. In Sec. III, they provide analyses of speaking styles in terms of acoustic dynamics using the modulation spectrum as well as traditional F0 and energy measures. Section IV presents results of style recognition experiments with some of these features. In Sec. V, several sampling methods and training strategies are investigated for ASR. Finally, a summary and conclusions are given in Sec. VI.

II. MODULATION ANALYSIS OF SPEECH

There is substantial evidence that many natural signals can be represented as low frequency modulators, which modulate higher frequency carriers. Many researchers have observed that this concept, loosely called "modulation frequency," is useful for describing, representing, and modifying broadband acoustic signals such as speech and music. Modulation frequency representations usually consist of a transform of a one-dimensional broadband signal into a two-dimensional joint frequency representation, where one dimension is a standard Fourier frequency and the other dimension is a modulation frequency.

In 1939, Dudley concluded his now famous paper on speech analysis[15] with

> … the basic nature of speech as composed of audible sound streams on which the intelligence content is impressed of the true message-bearing waves which, however, by themselves are inaudible.

In other words, he observed that speech and other audio signals, such as music, are actually low bandwidth processes that modulate higher bandwidth carriers. Over the years, research in auditory science has supported this idea, including findings that two-dimensional spectro-temporal modulation transfer functions can model many of the observed effects of auditory sensitivity to amplitude modulation[16] and that frequency and modulation periodicity are represented via orthogonal maps in the human auditory cortex.[17]

In signal processing, the modulation spectrum is a representation of speech that gives both acoustic and modulation frequency information.[18] In its simplest form, the modulation spectrum can be considered to be a Fourier transform (in time) of each row of the magnitude of the short-time Fourier transform (STFT) or the magnitude spectrogram. In general, modulation spectral analysis involves a base transform on short-term windows of speech, followed by a nonlinear detection operation, and then a second transform.

In the specific implementation used here, the modulation spectrum is obtained by first generating a STFT vector sequence, taking the magnitude of the result for each frequency bin, and then applying a second Fourier transform magnitude to each time series corresponding to a frequency bin. The result is a two-dimensional matrix with frequency and modulation axes. The parameters for modulation spectral analysis consists of those for the base STFT [e.g., window, overlap,

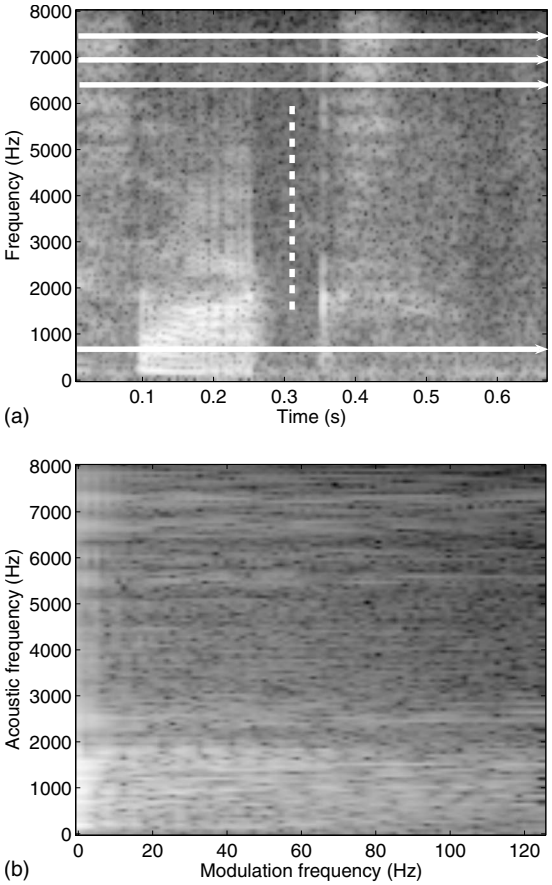


FIG. 1. A spectrogram of the word "socks" and its modulation spectrum. Modulation spectrum is obtained from a sequence of STFT vectors.

and fast Fourier transform (FFT) order], the length of the STFT sequence, and the parameters of the second Fourier transform.

Figure 1 shows an example of a spectrogram and its modulation spectrum, where the base analysis window length and overlap were 16 ms and 75%, respectively, and the second modulation window size was equal to the length of the spectrogram. The speech segment is an instance of an adult female pronouncing the word "socks" sampled at 16 kHz. As the figure illustrates, the second Fourier transform was performed on time sequences of subband energies, e.g., along the arrows overlaid on the spectrogram, to obtain the modulation spectrum.

Note that there are other versions of modulation spectral analysis and filtering that use either Hilbert transform[19] or coherent and distortion-free methods[20] to modify the modulation spectrum and then produce a new signal with filtered modulations. For this paper, the authors focus only on an energetic interpretation of the modulation spectrum, where, much as with a standard power spectral density estimate, there is no intent to modify the modulation content of a signal. The authors also use only the magnitude after the second Fourier transform—the modulation spectrum magnitude—leaving the phase of the modulation spectrum, which is also known to potentially have importance,[21] for future studies.

III. ACOUSTIC ANALYSIS

In this section, acoustic analyses on speaking styles are performed based on features extracted from utterances. Two different corpora are used, as described next, in order to have a variety of styles and to investigate analogies between the read/spontaneous and adult-directed/infant-directed contrasts.

A. Corpora

The Multi-Register speech corpus (MULTI-REG), which was collected at SRI, includes spontaneous speech and a read version of its transcription pronounced in a dictation manner.[9,10] The speech was recorded over telephone and high quality head mounted microphone channels. The telephone channel data were stored in 8 kHz u-law format, and the high quality channel was recorded with 16 kHz pulse code modulation (PCM) data. In the following experiments, a subset of the corpus was used that has consistent transcription across the speech types. It consisted of nine female speakers with 557 speech segments, where the authors restrict the analysis to female speech to match the second corpus which has only female speakers. Due to this constraint, the findings in this paper are biased to female speakers. Compared to male voice, the most prominent characteristics of female voice is higher fundamental frequency. It makes it, for example, difficult to accurately estimate formant frequency due to wider spacing of pitch harmonics.[22]

The Motherese corpus has infant-directed and adult-directed utterances provided by the Institute for Learning and Brain Sciences at the University of Washington. The infant-directed and adult-directed utterances are extracted from conversations that a mother has with her infant and with an adult experimenter, respectively. The authors' work used a subset of the data taken from the set used in the Kirchhoff–Schimmel study;[3] further details about the data are included therein. Specifically, 12k utterances (6.9k infant-directed utterances and 5.5k adult-directed utterances) from 32 female speakers were used. The data were designed to elicit keywords from the mothers, but the authors' analyses used complete utterances rather than just these keywords. The speech in this corpus was recorded with 16 kHz sampling and 16 bit PCM format using a far-field microphone. In the following experiments, wave forms with 8 kHz sampling frequency were made by down-sampling the original 16 kHz version with high cutoff frequency of 3.8 kHz.

B. F0 analysis

F0 contours were first estimated for each 10 ms frame using the getf0 command[23] from the ESPS package. Then, to reduce estimation error, a mixture model is used to characterize doubling and halving so as to more accurately determine F0, and the contour was stylized using the GRAPHTRACK program.[24] It is typical to normalize F0 to account for speaker variability, but it is important to use the same normalization factors for both styles recorded for a speaker. For the experiments here, the authors chose frame-wise mean and standard deviation as the normalization factors based on spontaneous or adult-directed utterances depending on the

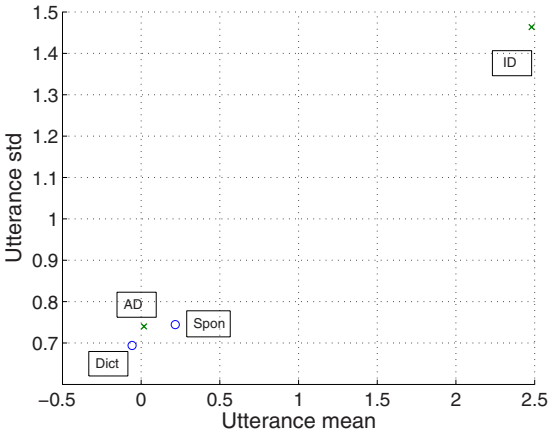


FIG. 2. (Color online) Normalized F0 statistics estimated for the MULTI-REG corpus ("Dict" and "Spon" are dictation and spontaneous speech, respectively) and the motherese corpus ("ID" and "AD" are infant-directed and adult-directed speech, respectively).

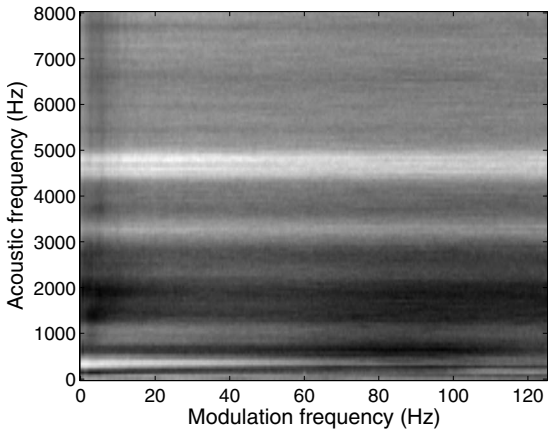


FIG. 3. Difference of averaged modulation spectrum of infant-directed utterances from adult-directed utterances.

corpora. The F0 features of both genres from a speaker were normalized by subtracting the mean and dividing by the standard deviation.

After the normalization, the mean and standard deviation within a segment were used as the features of that segment, excluding frames in unvoiced regions. Figure 2 shows F0 mean and standard deviation for each of the speech types that are averaged over the speakers in the corpus. It is observed that infant-directed utterances have a much higher F0 mean and variance than all other conditions, as expected. The differences between the other three cases are small in comparison.

C. Modulation spectrum analysis

As described in Sec. II, the modulation spectrum is obtained by applying the Fourier transform to a slice over time of the STFT, resulting in a two-dimensional matrix with acoustic frequency and modulation frequency axes. The dimensions of the matrix are the number of acoustic frequency bins (half the size of the first Fourier transform) and the number of modulation frequency bins (half the number of time frames in the second Fourier transform). In the studies presented here, the STFT base window width and overlap were 16 ms and 75%, respectively, and the FFT size was 128 or 256 depending on 8 and 16 kHz sampling rates, respectively. The modulation window size was equal to the length of the input STFT sequence, which varies with the signal being analyzed. In Figs. 3 and 4, for example, the sampling rate was 16 kHz and the length was 2.0 s, so the resulting modulation spectrum is a 128×256 array.

Figure 3 shows the difference of the modulation spectra for infant-directed and adult-directed speech. The figure was made by subtracting the log of averaged magnitude modulation spectrum of adult-directed speech from that of infant-directed speech using 1500 segments for each of the speech type. Similarly, Fig. 4 shows the difference of averaged modulation spectrum of the read and spontaneous speech given by female speakers from the MULTI-REG speech corpus using 700 segments for each of the speech type. The analysis indicates that both infant-directed and read speech have more energy than their counterparts at higher modulation frequencies for high acoustic frequencies, which the authors hypothesize to be due to a tendency to have more clearly articulated consonants in these genres. The analysis indicates that infant-directed speech tends to have more energy at low modulation frequencies in the low formant regions, particularly in the region of F0. This phenomenon is explored further in Sec. III D. In addition, for the read vs spontaneous contrast, a difference is observed at high modulation frequencies in the low acoustic frequency region. The authors have as yet no explanation for this difference.

D. Spectrogram analysis

To better understand the modulation spectrum differences for the adult- and infant-directed speech, the authors inspected several of the target words (covering the vowels /iy/, /uw/, and /aa/). They found that the mother's fundamental frequency frequently aligns with the first formant of the vowel in infant-directed speech, as illustrated in Figs. 5 and 6, which show spectrograms and a spectral slice for adult-directed and infant-directed speech segments. All spectrograms were made from speech segments with 8 kHz sampling, and the transformation window width was 32 ms. The vertical line in the spectrogram corresponds to the time po-

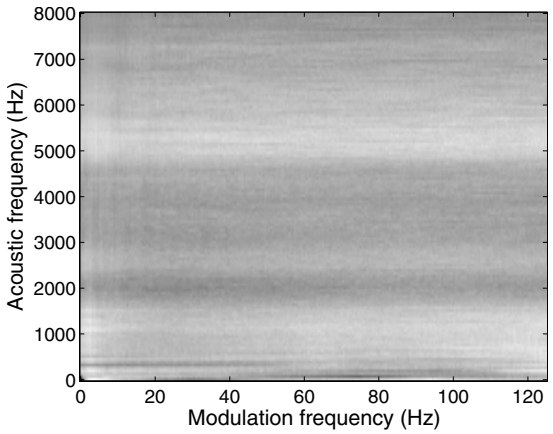


FIG. 4. Difference of averaged modulation spectrum of dictation utterances from spontaneous utterances.

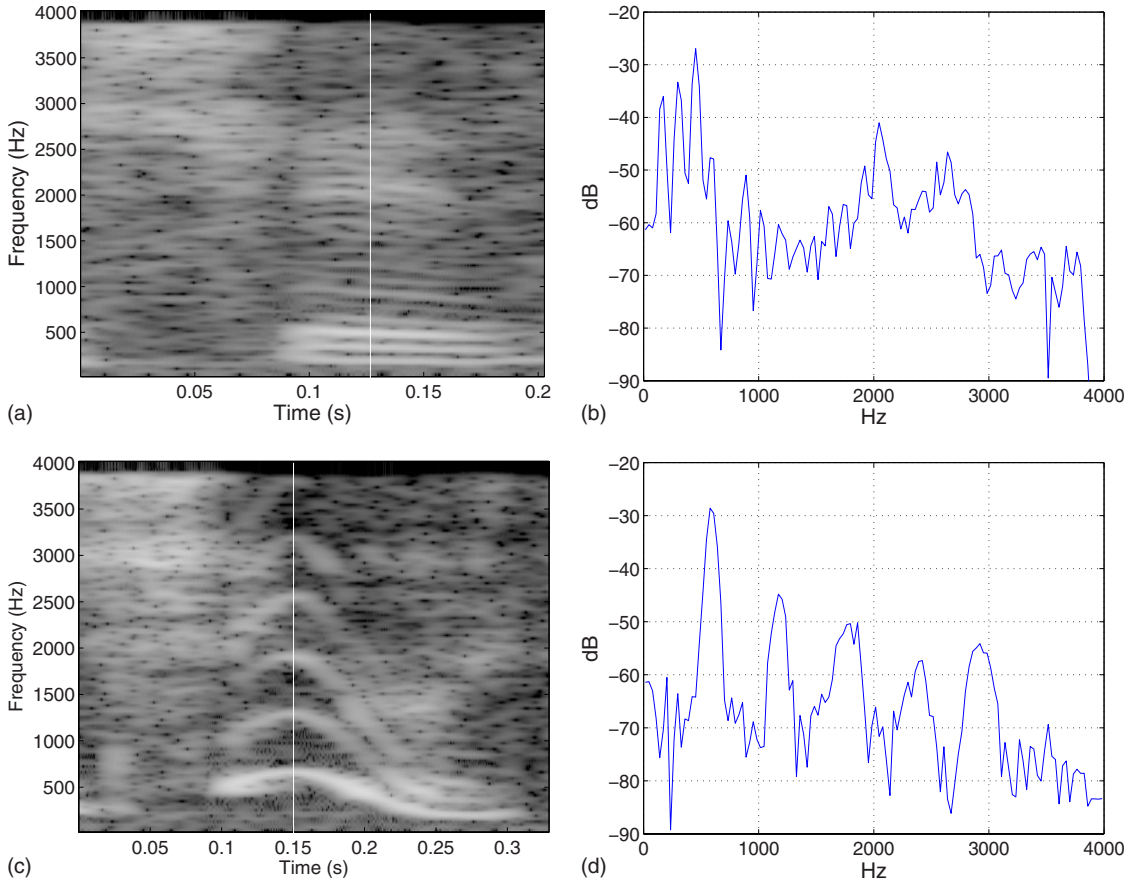


FIG. 5. (Color online) Spectrograms and their sections of "shoes" sound. In infant-directed /uw/ sound, fundamental frequency aligns with the first formant.

sition of the spectrum section. The peaks of the spectrum envelope are the formants caused by acoustic resonance of the vocal tract. It is known that most of the discriminative information between vowels is encoded in the lowest two formants F1 and F2. The finer peaks of the spectrogram correspond to the vibration of vocal folds whose fundamental frequency is referred to as F0.

Usually, the formant frequencies are much higher than F0, as can be seen in the adult-directed speech in Figs. 5(b) and 6(b). However, the authors found that F0 takes approximately the same frequency as F1 in highly exaggerated infant-directed utterances, as can be observed in Figs. 5(d) and 6(d). This phenomenon may be associated with a mother trying to draw her infant's attention to something. The meaning of the pitch-formant matching for infants is not known. Perhaps, it helps infants to learn how to discriminate vowels by giving simpler examples of the spectrum shapes. The authors conjecture that this phenomenon might be peculiar to female speakers because speaking F0 of adult male speakers is usually much lower than F1,[25] though the answer is not known yet since the Motherese corpus does not include male speakers.

E. MDS analysis

Multidimensional scaling (MDS) is a technique to arrange data points in a space that has lower dimension than the original data space.[26] The arrangement is determined so as to keep the distances between the data points as close as possible to the original space. In the MDS analysis, only the distances between points have meanings. Therefore, transformations that do not change distance, such as rotation, give an equivalent disposition in terms of MDS analysis, and the axes are arbitrary.

The MDS analysis was performed on a distance matrix of speakers defined by Euclidean distances of modulation-spectrum-based feature vectors. The feature vectors were made by first computing modulation spectrum for each utterance from the 8 kHz sampled wave form, reducing it to 5 $\times 5$ matrix, and re-ordering the matrix to form a 25-dimensional vector. For the reduction operation, element (i,j) in the original modulation spectrum matrix was assigned to block $(\lfloor 5(i-1)/128 \rfloor + 1, \lfloor 5(j-1)/128 \rfloor + 1)$, and an average in the block was used as the element of the new matrix. The utterance-level feature vectors were then averaged to make a speaker-level feature vector, and Euclidean distances were computed for all the pairs of speakers, both within and across corpora. Instances of the same speaker in different genres are included as well as cross-speaker pairs.

In the modulation spectrum estimation, the authors introduced a channel normalization to compensate for corpus-level recording effects, so that a comparison across corpora made sense. The proposed normalization algorithm works in the complex spectral domain as follows:

$$\hat{S}_i(\omega) = \frac{Y_i(\omega)}{C(\omega)} \approx \frac{Y_i(\omega)}{\exp(\langle \log(Y_i(\omega)) \rangle_1^n)}, \quad (1)$$

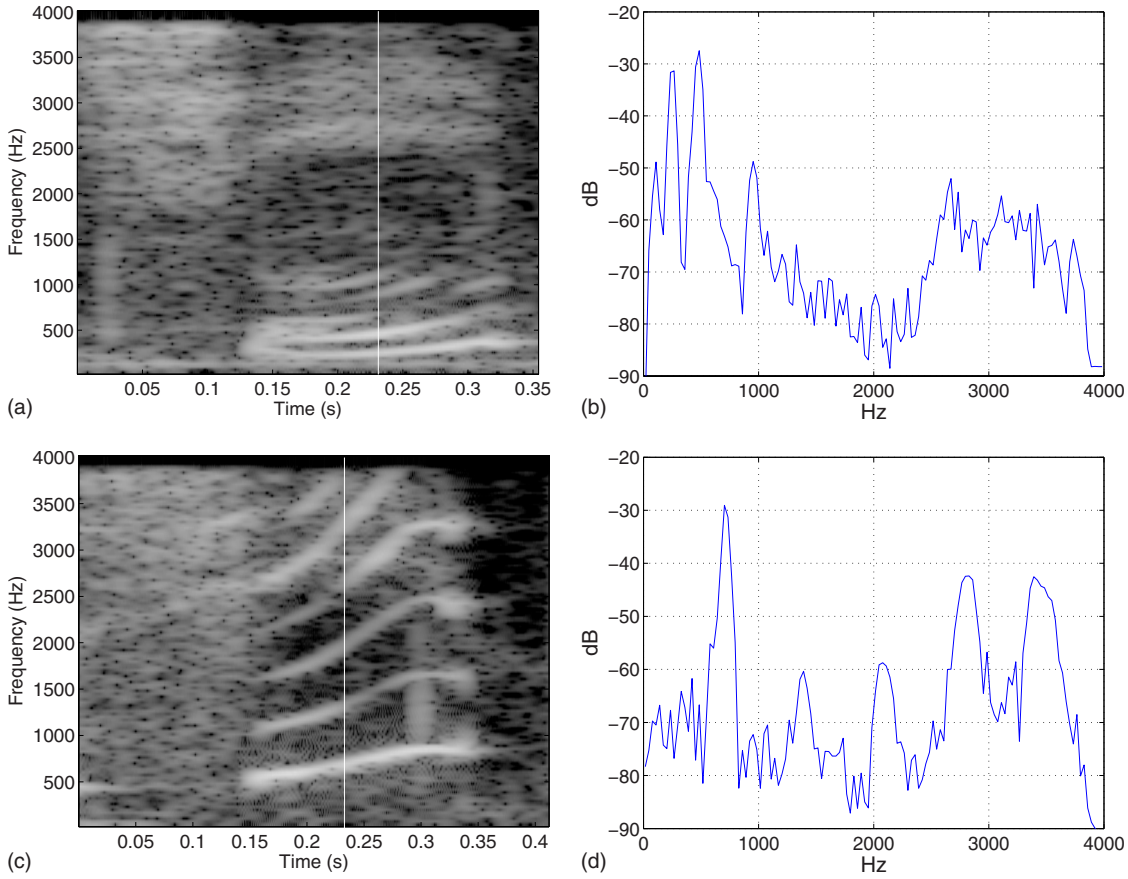


FIG. 6. (Color online) Spectrograms and their sections of "sheep" sound. In infant-directed /iy/ sound, fundamental frequency aligns with the first formant.

where $\hat{S}_i(\omega)$ is the estimated normalized signal associated with the ith time frame, $Y_i(\omega)$ is the corresponding observed signal, and $C(\omega)$ is the unknown constant channel effect which is approximated using an averaging operation $\langle\cdot\rangle_1^n$ on the observed n-length sequence of spectral vectors. (The same result is derived by applying cepstral mean normalization[27] with complex cepstra and inverting it to log spectrum domain. Since the cosine transformation from log spectrum to cepstrum is a linear transformation, it is canceled in the inversion.) The normalization is inserted between the first and the second Fourier transform of the modulation spectrum.

Figure 7 shows the two-dimensional MDS representation of the different speakers for each genre: Each point represents a speaker-genre combination, i.e., the transformation of the speaker's average modulation spectrum vector for either infant-directed/adult-directed or dictation/spontaneous manner. As can be seen, the spontaneous and adult-directed samples are arranged close to each other showing similarity of the styles. The dictated and infant-directed samples are on opposite sides of these two, representing different extremes of the range of styles examined here. The authors had expected some similarities between read speech and infant-directed speech, assuming that both would have more instances of well-articulated phonemes, but presumed that the difference was because of hyperarticulation that tended to be frequent in infant-directed speech but less so in read speech. The authors conjecture that the MDS dimensions capture the variability of the speech, with dictated (read) speech being the least variable and infant-directed speech being the most variable. The result is consistent with the F0 analysis in that the infant-directed speech is at one extreme, but the different genres seem better separated by the MDS analysis.

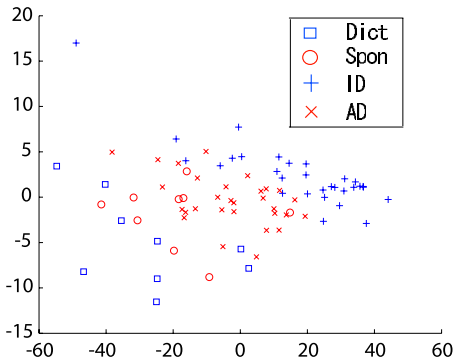


FIG. 7. (Color online) MDS representation of modulation spectrum features of speakers associated with dictated (Dict), spontaneous (Spon), infant-directed (ID), and adult-directed (AD) speech.

MDS is also used in an analysis of speaking style where distances between speakers are in terms of hidden Markov model state distribution parameters.[28] In that study, all of the speech is read, so it is difficult to make a comparison to the findings here. Further, the approach in that study requires transcribed speech, whereas that of the authors is strictly based on acoustic observations. An important difference between the two sets of findings is that speaking rate is an

TABLE I. Utterance features and associated classification error rates.

Feature	Error
F0	22.7
EN	40.3
MS	23.7
F0+EN	21.1
F0+MS	18.7
EN+MS	23.0
F0+EN+MS	17.9

important dimension in the Shozakai and Nagino study,[28] but in the authors' study the two slow rate cases (read speech and motherese) are at opposite ends of their distribution.

IV. AUTOMATIC CLASSIFICATION OF SPEAKING STYLES

The above analysis suggests that the modulation spectrum should provide useful features for recognizing speaking style, but the findings are based on averages of multiple utterances from a speaker, and classification might be better aimed at the utterance level. In this section, the authors compare the performance of different features used in automatic classification of speaking style, with the primary aim of assessing the relative utility of different features and secondarily to assess the performance that can be achieved with acoustic cues alone. The following three kinds of features and their combinations were used in the analysis.

(1) F0: average F0. Before the features are calculated for utterances, the F0 is normalized for each speaker by subtracting mean that is estimated using both genres. Only voiced segments are used.
(2) EN: average log energy. Before the features are calculated for utterances, the energy is normalized for each speaker by subtracting mean and dividing by standard deviation that are estimated using both genres.
(3) MS: modulation spectrum feature vector. The modulation spectrum matrix is first averaged and decimated to obtain a fixed, low-dimension (5×5) matrix represented as a 25-dimensional vector.

These features were calculated from utterance wave forms with 8 kHz sampling.

A linear discriminant function was used in a binary classification task: identifying adult- vs infant-directed speech in the Motherese corpus. The simple linear function was chosen because the training set size was not large and because the speaker-level data suggested that it would be effective. (In genre detection where more labeled data are available, it may be interesting to investigate more complex methods.) The parameters of the linear discriminant functions were estimated, and style classification performance was evaluated using tenfold cross-validation.

Table I shows the classification error rates. For comparison, guessing based on priors has an error rate of 49%. Among the three single features, F0 gave the lowest error rate of 22.7%, consistent with prior work showing high average F0 associated with infant-directed speech. (When F0 was normalized by both mean and standard deviation, the error rate was 22.9%, which was similar but slightly higher than the error rate of 22.7% when only the mean was normalized. This was probably because of the difficulty in estimating the normalization factor since infant-directed side has much larger variance than the adult-directed side.) The MS features gave a result slightly higher than that for F0. (The authors also looked at reduced dimension versions of the MS feature, but this led to increased error.) Using energy alone gave the highest error rate, which is partly because of the difficulty in removing channel effects from a far-field microphone recording. The lowest error rate (17.9%) was obtained by combining F0, energy, and modulation spectrum features.

V. SAMPLING FOR ASR TRAINING

While it is widely accepted that more data lead to improved speech recognition performance, it is also important to have data that are well matched to the target task. For new tasks (particularly new languages) where large amounts of transcribed data are not readily available, the cost of transcription can have a significant impact on overall development costs. In addition, for some tasks, experiments have shown that recognition performance can be improved by omitting certain samples from the training set. For that reason, researchers have investigated methods for selecting data for training. Early work in speech recognition[13] involved selecting based on a recognition error or speech recognizer confidence, specifically choosing to add those utterances with high error (or low confidence). With an error criterion, it was shown that improved performance could be achieved with less than the full amount of data, but this approach requires transcription for measuring errors and so cannot be used for initial selection. While recognizer confidence was less effective in that study, it has since been used to good effect in identifying utterances to remove from the training set (either because of poor transcriptions or noisy conditions, e.g., Ref. 29) and in active learning.[30] An alternative to measuring confidence of one system is to look at disagreement among multiple systems, as in a study applying hidden Markov models (HMMs) to part-of-speech tagging.[31] Another early study[32] looked at representing speech from groups of speakers with supervectors of average cepstral parameters from a subset of phones based on a force-alignment to the transcript, which are reduced in dimension with principal component analysis. These vectors are clustered, and data are selected to best represent the different clusters. A more recent related study seeks to sample the space characterized by distances between HMMs trained for different speakers, finding that the best results are obtained by sampling at the periphery of a small dimensional space learned via MDS,[33] building on the results of Shozakai and Nagino[28] and reducing the requirements for speech transcription by using adaptation for training the speaker-dependent models.

These different results are not entirely consistent in their recommendations. The clustering results suggest that one should sample to cover the space. Other results suggest that one should sample to emphasize outliers after some initial training phase, similar to the philosophy of boosting. Look-

ing at human language, learning would support the idea of using multiple strategies in the sense that mothers talk differently to their children when they are very young, but it also suggests the use of sampling in the first stage. In this work, the authors investigate methods of sampling in a single training phase and a two-stage approach. While they include average per-frame forced alignment likelihood (roughly equivalent to confidence) as a baseline criterion for comparison, the goal is to identify acoustic criteria that indicate utterances to select for transcribing for initial training. A secondary goal of investigating multipass strategies is to develop an efficient training strategy that gives a good ASR model with a lower computational cost for training.

A. Utterance features for sampling

The following one-dimensional features are used as the sampling criteria and compared to a random sampling baseline.

(1) LL: average per-frame log acoustic likelihood obtained by forced alignment to the reference transcript.
(2) F0: average normalized fundamental frequency. Only voiced segments are used.
(3) EN: average normalized log energy.
(4) MS: linear discriminant projection of modulation spectrum features designed to separate infant-directed vs adult-directed speech.

LL relies on an ASR model, but the others are acoustically based features. LL was estimated using a model developed for SRI's Decipher, a large vocabulary recognizer, trained on the same data it is used to sample from. LL is included because it is an indicator of "representative" (or "outlier") utterances in that higher (or lower) log likelihood occurs for utterances that are closer to (or farther from) the mean. However, there are some limitations of the LL measure. First, it cannot be used for the task of selecting data to transcribe since it requires transcriptions for forced alignments. Second, there is a bias introduced by using a previously trained ASR model to determine what is typical. Hence, LL mainly serves as a comparison point.

For the F0 and EN features, the raw features are extracted as described in the genre-classification study and are normalized for each speaker. For the MS feature, the authors first train a 25-dimensional linear classifier to distinguish between infant-directed and adult-directed speech, as in the genre-classification experiments, and then use the resulting score as the feature, which corresponds to a linear discriminant analysis projection.

B. Speech recognition systems

1. HTK-based system

A small CTS task defined in[34] was used to evaluate the authors' methods with a recognition system based on the HTK toolkit.[35] The acoustic model training set of the CTS task consisted of approximately 32 h of speech (16 h for each gender) coming from a mixture of Fisher[36] and Switchboard[37] training utterances. This baseline training set was selected by uniformly sampling the Switchboard and Fisher training sets, with the constraint that the two sources would comprise roughly 40% and 60% of each 16 h subset, respectively. In the following experiments, the male part of this baseline training set was used and compared to various sampled subsets from the same corpora.

The acoustic features were 12 perceptual linear predictive (PLP) coefficients[38] and energy, with their first two derivatives computed with vocal tract length normalization[39] and mean and variance normalization. The acoustic model was a set of three state left-to-right tried-state triphone HMMs with 32 mixtures per HMM state and had 2000 states across all triphones. The model was trained using the typical HTK "recipe" of initializing triphones from monophones, clustering single Gaussian triphones, and gradually increasing the number of mixtures after clustering. Models are updated with five iterations of expectation-maximization (EM) training at each step.

The small CTS task test sets were selected from the RT03 evaluation test set[40] based on constraining the out-of-vocabulary rate associated with a 1k-word vocabulary (the highest frequency words in the full corpus). (The original RT03 evaluation set contains about the same number of utterances from the Fisher and Switchboard corpus.) The male portions of the small CTS task test sets consist of 35 min of data for tuning and 32 min for testing. The dictionary for decoding contains multi-words and multiple pronunciations, so the overall size is 5.1k. A bigram language model was used, made by projecting the 2004 CTS evaluation language model onto the 1k vocabulary.

2. Decipher-based system

In order to investigate how the different sampling methods work in a large vocabulary system, experiments were also conducted using the SRI Decipher[41] system. The baseline training sets were randomly sampled from the Fisher and Switchboard corpus as for the HTK-based system. The test set is the RT04 development test set. The dictionary is based on 38k-word vocabulary and having 83k entries including multi-words and multiple pronunciations. Decoding involves rescoring a lattice of initial pass hypotheses with a speaker-adapted model (using maximum likelihood linear regression) and a 4-gram language model. Note that this system is different from the standard SRI recognition system in that it has only PLP cross-word triphone models and only ML training is used.

C. Sampling for data selection

For each feature considered, utterances in the training set are classified to three equally sized classes of "lowest," "middle," and "highest" in order to assess whether typical utterances (middle category) are more or less useful than outliers.

1. Small vocabulary results using HTK

For each of the utterance-level features, utterances from the Fisher and Switchboard corpora were sorted and partitioned into three subsets based on increasing feature ranges: lowest, middle, and highest. These ranges were chosen so

TABLE II. Sampling criteria and resulting WER. The baseline WER was 41.4.

Scoring measure	Lowest	Middle	Highest
LL	41.6	41.1	41.7
F0	41.3	41.0	41.1
EN	42.7	40.8	42.7
MS	41.1	40.4	41.9
F0+EN+MS	41.1	41.0	41.4

that the subsets were the same size in terms of duration. For each subset, 16 h of segments were randomly selected as training data. The ratio of durations of the two corpora within each of the sampled sets was kept the same as the original training set.

Table II shows the WER of the models trained using the subsets. Random sampling is used to provide a baseline. Generally speaking, the sampled subsets from the middle classes worked better than those from the lowest or highest classes, regardless of the scoring methods. All the samplings from the middle classes gave lower WER than the baseline, and among these MS gave the best result of 2.4% relative WER reduction. For the MS feature, a higher value meant that the utterance sounds more like infant-directed. For HMM training, however, it can be seen that utterances with mid-range features are more useful than higher scoring utterances that the authors presumed to be hyperarticulated. Either hyperarticulation is not as useful in machine learning as it is for human language learners, or the high MS score captures something other than hyperarticulation. The combination of F0, EN, and MS, which led to better classification of infant- vs adult-directed speech, was also tested but did not lead to lower WERs.

2. Large vocabulary results using Decipher

Both male and female triphone models were trained respectively using 16 h of the sampled training set. In the decoding, the system automatically decided which model to use by comparing likelihood of Gaussian mixture models (GMMs) associated with male and female speakers. The male/female GMMs were trained from HUB5 and had 256 mixtures each. Table III shows the WER of the subsets from the middle classes since this gave the best results in the HTK small task experiments. As can be seen, the feature-based sampling also results in lower WER than the baseline in this experiment using the large vocabulary Decipher systems. However, the LL criterion gives better results perhaps because of the match to the Decipher models.

TABLE III. WER for large vocabulary recognition using Decipher.

Sampling measure	WER
Baseline	26.1
LL	25.7
F0	25.9
EN	25.7
MS	25.9

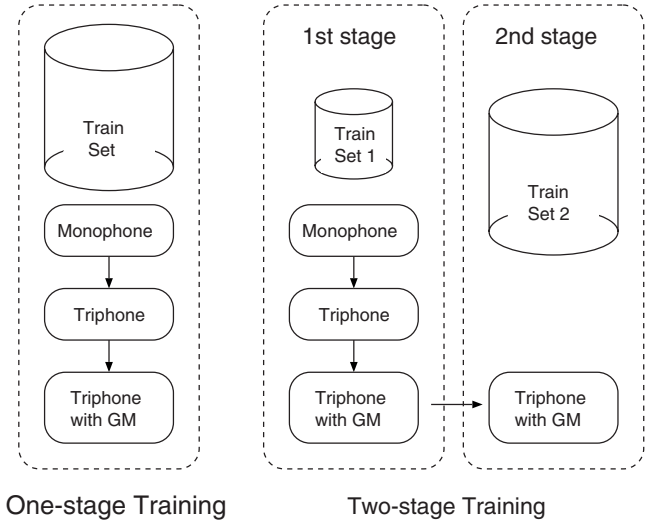


FIG. 8. One-stage and two-stage training procedures.

D. Sampling for two-stage training

Usually, acoustic models are trained using a locally optimal iterative method such as the EM algorithm.[42] By using an initial model trained on a subset of the data with better separated classes (either prototypical instances or data from the middle region only), the final model may avoid problems of local optima and require less training time (lower computational cost). Thus, to improve efficiency and hopefully performance, two-stage training methods are investigated. Figure 8 shows the procedure of conventional one-stage training and proposed two-stage training. In these experiments, the decision tree designed in stage 1 is fixed in stage 2, though the distributions are re-estimated. The experiments are designed to answer two questions: (1) Does two-stage training with increased amounts of data in the second stage yield improved performance over one-stage training? (2) Is it useful to constrain model means to the prototypes learned initially and update only the variances? The second question was motivated by work on child language acquisition showing early learning of prototypical vowel sounds.

1. Results using HTK-based systems

For two-stage training, 48 h of training data were randomly selected as a full set, using the same second stage random sampling for all different initial models. Acoustic models trained on the 16 h subset of the middle classes were used as the initial model, and their parameters were updated using two EM iterations in the second stage. Both the one-stage and two-stage models had 2k states and 32 mixtures per HMM state. Two types of EM training were conducted: updating only variances vs updating all the parameters. Note that the subset used to train the initial model was not added in the second stage to keep the experimental condition the same among the sampling methods excepting the parameter initialization.

Table IV shows the WER. In the table, results of mean only update are also shown for a reference purpose. The baseline "one-stage" model was trained from scratch on the 48 h of the full set. While some models gave slightly better

TABLE IV. WER of one vs two-stage training using HTK, with and without update constraints.

Sampling	Variance	Mean	All
One-stage		41.0	
Random	41.6	40.9	40.9
LL	41.4	40.8	40.6
F0	40.6	40.8	40.7
EN	40.3	40.1	40.5
MS	41.5	40.4	40.1

results when only variance or mean was updated, it was more often better to update all parameters. When all parameters were updated, all two-stage strategies improved performance over the one-stage baseline, which supports the hypothesis that initial training with better separated classes is helpful. As before in the HTK-based experiments, the MS-based initialization gave the best result. The computation time of these two-stage EM training was only 40% compared to one-stage training.

2. Results using Decipher-based systems

The two-stage training strategy was also evaluated using the Decipher system using 32 h of training data in the first stage and 64 h in the second stage. The results are shown in Table V. When the randomly sampled set was used in the first stage, the WER was increased compared to the one-stage baseline. On the other hand, there were small reductions in WER when the initial set was selected based on the utterance features, suggesting that better initialization can improve overall system performance. More significantly, the cost of two-stage training is 65% of the one-stage approach.

VI. CONCLUSIONS

Motivated by insights from human language acquisition and the high cost of transcribing speech when moving to a new domain, the authors analyzed the characteristics of speaking styles and investigated sampling methods for ASR.

The analyses were performed in the first half of the paper using the MULTI-REG corpus and the motherese corpus. While both the infant-directed and dictation utterances tend to have more clearly articulated utterances and slower speaking rates, the authors found that these appear on opposite ends of a scale learned through MDS of the modulation spectrum. They also discovered the pitch-formant matching phenomenon in highly exaggerated infant-directed utterances. The modulation spectrum was also shown to be useful in automatic classification of speaking style.

TABLE V. WER of one vs two-stage training using Decipher.

Sampling	WER
One-stage	23.8
Random	24.2
LL	23.6
F0	23.7
EN	23.7
MS	23.7

In the second half of the paper, the authors have investigated sampling methods for ASR based on features used in the analyses. For the small CTS task, it was shown that sampling from middle range classes gave lower WER than from lowest or highest classes for all of the utterance features. Among the sampling criteria, the lowest WER (2.4% reduction) was obtained by mid-range sampling using a modulation-spectrum-based feature. Sampling was useful in both a small vocabulary simple HTK-based system and a more complex large vocabulary Decipher system. While acoustic measures did well with the HTK system, the better matched likelihood criterion was most useful for the case where a model is trained in advance.

In addition, sampling was useful for improving the training schedule, not only reducing the computational cost, but also leading to gains in WER in some cases. In the two-stage training experiment using HTK, it was shown that better performance was obtained by initializing with a model trained on data selected to include utterances near the mean of the modulation spectrum. In other words, using lower variance data in initializing the model is helpful. In addition, updating all parameters was better than updating only variances. The computational cost was about 40%–65% compared to the conventional one-stage training. The cost savings may increase in scaling to larger data sets, but there are issues to explore related to incrementing model complexity. An open question is how sampling interacts with discriminative training.

ACKNOWLEDGMENT

This work was supported by DARPA Grant No. MDA972-02-1-0024.

[1]P. K. Kuhl, J. E. Andruski, I. A. Chistovich, L. A. Chistovich, E. V. Kozhevnikova, V. L. Ryskina, E. I. Stolyarova, U. Sundberg, and F. Lacerda, "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686 (1997).
[2]D. Burnham, C. Kitamura, and U. Vollmer-Conna, "What's new, pussycat? On talking to babies and animals," Science **296**, 1435 (2002).
[3]K. Kirchhoff and S. Schimmel, "Statistical properties of infant-directed vs. adult-directed speech: Insights from speech recognition," J. Acoust. Soc. Am. **117**, 2238–2246 (2005).
[4]H. M. Liu, P. K. Kuhl, and F. M. Tsao, "An association between mothers' speech clarity and infants' speech discrimination skills," Dev. Sci. **6**, F1–F10 (2003).
[5]V. Hazan and A. Simpson, "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects," Lang. Speech **43**, 273–294 (2000).
[6]R. P. Lippmann, "Speech recognition by machines and humans," Speech Commun. **22**, 1–15 (1997).
[7]D. A. van Leeuwen, L. G. V. den Berg, and H. J. M. Steeneken, "Human benchmarks for speaker independent large vocabulary recognition performance," in Proceedings of Eurospeech (1995), Vol. **2**, pp. 1461–1464.
[8]O. Scharenborg, "Reaching over the gap: A review of efforts to link human and automatic speech recognition research," Speech Commun. **49**, 336–347 (2007).
[9]M. Weintraub, K. Taussig, K. Hunicke-Smith, and A. Snodgrass, "Effect of speaking style on LVCSR performance," in Proceedings of ICSLP, Philadelphia, PA (1996), pp. 16–19.
[10]M. Saraclar, H. Nock, and S. Khudanpur, "Pronunciation modeling by sharing Gaussian densities across phonetic models," Comput. Speech

Lang. **14**, 137–160 (2000).
[11]T. Kawahara, H. Nanjo, T. Shinozaki, and S. Furui, "Benchmark test for speech recognition using the Corpus of Spontaneous Japanese," in Proceedings of SSPR2003 (2003), pp. 135–138.
[12]K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," J. Acoust. Soc. Jpn. (E) **20**, 199–206 (1999).
[13]T. M. Kamm and G. G. L. Meyer, "Selective sampling of training data for speech recognition," in *Proceedings of Human Language and Technology*, San Francisco, CA (2002), pp. 20–24.
[14]G. Zweig and M. Padmanabhan, "Boosting Gaussian mixtures in a LVCSR system," in *Proceedings of ICASSP*, Istanbul, Turkey (2000), pp. 1527–1530.
[15]H. Dudley, "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177 (1939).
[16]N. Kowalski, D. Depireux, and S. Shamma, "Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra," J. Neurophysiol. **76**, 3503–3523 (1996).
[17]G. Langner, M. Sams, P. Heil, and H. Schulze, "Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: Evidence from magnetoencephalography," J. Comp. Physiol. **181**, 665–676 (1997).
[18]L. Atlas and S. Shamma, "Joint acoustic and modulation frequency," EURASIP J. Appl. Signal Process. **2003**, 668–675 (2003).
[19]R. Drullman, J. M. Festen, and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680 (1994).
[20]S. Schimmel and L. Atlas, "Target talker enhancement in hearing devices," in *Proceedings of ICASSP*, Las Vegas, NV (2008), pp. 4201–4204.
[21]S. Greenberg and T. Arai, "The relation between speech intelligibility and the complex modulation spectrum," in *Proceedings of Eurospeech*, Aalborg, Denmark (2001), pp. 473–476.
[22]J. Darch, B. Milner, I. Almajai, and S. Vaseghi, "An investigation into the correlation and prediction of acoustic speech features from MFCC vectors," in *Proceedings of ICASSP* (2007), Vol. **IV**, pp. 465–468.
[23]D. Talkin, "A robust algorithm for pitch tracking (RAPT)," in *Speech Coding and Synthesis*, edited by W. Kleijn and K. Paliwal (Elsevier Science, Amsterdam, 1995), pp. 495–518.
[24]K. Sonmez, E. Shriberg, L. Heck, and M. Weintraub, "Modeling dynamic prosodic variation for speaker verification," in *Proceedings of ICSLP*, Sydney, Australia (1998), Vol. **7**, pp. 3189–3192.
[25]E. C. Willis and D. T. Kenny, "Effect of voice change on singing pitch accuracy in young male singers," J. Interdisciplinary Music Studies **2**, 111–119 (2008).
[26]*Modern Multidimensional Scaling: Theory and Applications*, edited by I. Borg and P. Groenen (Springer-Verlag, New York, 1997).
[27]B. Atal, "Effectiveness of linear prediction characteristics of of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Am. **55**, 1304–1312 (1974).
[28]M. Shozakai and G. Nagino, "Analysis of speaking styles by two-dimensional visualization of aggregate of acoustic models," in *Proceedings of ICSLP*, Jeju, Korea (2004), Vol. **I**, pp. 717–720.
[29]H. Y. Chan and P. C. Woodland, "Improving broadcast news transcription by lightly supervised discriminative training," in *Proceedings of ICASSP*, Quebec, Canada (2004), Vol. **I**, pp. 737–740.
[30]G. Riccardi and D. Hakkani-Tur, "Active learning: Theory and applications to automatic speech recognition," IEEE Trans. Speech Audio Process. **13**, 504–511 (2005).
[31]I. Dagan and S. Engelson, "Committee-based sampling for training probabilistic classifiers," in *Proceedings of ICML*, Tahoe City, CA (1995), pp. 150–157.
[32]A. Nagorski, L. Boves, and H. Steeneken, "Optimal selection of speech data for automatic speech recognition systems," in *Proceedings of ICSLP*, Denver, CO (2002), pp. 2437–2440.
[33]G. Nagino and M. Shozakai, "Building an effective corpus by using acoustic space visualization (COSMOS) method," in *Proceedings of ICASSP*, Philadelphia, PA (2005), Vol. **I**, pp. 449–452.
[34]B. Chen, O. Cetin, G. Doddington, N. Morgan, M. Ostendorf, T. Shinozaki, and Q. Zhu, "A CTS task for meaningful fast-turnaround experiments," in NIST RT-04 Workshop, Palisades, NY (2004).
[35]S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. A. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book* (Cambridge University Engineering Department, Cambridge, 2006).
[36]C. Cieri, D. Miller, and K. Walker, "The Fisher corpus: A resource for the next generations of speech-to-text," in *Proceedings of LREC*, Lisbon, Portugal (2004), pp. 69–71.
[37]J. J. Godfrey, E. C. Holliman, and J. McDaniel, "Switchboard: Telephone speech corpus for research and development," in *Proc. ICASSP*, San Francisco, CA (1992), Vol. **I**, pp. 517–520.
[38]H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Am. **87**, 1738–1752 (1990).
[39]E. Eide and H. Gish, "A parametric approach to vocal tract length normalization," in *Proceedings of ICASSP*, Atlanta, GA (1996), Vol. **I**, pp. 346–348.
[40]http://www.nist.gov/speech/tests/rt/ (Last viewed April 21, 2008).
[41]A. Stolcke, H. Bratt, J. Butzberger, H. Franco, V. R. R. Gadde, M. Plauche, C. Richey, E. Shriberg, K. Sonmez, F. Weng, and J. Zheng, "The SRI March 2000 Hub-5 conversational speech transcription system," in *Proceedings of NIST Speech Transcription Workshop*, College Park, MD (2002).
[42]A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Stat. Soc. Ser. B (Methodol.) **39**, 1–38 (1977).

Pitch bending and *glissandi* on the clarinet: Roles of the vocal tract and partial tone hole closure

Jer-Ming Chen,[a] John Smith, and Joe Wolfe
School of Physics, The University of New South Wales, Sydney, New South Wales 2052, Australia

(Received 9 March 2009; revised 15 June 2009; accepted 17 June 2009)

Clarinettists combine non-standard fingerings with particular vocal tract configurations to achieve pitch bending, i.e., sounding pitches that can deviate substantially from those of standard fingerings. Impedance spectra were measured in the mouth of expert clarinettists while they played normally and during pitch bending, using a measurement head incorporated within a functioning clarinet mouthpiece. These were compared with the input impedance spectra of the clarinet for the fingerings used. Partially uncovering a tone hole by sliding a finger raises the frequency of clarinet impedance peaks, thereby allowing smooth increases in sounding pitch over some of the range. To bend notes in the second register and higher, however, clarinettists produce vocal tract resonances whose impedance maxima have magnitudes comparable with those of the bore resonance, which then may influence or determine the sounding frequency. It is much easier to bend notes down than up because of the phase relations of the bore and tract resonances, and the compliance of the reed. Expert clarinettists performed the *glissando* opening of Gershwin's *Rhapsody in Blue*. Here, players coordinate the two effects: They slide their fingers gradually over open tone holes, while simultaneously adjusting a strong vocal tract resonance to the desired pitch.
 [DOI: 10.1121/1.3177269]

PACS number(s): 43.75.Pq, 43.75.St [NHF] Pages: 1511–1520

I. BACKGROUND

A. Pitch bending and *glissandi*

Pitch bending refers to adjusting the musical pitch of a note. Usually it means a smooth variation in pitch and can include *portamento* and *glissando*, which refer to continuous variation of pitch from one note to the next. The pitch of some instruments can be varied continuously over a wide range by adjusting the position of the hands and fingers, e.g., the slide trombone or members of the violin family. The pitch of some fretted string instruments, e.g., the guitar and sitar, can also be varied smoothly over a restricted range by moving the finger that stops the string on the fingerboard, thereby changing the string tension. The pitch of lip-valve instruments can be altered via changes in lip tension (lipping). In woodwind instruments, each particular configuration of open and closed tone holes is called a fingering, and each fingering is associated with one or more discrete notes. Woodwinds are, however, capable of pitch bending, either by partially opening/closing tone holes or by playing techniques that involve the player's mouth, vocal tract, breath, and in the case of some air-jet woodwinds such as the flute and shakuhachi, adjusting the extent of baffling with the face.

On the clarinet, partially covering a tone hole can be used to achieve pitch bending when a transition between notes uses a tone hole covered directly by a finger rather than by a pad. Seven tone holes are covered directly by the fingers, allowing bending by this method over the range G3 (175 Hz) to G4 (349 Hz) and from D5 (523 Hz) upwards. (The clarinet is a transposing instrument; clarinet written pitch is used in this paper—one musical tone above sounding pitch.) Substantial pitch bending using the vocal tract, however, is usually possible only over the upper range of the instrument (Pay, 1995), typically above about D5 (523 Hz), although the actual range depends on the player. Further, this bending is asymmetric: Although expert players can use their vocal tract and embouchure to lower the pitch by as much as several semitones, they can only raise the pitch slightly (Rehfeldt, 1977). Similar observations apply to saxophones, whose reed and mouthpiece are somewhat similar to those of clarinets.

Pitch bending on the clarinet is used in several musical styles including jazz and *klezmer*. In concert music, the most famous example is in the opening bar of Gershwin's *Rhapsody in Blue* (Fig. 1), which features a clarinet playing a musical scale over two and half octaves. At the composition's first rehearsal, the clarinettist replaced the last several notes in the scale with a *glissando* (Schwartz, 1979). This delighted the composer and started a performance tradition. Figure 1 shows the spectrogram of a performance in this style, in which the *glissando* spans an octave from C5 (466 Hz) to C6 (932 Hz) at the end of the run. It is explained in Sec. III B why the *glissando* usually replaces only the last several notes, as shown in Fig. 1.

This solo in Rhapsody in Blue (including the performance tradition) is such a standard part of the clarinet repertoire that it is well known to most professional clarinettists. It is therefore used as the context for part of this study into the roles that a player's vocal tract and partial covering of tone holes can play in pitch bending. For the rest of the study, an artificial exercise was used: Players were simply asked to bend down the pitch of standard notes on the clari-

[a]Author to whom correspondence should be addressed. Electronic mail: jerming@phys.unsw.edu.au

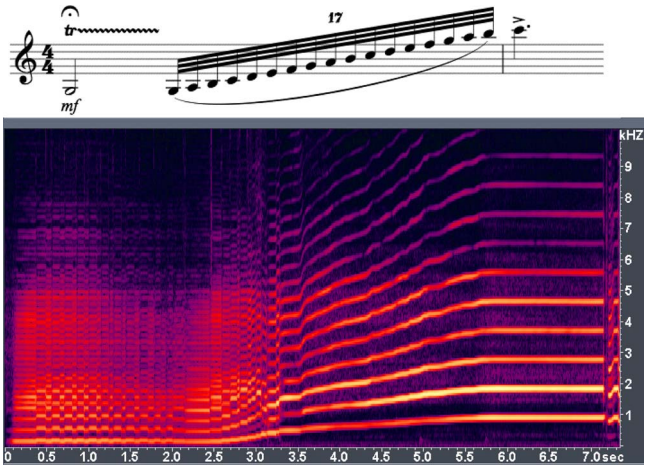


FIG. 1. (Color online) The opening of Gershwin's Rhapsody in Blue. The upper figure shows the 2.5 octave run as it is written—but not as it is usually played. In traditional performance, the last several notes of the scale are replaced with a smooth *glissando*. The lower figure is a spectrogram of such a performance. The opening trill on G3 (written, 174 Hz) is executed from 0 to 2 s and followed by a scale-like run at 2.5–3.5 s that becomes smooth pitch rise from C5 (466 Hz) to C6 (932 Hz) at 3.5–5.7 s.

net, using only their vocal tracts. Acoustic impedance spectra of the clarinet bore were measured using techniques reported previously (Dickens *et al.*, 2007a, 2007b), while impedance spectra inside the player's mouth were measured using an impedance head built into the mouthpiece of the clarinet so that the player could perform with very little perturbation.

B. The sounding frequency of the clarinet

The sounding frequency f_0 of the clarinet is determined by several interacting effects and thus depends on a number of parameters. Some of these effects are modest (such as lip damping, bite configuration, jaw force, and blowing pressure) and, in this study, the aim was to hold them constant rather than to examine them in detail. For example, f_0 depends on the natural frequency of the reed. For a given reed, this can be adjusted by the player during performance by applying greater or lesser force with the lower jaw, thereby varying its effective length and vibrating mass and thus influencing the sounding pitch slightly.

The clarinet has a range of rather more than three octaves and, to first order, stable reed oscillation (at the sounding frequency f_0) occurs near one of the maxima in the acoustic impedance Z_{load} that loads the reed generator, which, along with the pressure difference between mouthpiece and bore, determines the airflow into the instrument (Fletcher and Rossing, 1998). The acoustic pressure difference across the reed is $\Delta p = p_{\text{tract}} - p_{\text{bore}}$ where p_{tract} and p_{bore} are, respectively, the acoustic pressures in the mouth near the reed (upstream) and in the clarinet mouthpiece near the reed (downstream). If U is the acoustic flow passing through the aperture between the reed and the mouthpiece, then $Z_{\text{load}} = \Delta p / U$. Benade (1985) offered a considerable simplification of processes at the reed junction, applied continuity of volume flow, and assumed that p_{tract} and p_{bore} both act on equal areas of the reed. He then showed that the impedance loading the reed generator is given by

$$Z_{\text{load}} = \frac{Z_{\text{reed}}(Z_{\text{bore}} + Z_{\text{tract}})}{Z_{\text{reed}} + Z_{\text{bore}} + Z_{\text{tract}}} = (Z_{\text{bore}} + Z_{\text{tract}}) \parallel Z_{\text{reed}}. \quad (1)$$

This impedance includes contributions from the bore of the instrument (Z_{bore}) and the player's vocal tract (Z_{tract}) where both are measured near the reed. It also includes the effective impedance of the clarinet reed (Z_{reed}) itself: Δp divided by the volume flow due to reed vibration. The second expression in Eq. (1) is included to show explicitly that, in this rudimentary model, Z_{tract} and Z_{bore} are in series, and their sum is in parallel with Z_{reed}. Consequently, under conditions in which the vocal tract impedance is small compared to the bore impedance, Z_{load} depends only on Z_{bore} and Z_{reed} alone—indeed, the maximum in measured impedance of the clarinet bore in parallel with the reed corresponds closely to the pitch in normal playing (Benade, 1985). On the other hand, if the player were able to make Z_{tract} large and comparable to Z_{bore}, the player's vocal tract could significantly influence, or even determine, the sounding frequency of the player-instrument system.

The interaction of bore, reed, and airflow is inherently nonlinear and the subject of a number of analyses and experimental studies (e.g., Backus, 1963; Wilson and Beavers, 1974; Benade, 1985; Grand *et al.*, 1996; Fletcher and Rossing, 1998; Silva *et al.*, 2008). Although the nonlinear effects must be considered to understand the threshold pressure for blowing and features of the waveform and spectrum, there is agreement that the playing frequency can be explained to reasonable precision in terms of the linear acoustics of the bore, vocal tract, and reed, using Benade's (1985) model. Specifically, the operating frequency lies close to the frequency at which the imaginary part of the acoustical load is zero, which in turn is very near or at a maximum in the magnitude of the impedance. The present experimental paper considers only the linear acoustics of the bore, the vocal tract and the compliance of the reed.

C. Influence of the player's vocal tract

Some pedagogical studies on the role of the player's vocal tract in woodwind performance report musicians' opinions that the tract affects the pitch (Pay, 1995; Rehfeldt, 1977). Scientific investigations to date include numerical methods, modeling the tract as a one-peak resonator (Johnston *et al.*, 1986), electro-acoustic analog simulations in the time domain (Sommerfeldt and Strong, 1988), and digital waveguide modeling (Scavone, 2003). Clinch *et al.* (1982) used x-ray fluoroscopy to study directly performing clarinettists and saxophonists and concluded that "vocal tract resonance frequencies must match the frequency of the required notes" played. Backus (1985) later observed that the player's vocal tract impedance maxima must be similar or greater in magnitude than the instrument bore impedance in order to influence performance, but concluded that "resonances in the vocal tract are so unpronounced and the impedances so low that their effects appear to be negligible." On the other hand, Wilson (1996) while investigating pitch bending concluded that the upstream vocal tract impedance at the fundamental frequency in pitch bending must be large and comparable to the downstream bore impedance, but did

not report details on the vocal tract resonance frequency or the magnitude of its impedance. Watkins (2002) summarized several empirical studies of the use of the vocal tract and its reported or measured geometry in saxophone performance.

The environment in the clarinettist's mouth when playing poses challenges for direct measurements of vocal tract properties during performance. The vibrating reed generates high sound pressure levels in the mouth: Backus (1961) and Boutillon and Gibiat (1996) reported sound levels of 166 dB and exceeding 170 dB inside the mouthpiece of a clarinet and saxophone, respectively, when artificially blown. The static pressure and humidity in the mouth complicate measurements. To avoid these difficulties, some previous measurements of the musician's vocal tract (Wilson, 1996) were made under conditions that were somewhat different from normal performance. Fritz and Wolfe (2005) made acoustic impedance measurements inside the mouth by having the musician mime with the instrument for various musical gestures, including pitch bending. The peaks in impedance measured in the mouth were as high as a few tens of MPa s m^{-3} and so comparable with those of the clarinet bore, but no simple relation between the frequencies of the peak and the note played was reported.

More recently, Scavone *et al.* (2008) developed a method to provide a real-time measurement of vocal tract influence on saxophones while the player performs a variety of musical effects. This method, developed from Wilson's (1996) indirect technique, uses microphones in the mouthpiece, one on the tract side and one on the bore side. It uses the played note and its harmonics as the measurement signal, and so measures impedance ratios of the upstream vocal tract to downstream saxophone bore at harmonics of the sounding reed. This has the advantage of simplicity, a strong signal, and is fast enough to provide real-time feedback to players. However, it does not measure vocal tract resonance frequency or the magnitude of impedance at the tract resonance. Using this method, they found that during pitch bending on the alto saxophone, the pressure component at the playing frequency was larger in the player's mouth than in the bore, from which it may be surmised that a strong vocal tract resonance influences behavior of the reed.

At the same time, the authors (Chen *et al.*, 2008) reported direct impedance spectra measurements made inside the mouth during performance and described how expert tenor saxophonists can produce maxima in Z_{tract} and tune them to produce notes in the very high (*altissimo*) range of that instrument. Several amateur saxophonists, on the other hand, were found not to exhibit tuning, and consequently were unable to play in the altissimo range. The saxophone has a single reed mouthpiece, somewhat similar to that of a clarinet. A major difference, however, is that the saxophone's bore is nearly conical, while that of the clarinet is largely cylindrical. This difference has the result that, for high notes, the maxima in Z_{bore} on the saxophone (Chen *et al.*, 2009) are rather smaller than those in the clarinet (Backus, 1974; Dickens *et al.*, 2007b), so saxophone players can achieve the condition $Z_{\text{tract}} \approx Z_{\text{bore}}$ discussed above with weaker resonances of the vocal tract.

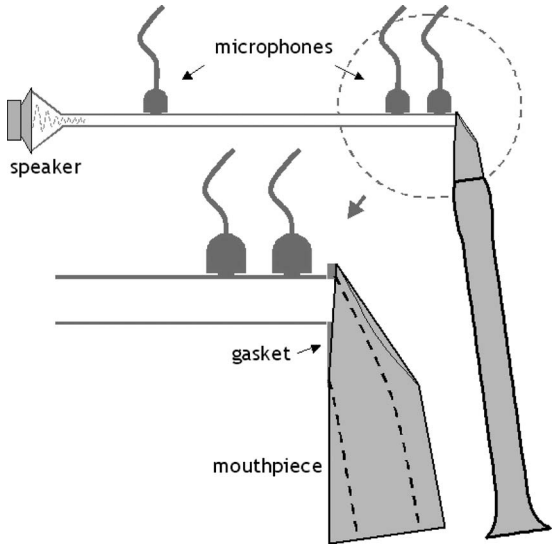


FIG. 2. Schematic of the 3M2C technique used to measure the acoustic impedance of the clarinet bore. The loudspeaker provides a stimulus synthesized as the sum of 2900 sine waves, while three microphones at known spacings record the response from the clarinet being measured. Not to scale.

II. MATERIALS AND METHODS

A. Measurements of bore impedance

Measurements of the acoustic impedance of the clarinet bore for the fingering positions employed in pitch bending were made on a B-flat soprano clarinet, the most common member of the clarinet family (Yamaha model CX). Z_{bore} was measured using the three-microphone-two-calibration (3M2C) method with two non-resonant loads for calibration (Dickens *et al.*, 2007a): one was an open circuit (nearly infinite impedance) and the other an acoustically infinite waveguide (purely resistive impedance). This method allows the measurement plane to be located near the reed tip "looking into" the clarinet bore (Fig. 2) without the involvement of a player. Clarinet bore impedances were then measured for standard fingering positions, as well as for some with partial uncovering of the tone holes. The excitation signal for these measurements was synthesized as the sum of sine waves from 100 to 4000 Hz with a spacing of 1.35 Hz. The measurements of standard fingerings gave results similar to those reported previously (Dickens *et al.*, 2007b). A professional clarinettist was engaged for measurements involving partial uncovering of tone holes. Fingering gestures that would be typical when executing the *glissando* in Rhapsody in Blue were used, involving the gradual and consecutive sliding of one's fingertips off the keys in order to uncover smoothly the open finger holes of the clarinet, starting from the lowest. Acoustic impedance spectra of the clarinet bore were measured at varying stages of the finger slide while the clarinettist temporarily halted the slide for several seconds.

B. Measurements of effective reed compliance

Representative values for the effective compliance of the clarinet reed during playing were measured using Benade's (1976) technique: measuring the sounding frequency produced by a clarinet mouthpiece (with reed) that is attached to

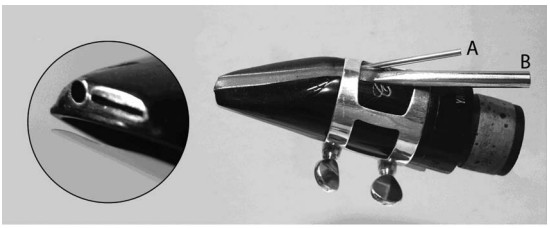


FIG. 3. Photograph of the modified mouthpiece used to measure the acoustic impedance of the player's vocal tract during performance. Tube A is attached to the microphone whereas tube B is attached to the calibrated source of acoustic current. The circular inset to the left shows a magnified view of the mouthpiece tip. The circular end of tube A and the rectangular cross section of tube B are visible just above the reed.

various lengths of pipe and played *mezzoforte* in a "normal" style, i.e., without using the vocal tract to adjust the pitch. Comparing the lengths of pipe for each tone with the calculated lengths of simple tubes (closed at one end) having a natural frequency matching the played one, the equivalent compliance and volume of the mouthpiece under playing conditions for a wide range of frequencies can be calculated by reducing the mouthpiece volume (having the tubes as deep as possible), then treating the calculated compliance as the sum of the compliance of the remaining mouthpiece volume and the compliance of the reed. Here, Benade's (1976) technique relies on the assumption that the impedance of the air in the gap between reed and mouthpiece does not contribute to the end effect. The magnitude of the ac component of the flow resistance is discussed in Sec. III B.

Three Légère synthetic reeds of varying hardness ($1\frac{3}{4}$, $2\frac{1}{2}$, and 3) were used in combination with cylindrical metal pipes with internal diameter of 14.2 mm and external diameter of 15.9 mm, and lengths of 99, 202, 299, and 398 mm. The reed compliance thus calculated was consistent for all pipe lengths listed. A reed of hardness 3 played with a tight embouchure, typical for normal playing, yielded an average equivalent volume of 1.1 ml, which equals the value given by Nederveen (1998). This value, treated as a pure compliance, is used for Z_{reed} in calculations here, i.e., reed losses are neglected. (The effect of different reeds and of the lip force applied to them was not studied in detail here. However, it is worth nothing that a reed of hardness 3 played with a relaxed embouchure yielded an equivalent volume of 1.7 ml, and that the softest reed (hardness $1\frac{3}{4}$) played with a relaxed embouchure gave 2.7 ml. Clarinettists are well aware that one can play flat with a relaxed embouchure, and especially with a soft reed.)

C. Measurements of vocal tract impedance

The acoustic impedance of the player's vocal tract was measured directly during performance using an adaptation of a technique reported previously (Tarnopolsky *et al.*, 2006; Chen *et al.*, 2008) and based on the capillary method [methods reviewed by Benade and Ibisi (1987) and Dickens *et al.* (2007a)]. An acoustic current is injected into the mouth via a narrow tube incorporated into a standard clarinet mouthpiece (Yamaha 4C)—see Fig. 3. The internal cross section of this narrow tube is approximately rectangular with an area of 2 mm^2, giving a characteristic source impedance around 200 MPa s m^{-3}. The sound pressure inside the mouth is measured via an adjacent tube embedded into the mouthpiece. This cylindrical tube (internal diameter of 1.2 mm) is connected to a microphone (Brüel & Kjær 4944A) located just outside the mouthpiece to form a probe microphone. The system thus measures the impedance looking into the vocal tract from a location inside the mouth just past the vibrating reed. It is calibrated by connecting the modified mouthpiece to the quasi-infinite tube used as a standard (Smith *et al.*, 1997), which has an internal diameter of 26.2 mm (comparable in size with the vocal tract) and length 197 m. To minimize the perturbations to the players, they used their own clarinet and a reed of their choice with the experimental mouthpiece. They reported only moderate perturbation to their playing—presumably because the mouthpiece geometry remains almost unchanged, except for an increase in thickness of about 1.5 mm at the bite point. (Indeed, some players routinely add a pad of up to 0.8 mm thickness at this point.)

The acoustic impedance spectrum for each particular vocal tract configuration was measured by injecting a calibrated acoustic current (synthesized as the sum of 336 sine waves from 200 to 2000 Hz with a spacing of 5.38 Hz) during playing. Each configuration was measured for 3.3 s to improve the signal-to-noise ratio. Because measurements are made during playing, the signal measured by the microphone will necessarily include the pressure spectrum of the reed sounding in the mouth. This produces clearly recognizable harmonics of the note played in the raw impedance spectrum with amplitudes much higher than those of the vocal tract itself, and allows the sounding frequency f_0 to be determined. To determine the $Z(f)$ measured in the mouth, these broadly spaced, narrow peaks are removed and replaced with interpolation. There are also high levels of broad band noise produced in the mouth. For this reason, the spectrum is then smoothed using a third order Savitsky–Golay filter typically over 11 points ($\pm$30 Hz) for magnitude values and 15 points ($\pm$40 Hz) for phase values. An example is shown in Fig. 4.

D. Players and protocols

Five expert clarinettists (four professional players and 1 advanced student) were engaged for the measurements on the players' vocal tract. They used the modified mouthpiece on their own clarinet and performed the following tasks.

(i) They played the opening bar of George Gershwin's Rhapsody in Blue (Fig. 1), first normally, then later pausing for a few seconds at various points in the *glissando* while their vocal tract impedance was measured.
(ii) They were asked to play chromatic notes in the range G4 (349 Hz) to G6 (1397 Hz) using standard fingerings and their normal embouchure and tract configuration, and to hold each note while the impedance in their mouth was measured.
(iii) They used standard fingerings for chromatic notes in the range G4 (349 Hz) to G6 (1397 Hz) and were asked to bend each sounding pitch progressively

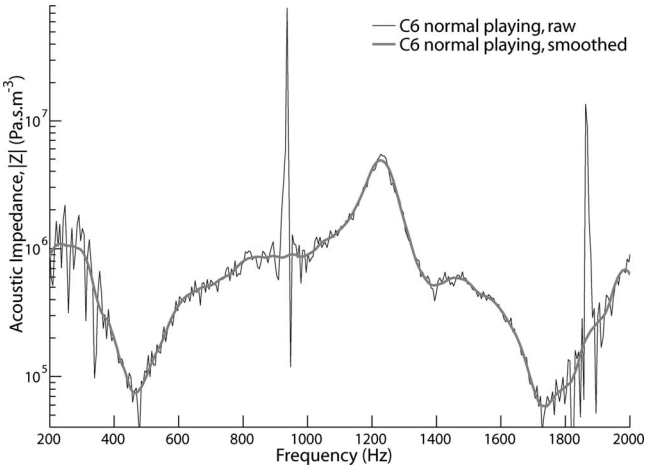


FIG. 4. The magnitude of the acoustic impedance spectrum measured in the mouth for a normal vocal tract configuration (and embouchure) playing the written pitch C6 (932 Hz) with standard fingering. Because measurements (thin line) are made during playing, harmonics of the note sounded appear added to the impedance spectrum—at 935 and 1870 Hz in this case. These are removed and replaced with interpolation, and the data then smoothed (thick line) to produce the vocal tract impedance spectra used in this paper. Here, the measured resonance is at 1225 ± 25 Hz.

down from its standard pitch and to hold it steady while the impedance in their mouth was measured.

III. RESULTS AND DISCUSSION

A. Effects of partial tone hole closure on the clarinet bore impedance

Acoustic impedance measurements of the clarinet bore (Z_{bore}) for standard fingerings show the expected, well-spaced maxima indicating the bore resonances near which the clarinet reed operates in normal playing (Fig. 5). This same figure also shows measurements using a fingering with one tone hole partly uncovered, to different extents, by a finger slide. These fingerings with different extents of hole

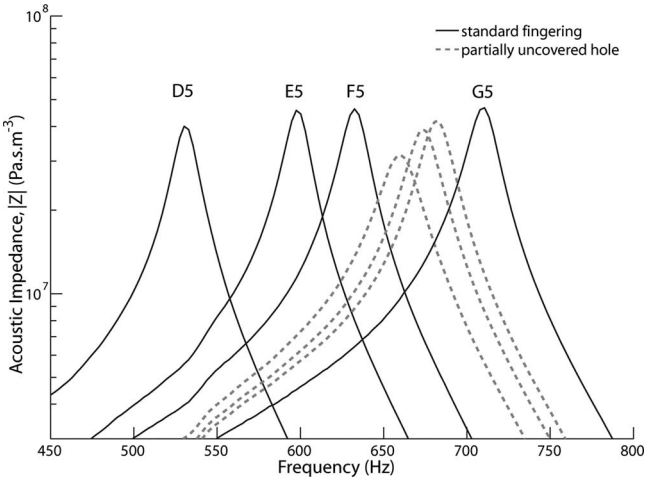


FIG. 5. The measured acoustic impedance at the mouthpiece of the clarinet (Z_{bore}) for different fingerings. The second maxima only shown here. Solid lines indicate the standard fingerings for the written pitches D5 (523 Hz), E5 (587 Hz), F5 (622 Hz), and G5 (698 Hz), and show maxima spaced at discrete frequencies corresponding to those notes. Dashed lines indicate a sequence of fingerings that use partial covering of the hole that is opened to change from F5 to G5.

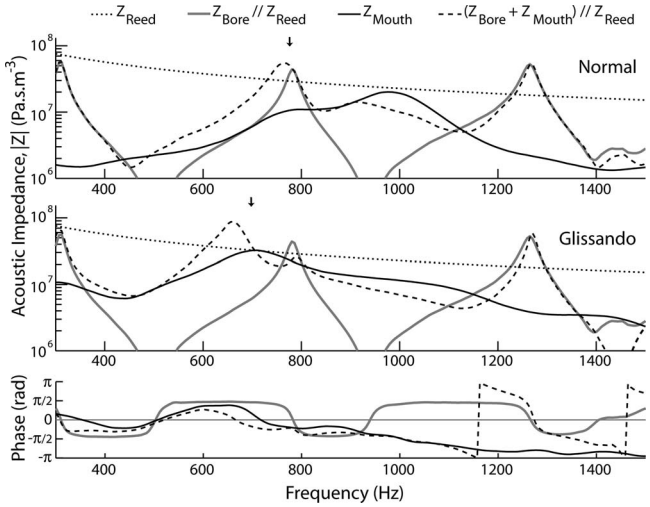


FIG. 6. Measured input impedance of the clarinet, Z_{bore}, shown here with the reed compliance in parallel, $Z_{bore} \| Z_{reed}$ (pale line), and the impedance measured in the mouth, Z_{mouth} (dark line). The impedance of the reed, Z_{reed} (dotted line), was calculated from the measured reed compliance. $Z_{load} = (Z_{mouth} + Z_{bore}) \| Z_{reed}$ is plotted as a dashed line. In both cases the fingering is for the note written A5 (784 Hz). Arrows indicate the frequency f_0 of the sounded note. The top graph shows the impedance magnitude for the note played normally. The middle graph shows that for the same fingering, as played in the *glissando* exercise. At this stage of the *glissando*, the sounding frequency is 76 Hz (190 cents, almost a whole tone) below that of the normal fingering. The bottom graph shows the phases of the impedances whose magnitudes are shown in the middle graph.

uncovering show impedance maxima at frequencies intermediate between those used for notes on the diatonic scale. The peaks, however, are lower than those for standard fingerings, particularly when the hole is nearly closed: The difference can be as large as tens of percent (Fig. 5). Thus the finger slide allows sounding pitch to increase smoothly by gradually raising the bore resonance, instead of moving in discrete steps as in a normal musical scale. Over part of the clarinet's range, this technique alone, with no vocal tract or embouchure adjustments, could contribute to the smoothly varying sounding pitch. However, musicians report that this is only one of the effects used to produce a *glissando*.

B. Vocal tract resonance and *glissandi*

Figure 6 shows the measured impedance of the clarinet bore (shown here with Z_{reed} in parallel), and that of a typical result for the impedance measured in the mouth (Z_{mouth}) during normal playing (top). For comparison, it also shows a typical measurement of the acoustic impedance in the mouth of a player performing a *glissando* (middle and bottom). This measured impedance in each case is simply added in series with the clarinet bore impedance, and then the effective reed impedance added in parallel to obtain an estimate of the effective acoustic impedance of the tract-reed-bore system according to the simple model represented by Eq. (1). In both cases, the player's fingers were fixed at the fingering used for A5.

In some cases, the impedance measured in the mouth Z_{mouth} is expected to be a good approximation to that of the vocal tract Z_{tract}. It is complicated, however, by the presence in the mouth of the reed and the acoustic component of vol-

ume flow past the reed. In the simple Benade (1985) model, the measured impedance would be Z_{tract} in parallel with $(Z_{bore}+Z_c)$, where Z_c is the combined impedance of the reed and the intermittent air gap beside it. The impedance of the reed, assumed largely compliant, is discussed above, and has a magnitude of a few tens of MPa s m^{-3} in the range of interest. The impedance through the intermittent gap includes both the inertance of the air in the gap and the flow resistance across it. For a typical blowing pressure of a few kilopascal and a cross section of about 10 mm^2, and considering only the Bernoulli losses, this is also some tens of MPa s m^{-3}. Despite their geometry, the two components are not simply in parallel, because when the reed's motion produces a volume flow into the bore, it also tends to reduce the aperture, and conversely.

Because Z_{mouth} is the parallel combination of Z_{tract} and (Z_c+Z_{bore}), it may sometimes be an underestimate of Z_{tract}, especially when Z_{tract} is large. An estimate of the lower bound for Z_c can be gained from the magnitude of Z_{mouth} measured when Z_{bore} is small. In Fig. 6, for example, Z_{mouth} is 17 MPa s m^{-3} at 942 Hz, when Z_{bore} is a minimum at 0.4 MPa s m^{-3}, over 40 times smaller. In the cases of pitch bending, Z_{mouth} is measured as high as 60 MPa s m^{-3} at frequencies when Z_{bore} is a few MPa s m^{-3}. Hence Z_{mouth} is likely to be a significant underestimate of Z_{tract} only when the latter is several tens of MPa s m^{-3}. For this reason, Z_{mouth} is simply shown in Fig. 6 as an estimate of Z_{tract} and is used in estimating Benade's (1985) effective impedance $Z_{load} \cong (Z_{mouth}+Z_{bore})\|Z_{reed}$.

In normal playing (top graph in Fig. 6), the magnitudes of the peak in Z_{mouth} (20 MPa s m^{-3} in this example) is about half that of the peaks of Z_{bore} (41 MPa s m^{-3} here), smaller than the effective impedance of the reed (~25 MPa s m^{-3} at these frequencies), and also smaller than those of the peak in $Z_{bore}\|Z_{reed}$ (44 MPa s m^{-3} here).[1] Consequently, according to the simple Benade (1985) model expressed in Eq. (1), the combined acoustic impedance for normal playing yields a resulting maximum determined largely by the maximum in Z_{bore}: The reed vibrates at a frequency (781 Hz) matching the strongest peak in $Z_{bore}\|Z_{reed}$ (which is close to the peak in Z_{bore}).

In the *glissando* exercise, however, the maximum impedance measured in the mouth is consistently comparable in magnitude with that of the maximum of the bore impedance and the effective impedance of the reed. The middle graph in Fig. 6 shows that, because here the peak in Z_{mouth} is no longer small compared with Z_{bore}, the peak in $(Z_{mouth}+Z_{bore})\|Z_{reed}$ is no longer determined solely by a peak in Z_{bore}. In this example, the Z_{mouth} maximum (32 MPa s m^{-3}) centered at 705 Hz is more comparable in magnitude with the corresponding $Z_{bore}\|Z_{reed}$ maximum (44 MPa s m^{-3}). Here, the sounding frequency during the *glissando* (indicated by an arrow) is about 76 Hz (190 cents or about one whole tone) lower than that produced for normal playing, while the peak in $(Z_{mouth}+Z_{bore})\|Z_{reed}$ is calculated to fall at 662 Hz, 119 Hz lower than the peak in $Z_{bore}\|Z_{reed}$ (781 Hz). In most of the *glissando* examples studied, sounding frequency f_0 did not coincide with the peak in $Z_{bore}\|Z_{reed}$ but instead occurred closer to the peak in $(Z_{mouth}+Z_{bore})\|Z_{reed}$. However, f_0 was usually about 10–40 Hz above the peak in $(Z_{mouth}+Z_{bore})\|Z_{reed}$. This difference might be due to the simplicity of the model used to derive Eq. (1). Also, a smaller value of Z_{reed} would give rise to a higher frequency for the peak in $(Z_{mouth}+Z_{bore})\|Z_{reed}$, so the difference might also be explained if the compliance of the reed in this situation were higher than in normal playing condition from which it was estimated. (The compliance of the reed could, in principle, be reduced by biting harder on the reed. However, players report no changes in the lip force during this exercise. Unfortunately, it would be difficult to determine independently the value of the reed compliance for the pitch bending playing condition, because Benade's (1976) technique assumes a non-negligible vocal tract impedance.)

As explained above, the clarinet's resonance dominates in normal playing and the player's tract has only a minor effect, as shown in the top part of Fig. 6. However, the region of the *glissando* in Rhapsody in Blue (from C5 to C6, written—i.e., 466–932 Hz) lies in the clarinet's second register, a range where the clarinet resonances have somewhat weaker impedance peaks than those in the lower register. In this range, experienced players can produce a resonance in the vocal tract whose measured impedance peak is comparable with or sometimes even larger in magnitude than those of the clarinet. Consequently, by tuning a strong resonance of the vocal tract and skillfully adjusting the fingering simultaneously, expert clarinettists can perform a *glissando* and smoothly control the sounding pitch continuously over a large pitch range. In performing this *glissando*, the sounding pitch here need only deviate from that of the fingered note by a semitone or so. However, greater deviations are possible. To study this, a simple pitch bending exercise was used.

C. Vocal tract resonance in normal playing and pitch bending

Figure 7 shows the resonance frequency measured in the mouth for the five test subjects as they played. It is plotted against the sounding pitch for both normal playing and pitch bending in the range between G4 (349 Hz) and G6 (1397 Hz). This plot shows the extent of vocal tract tuning: If the players tuned a resonance of the vocal tract to the note played, then the data would lie close to the tuning line $y=x$, which is shown as a gray line. If players maintained a constant vocal tract configuration with a weak resonance and the sounding pitch were determined solely by $(Z_{bore}\|Z_{reed})$, the data would form a horizontal line. The magnitude of the impedance peak is indicated on this graph by the size of the symbol used, binned in half decades as indicated by the legend.

Above about 600 Hz (written E5), the data for pitch bending (black circles) show clear tuning: The sounding frequency f_0 is always close to that of an impedance peak measured in the player's mouth. Below this frequency, the examples where the peaks in Z_{mouth} are large (indicated by large circles) also follow the tuning line. In the range below 600 Hz, examples of (intended) pitch bending with relatively small peaks (small black circles) sometimes deviate from the tuning line: In these cases, the player has not succeeded in

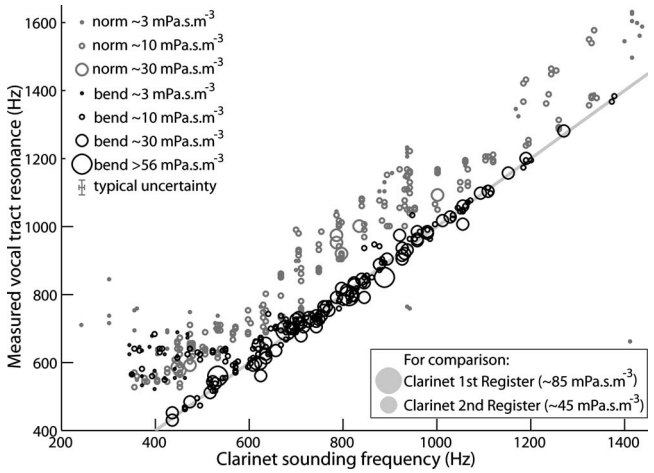


FIG. 7. Measured vocal tract resonance frequency plotted against clarinet sounding frequency. Data from the five players include both normal playing (pale circles) and pitch bending together with the *glissando* exercise (dark circles) in the range between written G4 (349 Hz) and G6 (1397 Hz). The size of each circle represents the magnitude of the acoustic impedance for that measurement, binned in half decade bands. For comparison two circles at bottom right show typical magnitudes of Z_{bore} for fingerings in the first and second registers of the clarinet. The gray line indicates the hypothetical relationship (frequency of vocal tract resonance)=(frequency of note played).

having the instrument play at the frequency determined by a resonance in the mouth. The legend shows, for comparison, the magnitude of the peaks in Z_{bore} for fingerings in the first and second registers of the clarinet. [For the purposes of this discussion, the second register can be defined to be the notes played using the second peak in impedance. Loosely speaking, these are the notes that use the second mode of the bore. Thus the second register goes from written B4 (440 Hz) to C6 (932 Hz).] Comparison with the size of the peaks in the bore and the tract impedance gives one reason why pitch bending is easier in the second register and higher, where peaks in Z_{bore} are smaller.

The range of frequencies over which the vocal tract is used for pitch bending in the second register of the clarinet (well within the normal range of the instrument) is comparable with the range for which Scavone *et al.* (2008) reported vocal tract effects for the alto saxophone (520–1500 Hz). This range is also comparable with that reported for vocal tract tuning in a study on tenor saxophones: To play in the very high (*altissimo*) range, saxophonists tune their vocal tract resonance (Chen *et al.*, 2008), but they do not do so in the normal range. However, the tenor saxophone is a tenor instrument and its *altissimo* range (above written F#6, sounding E5, 659 Hz) corresponds approximately to the upper second and third registers on the clarinet and to the range over which tract tuning is shown in Fig. 7.

The results for normal playing (gray symbols in Fig. 7) are more complicated. First playing at low frequencies (which were not the principal object of this study) is discussed. At frequencies below about 600 Hz, the results are approximately as expected for a configuration of the tract, which did not vary with pitch. Above about 600 Hz, however, and in normal playing, the resonance measured in the player's mouth occurs at frequencies about 150 Hz higher (on average) than the sounding pitch. The magnitudes of these vocal tract resonances formed during normal playing are modest ($|Z_{mouth}|$ about 9 MPa s m^{-3} on average) when compared with those of the operating clarinet impedance peak (~40–90 MPa s m^{-3}), so they contribute little to the series combination and thus are expected to have only a small effect on the sounding frequency of the reed; here the clarinet bore resonance dominates as expected (Nederveen, 1998; Dickens *et al.*, 2007b). Nevertheless, even though the players are not tuning their vocal tract to the note produced, they are adjusting it as a function of the note produced. Why might this be?

First it is noted that, in normal playing, a strong resonance of the vocal tract is not needed, to first order, to determine the sounding frequency f_0: Here the player can usually allow the clarinet bore resonance to determine, at least approximately, the appropriate sounding pitch. Indeed, calculations show that the magnitude and frequencies of these vocal tract impedance peaks change $(Z_{mouth}+Z_{bore})\|Z_{reed}$ by only several hertz at most. (While even a few hertz difference is important in accurate intonation, a raise in pitch over the whole range can be achieved by adjusting the mouthpiece on the barrel.) One possibility is that, in this range, experienced players learn, presumably implicitly, to keep their vocal tract resonance *away* from the sounding pitch to prevent it from interfering with the bore resonance during normal playing.

For f_0 below about 450 Hz, the resonances of the bore are stronger and unintended bending is less of a danger. In this range, players may keep the tract resonance at a constant frequency (near 600 Hz for these players). For the range 450 Hz to at least 1400 Hz, they raise the frequency of their tract resonance to keep it substantially above that of the bore. As will be discussed below, it is easier to bend a note down than up on the clarinet. Perhaps having a tract resonance "nearby" (100–200 Hz away) makes for a good performance strategy: Tuning assistance from the vocal tract can be quickly and easily engaged by adjusting the resonance frequency and strength appropriately, should the need arise. And perhaps having a resonance slightly below the played note is just too dangerous, because of the potential effects on the pitch, which will be discussed later. This strategy of keeping the tract resonance at a frequency somewhat above that of the bore resonance for normal playing may explain the results of Clinch *et al.* (1982), who observed a gradual variation of vocal tract shape with increasing pitch over the range of notes studied. These researchers used x-ray fluoroscopy to study the vocal tract during playing and concluded that players were tuning the tract resonance to match the note played. However, as this technique can only give qualitative information about the tract resonance, it is possible that the subject of their study was also keeping the tract resonance frequency somewhat above that of the note played.

In contrast to the results for normal playing, the measurements made during pitch bending show tight tuning of the sounding pitch to the vocal tract resonance, the difference in frequency being typically less than 30 Hz. Here, a strong resonance measured in the mouth (average $|Z_{mouth}|$ ~20 MPa s m^{-3}) is generated by the player and competes with the clarinet bore resonance. This changes (Z_{mouth}

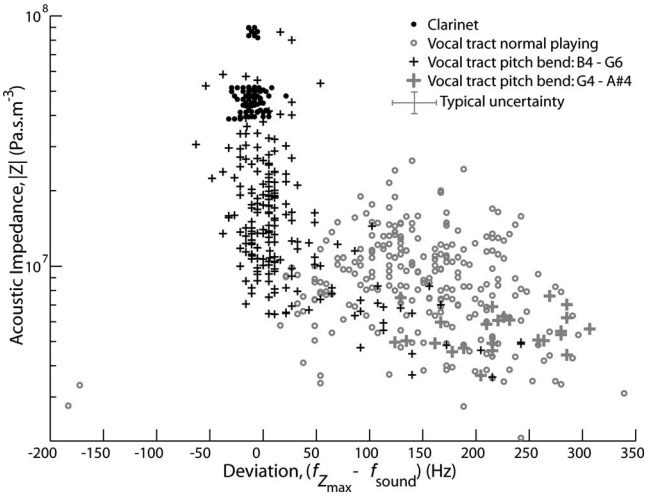


FIG. 8. The magnitudes of the maxima in acoustic impedance of $Z_{bore} \| Z_{reed}$, the clarinet bore in parallel with the reed (dark dots) and of that measured in the player's mouth, Z_{mouth}, are plotted as a function of the difference between the frequency of the relevant maximum and that of the pitch played: In other words as a function of the deviation of resonance frequency from the sounding pitch. Crosses indicate pitch bending together with the *glissando* exercise; open circles indicate normal playing.

$+Z_{bore}) \| Z_{reed}$ and, as predicted by the simple model, the resonance frequency of the player's vocal tract begins to influence the sounding frequency of the reed (normally determined by the bore resonance). This can be observed for sounding notes above 600 Hz, in agreement with Rehfeldt (1977) who suggested that the lower limit to large pitch bending on the clarinet lies about D5 (587 Hz). This would also explain why the *glissando* is usually only played over the last several notes of the scale in Rhapsody in Blue.

Below written E5 (~600 Hz), there is less strict tuning of vocal tract resonance. This might be because it is difficult to produce a vocal tract resonance with a sufficiently large impedance peak at frequencies below this range. Scavone *et al.* (2008) placed the lower limit for adjusting the relevant vocal tract influence at about 520 Hz. Further, clarinet bore resonances in this lower playing range are rather stronger. However, although the extent of pitch bending using the vocal tract resonance is limited in this range, other strategies are used, including partial uncovering of tone holes and techniques that are not studied here, such as changing the bite force on the reed and adjusting lip damping.

Figure 8 plots the same data used for Fig. 7 to show the magnitude of Z_{mouth} and $Z_{bore} \| Z_{reed}$ explicitly. Here, these quantities are plotted as a function of the deviation of the impedance maximum from sounding frequency f_0. For $Z_{bore} \| Z_{reed}$ (black dots), two tight clusters of data are seen. One cluster is those of the bore resonances involved in producing notes in the clarinet's first register: These have magnitudes of about 90 MPa s m^{-3}, the other corresponds to the resonances that produce the clarinet's second register (about 40–50 MPa s m^{-3}). These bore resonances are of course centered on the normal sounding pitch.

For the vocal tract resonances, two contrasting regimes are seen in Fig. 8: For vocal tract resonances with impedance peaks above about 20–25 MPa s m^{-3}, the sounding frequency is tuned closely to the resonance frequency measured in the mouth, with typically less than 30 Hz deviation. Only measurements made during pitch bending fall into this region. For tract resonances with smaller magnitude, the sounding frequency is not necessarily tuned to the vocal tract resonance.

Normal playing (light circles) over this pitch range (G4–G6) shows a broad scattering of weak mouth resonances that deviate from the sounding frequency by typically 100–200 Hz. The asymmetry is striking: They are nearly always above the sounding frequency. These weak tract resonances ($|Z_{tract}| \ll |Z_{bore}|$) do not affect the correct sounding pitch. For pitch bending (crosses), however, a strong vocal tract resonance can influence the sounding frequency. When the peak in Z_{mouth} is large (for notes in the high range of Z: dark crosses), the tract resonance dominates and consequently there is little deviation from the sounding pitch. The pitch bending points in the right of the graph largely correspond to the lower range of the clarinet (below about 600 Hz—pale crosses) where bore resonances are very strong (dots, top left). It may be that it is difficult to produce a vocal tract resonance with a strong peak in this frequency range. Here, the tract resonances are weak and do not determine the sounding frequency directly.

D. Example: Pitch bending exercise on fingered C6

To elucidate the relative influence of bore, tract, and reed on combined impedance and sounding frequency, players were further asked to finger a standard note (written C6, 932 Hz) and to bend its sounding pitch progressively down from the standard pitch while their vocal tract impedance was measured. Players were able to bend the normal pitch C6 (932 Hz) smoothly down by as much as a major third to G#5, a deviation of 400 cents or a third of an octave. This is similar to the average maximum downward pitch bend of 330 cents found by Scavone *et al.* (2008) for the alto saxophone.

Figure 9 shows calculations of the combined impedance ($Z_{mouth}+Z_{bore}$) for bending a note down. A single measured clarinet impedance spectrum (for C6 fingering) is added in series to three different impedance spectra measured in a player's vocal tract for normal playing and two varying degrees of pitch bending, all using the same fingering. To increase pitch bending, the tract resonance moves to successively lower frequencies. It also has successively increased magnitude. The resulting decrease in frequency of the maximum in the series impedance correlates with successively lower sounding frequencies (indicated by the arrows).

E. Why is it easier to bend pitch down rather than up?

On the clarinet and saxophone, it is possible to bend pitches down, whereas on the clarinet "only slight upward alterations are possible" (Rehfeldt, 1977) and similarly "significant upward frequency shifts… are not possible" on the alto saxophone (Scavone *et al.*, 2008). It can be shown here that, according to the simple model of Benade (1985), this is simply because although X_{reed} is always compliant, Z_{bore} can be either inertive or compliant.

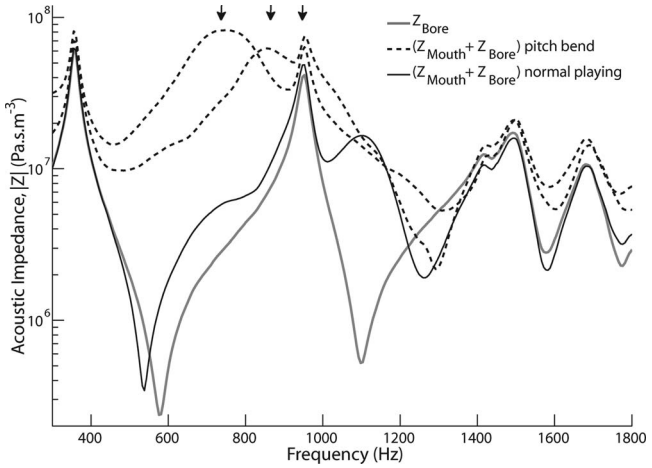


FIG. 9. The clarinet bore impedance (Z_{bore}) for the fingering C6 (gray line) is shown with the series impedance ($Z_{mouth}+Z_{bore}$) for normal playing (solid line) and increasing degrees of pitch bending (dashed lines), all while maintaining the same fingering for C6 (932 Hz). Vertical arrows indicate the sounding pitch for the three cases: The right-hand arrow shows a normal sounding pitch at C6+8 cents (937 Hz), while the left-hand arrow denotes the lowest pitch sounding at G#5+7 cents (743 Hz), a deviation of 400 cents or a major third.

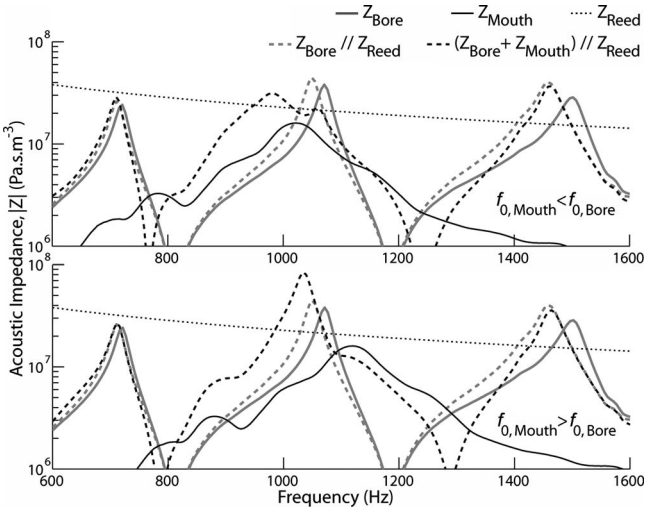


FIG. 10. To examine the effect of vocal tract resonances tuned above and below that of the bore, two hypothetical vocal tract values of equal magnitude (11 MPa s m^{-3}) are shown (dark lines), with resonance frequencies 1020 (shown above) and 1120 Hz (shown below), along with Z_{reed} (dotted line), Z_{bore} (peak at 1070 Hz, pale line), and $Z_{bore}\|Z_{reed}$ (peak at 1050 Hz, pale dashes). Total impedance of the tract-reed-bore system ($Z_{bore}+Z_{mouth}$)$\|Z_{reed}$ for both cases are shown (dark dashes), with respective maxima at 980 (above) and 1035 Hz (below).

Nederveen (1998) and the experiments reported here give an effective clarinet reed compliance C (in a typical clarinet embouchure on a reed of hardness 3) as about 7×10^{-12} m^3 Pa^{-1} (equivalent to an air volume of 1.1 ml). At a frequency of 1 kHz, this gives a reactance X_{reed} ($=-1/2\pi fC$) of about -20 MPa s m^{-3}. Its dependence on frequency is weak compared with that of the tract and bore impedances near resonances. For the purposes of this simple model, and as argued above, the sounding frequency f_0 occurs near the maximum in $Z_{load}=(Z_{tract}+Z_{bore})\|X_{reed}$, where the reactance (i.e., the imaginary part) is zero, i.e., when

$$X_{tract}(f)+X_{bore}(f)=-X_{reed}(f). \quad (2)$$

Because X_{reed} is always negative and moderately large, the net sign of $(X_{tract}+X_{bore})$ must always be positive (and equally large) for resonance to occur.

In normal playing, Z_{tract} is generally small in comparison with the maxima in Z_{bore}, and initially, its effect can be neglected. The condition that $X_{bore}(f)=-X_{reed}(f)$ requires that X_{bore} be positive and so the sounding frequency f_0 must lie on the low frequency (inertive) side of the resonance peak in Z_{bore}. A soft reed or a more relaxed embouchure will produce a larger compliance C, a decrease in X_{reed}, and consequently a decrease in sounding frequency.

Now, consider the effect of including Z_{tract} with maxima of similar magnitude to that of Z_{bore} and, initially, at the same resonance frequency, which will be greater than f_0. If the resonant frequencies of Z_{tract} and Z_{bore} are similar, both $X_{tract}(f)$ and $X_{bore}(f)$ will be positive at frequencies below their resonances, including f_0. If the magnitude of the resonance in Z_{tract} is now increased, $X_{tract}(f)$ will also increase and consequently the sum $X_{tract}(f)+X_{bore}(f)$ will increase and so the sounding frequency f_0 will decrease so that X_{reed} satisfies Eq. (2). If the player now decreases the resonance frequency of the tract, the value of $X_{tract}(f)$ around f_0 will increase and so f_0 must again decrease to satisfy Eq. (2). However, decreasing f_0 by continuing to decrease the resonance frequency of the tract will become increasingly difficult. This is because the contribution of $X_{bore}(f)$ to the sum $X_{tract}(f)+X_{bore}(f)$ will decrease as f_0 moves further away from the resonance frequency of Z_{bore} and yet $X_{tract}(f)+X_{bore}(f)$ must increase to match the increase in $X_{reed}(f)$ as f_0 decreases. Eventually a player will be unable to increase $Z_{tract}(f)$ sufficiently to match $X_{reed}(f)$ and further downward pitch bending will not be possible.

The situation is quite different, however, if a player wishes to bend the pitch upwards. Again, imagine a vocal tract resonance, comparable in magnitude to that in the bore, and with initially the same resonance frequency as the bore, again above f_0. If the resonant frequency of the tract is then increased, the sum of $X_{tract}(f)+X_{bore}(f)$ in the frequency range where this sum is inertive will decrease and consequently f_0 will increase as predicted by Eq. (2). However, once f_0 exceeds the resonance frequency of the bore, $X_{bore}(f)$ will suddenly change sign and the sum $X_{tract}(f)+X_{bore}(f)$ will decrease dramatically to a value much smaller than possible for $X_{reed}(f)$. Players are probably unable to increase $Z_{tract}(f)$ sufficiently to increase f_0 past this point, unless they can produce a peak in Z_{tract} that is significantly greater than that in Z_{bore}, which Fig. 8 shows is relatively rare. (In those rare cases where the player can produce such a peak in Z_{tract}, then f_0 is determined largely by the tract and depends less strongly on the bore, much as is the case in the altissimo region of the saxophone.) Thus, in normal situations, the maximum increase in sounding frequency will be of the same order in magnitude as the decrease in resonance frequency of the bore due to the compliance of the reed, possibly not more than 50 cents.

What then are the effects of vocal tract resonances tuned above and below that of the bore? Figure 10 shows the impedance of Z_{bore} for the note written D6, with a peak at 1070

Hz, and that of $Z_{bore} \| Z_{reed}$, which has a peak at 1050 Hz (near the nominal frequency, 1047 Hz, for that note). A single measured vocal tract impedance spectrum Z_{mouth} was then numerically shifted in frequency so that the "same" tract impedance peak now lay at either 50 Hz above or 50 Hz below the peak in 1070 Hz. Then $(Z_{bore}+Z_{mouth}) \| Z_{reed}$ is plotted for the two cases. In both cases, the frequency of the peak $(Z_{bore}+Z_{mouth}) \| Z_{reed}$ lies below that of the peak in $Z_{bore} \| Z_{reed}$, but the downward pitch bend is larger for the lower frequency tract resonance. This figure also shows what happens when Z_{mouth} is much smaller than Z_{bore}, because this is the case for the other two bore resonances that occur in the frequency range shown. Finally, it is worth observing another peak in Z_{bore}, that at about 700 Hz. At this frequency, Z_{mouth} is small compared with Z_{bore}, and both are small compared with Z_{reed}; consequently, the peak in $(Z_{bore}+Z_{mouth}) \| Z_{reed}$ here nearly coincides with that in Z_{bore}.

IV. CONCLUSION

For normal clarinet playing, resonances in the clarinet bore (determined by the fingering used) dominate to drive the reed to oscillate at a frequency very close to that of the bore and reed in parallel. However, if the upstream resonance in the player's vocal tract is adjusted to have a sufficiently high impedance peak at the appropriate frequency, the vocal tract resonance competes with or dominates the clarinet resonance to determine the reed's sounding frequency.

By skillfully coordinating the fingers to smoothly uncover the clarinet finger holes and simultaneously tuning strong vocal tract resonances to the continuously changing pitch, expert players are able to facilitate a smooth trombone-like *glissando*, of which a famous example is the final octave of the run that opens Gershwin's Rhapsody in Blue.

ACKNOWLEDGMENTS

The authors thank the Australian Research Council for support of this project and Neville Fletcher for helpful discussions. They thank Yamaha for the clarinets, Légère for the synthetic reeds, and the volunteer clarinettists.

[1]The peak in $Z_{bore} \| Z_{reed}$ has a larger magnitude than that of either Z_{bore} or Z_{reed} because it is a parallel resonance between the reed compliance and the bore in its inertive range, i.e., at frequencies a little below the peak in Z_{bore}.

Backus, J. (**1961**). "Vibrations of the reed and the air column in the clarinet," J. Acoust. Soc. Am. **33**, 806–809.

Backus, J. (**1963**). "Small-vibration theory of the clarinet," J. Acoust. Soc. Am. **35**, 305–313.

Backus, J. (**1974**). "Input impedance curves for the reed woodwind instruments," J. Acoust. Soc. Am. **56**, 1266–1279.

Backus, J. (**1985**). "The effect of the player's vocal tract on woodwind instrument tone," J. Acoust. Soc. Am. **78**, 17–20.

Benade, A. H. (**1976**). *Fundamentals of Musical Acoustics* (Oxford University Press, New York), pp. 465–467.

Benade, A. H. (**1985**). "Air column, reed, and player's windway interaction in musical instruments," in *Vocal Fold Physiology, Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. C. Scherer (Denver Center for the Performing Arts, Denver, CO), Chap. 35, pp. 425–452.

Benade, A. H., and Ibisi, M. I. (**1987**). "Survey of impedance methods and a new piezo-disk-driven impedance head for air columns," J. Acoust. Soc. Am. **81**, 1152–1167.

Boutillon, X., and Gibiat, V. (**1996**). "Evaluation of the acoustical stiffness of saxophone reeds under playing conditions by using the reactive power approach," J. Acoust. Soc. Am. **100**, 1178–1889.

Chen, J.-M., Smith, J., and Wolfe, J. (**2008**). "Experienced saxophonists learn to tune their vocal tracts," Science **319**, 726.

Chen, J.-M., Smith, J., and Wolfe, J. (**2009**). "Saxophone acoustics: Introducing a compendium of impedance and sound spectra," Acoust. Aust. **37**, 18–23.

Clinch, P. G., Troup, G. J., and Harris, L. (**1982**). "The importance of vocal tract resonance in clarinet and saxophone performance, a preliminary account," Acustica **50**, 280–284.

Dickens, P., Smith, J., and Wolfe, J. (**2007a**). "High precision measurements of acoustic impedance spectra using resonance-free calibration loads and controlled error distribution," J. Acoust. Soc. Am. **121**, 1471–1481.

Dickens, P., France, R., Smith, J., and Wolfe, J. (**2007b**). "Clarinet acoustics: Introducing a compendium of impedance and sound spectra," Acoust. Aust. **35**, 17–24.

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer, New York), pp. 470–480.

Fritz, C., and Wolfe, J. (**2005**). "How do clarinet players adjust the resonances of their vocal tracts for different playing effects?," J. Acoust. Soc. Am. **118**, 3306–3315.

Grand, N., Gilbert, J., and Laloë, F. (**1996**). "Oscillation threshold of woodwind instruments," Acust. Acta Acust. **82**, 137–151.

Johnston, R., Clinch, P. G., and Troup, G. J. (**1986**). "The role of the vocal tract resonance in clarinet playing," Acoust. Aust. **14**, 67–69.

Nederveen, C. J. (**1998**). *Acoustical Aspects of Wind Instruments*, 2nd ed. (Northern Illinois University, De Kalb, IL), pp. 35–37.

Pay, A. (**1995**). "The mechanics of playing the clarinet," in *The Cambridge Companion to the Clarinet*, edited by C. Lawson (Cambridge University Press, Cambridge), pp. 107–122.

Rehfeldt, P. (**1977**). *New Directions for Clarinet* (University of California Press, Berkeley, CA), pp. 57–76.

Scavone, G., Lefebre, A., and da Silva, A. R. (**2008**). "Measurement of vocal-tract influence during saxophone performance," J. Acoust. Soc. Am. **123**, 2391–2400.

Scavone, G. P. (**2003**). "Modeling vocal-tract influence in reed wind instruments," in Proceedings of the 2003 Stockholm Musical Acoustics Conference, Stockholm, Sweden, pp. 291–294.

Schwartz, C. (**1979**). *Gershwin: His Life and Music* (Da Capo, New York, NY), pp. 81–83.

Silva, F., Kergomard, J., Vergez, C., and Gilbert, J. (**2008**). "Interaction of reed and acoustic resonator in clarinetlike systems," J. Acoust. Soc. Am. **124**, 3284–3295.

Smith, J. R., Henrich, N., and Wolfe, J. (**1997**). "The acoustic impedance of the Boehm flute: Standard and some non-standard fingerings," Proc. Inst. Acoustics **19**, 315–330.

Sommerfeldt, S. D., and Strong, W. J. (**1988**). "Simulation of a player-clarinet system," J. Acoust. Soc. Am. **83**, 1908–1918.

Tarnopolsky, A., Fletcher, N., Hollenberg, L., Lange, B., Smith, J., and Wolfe, J. (**2006**). "Vocal tract resonances and the sound of the Australian didjeridu (yidaki) I: Experiment," J. Acoust. Soc. Am. **119**, 1194–1204.

Watkins, M. (**2002**). "The saxophonist's vocal tract," The Saxophone Symposium: J. North Am. Saxophone Alliance **27**, 51–75.

Wilson, T. A., and Beavers, G. S. (**1974**). "Operating modes of the clarinet," J. Acoust. Soc. Am. **56**, 653–658.

Wilson, T. D. (**1996**). "The measured upstream impedance for clarinet performance and its role in sound production," Ph.D. thesis, University of Washington, Seattle, WA.

The kinetics and acoustics of fingering and note transitions on the flute

André Almeida,[a] Renee Chow, John Smith, and Joe Wolfe
School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia

(Received 12 March 2009; revised 17 June 2009; accepted 22 June 2009)

Motion of the keys was measured in a transverse flute while beginner, amateur, and professional flutists played a range of exercises. The time taken for a key to open or close was typically 10 ms when pushed by a finger or 16 ms when moved by a spring. Because the opening and closing of keys will never be exactly simultaneous, transitions between notes that involve the movement of multiple fingers can occur via several possible pathways with different intermediate fingerings. A transition is classified as "safe" if it is possible to be slurred from the initial to final note with little perceptible change in pitch or volume. Some transitions are "unsafe" and possibly involve a transient change in pitch or a decrease in volume. Players, on average, used safe transitions more frequently than unsafe transitions. Delays between the motion of the fingers were typically tens of milliseconds, with longer delays as more fingers become involved. Professionals exhibited smaller average delays between the motion of their fingers than did amateurs.
 [DOI: 10.1121/1.3179674]

PACS number(s): 43.75.Qr, 43.75.St [NHF] Pages: 1521–1529

I. INTRODUCTION

In wind instruments, a transition between two successive notes often requires the coordinated movement of two or more fingers (for simplicity, all digits including thumbs will be referred to as fingers). One of the reasons why players practice scales, arpeggios, and exercises is to learn to make smooth and well-controlled transitions between notes without undesired transients. For players, this motivation is particularly important for slurred notes, where the transition involves no interruption to the flow of air and should ideally produce no wrong notes and no detectable silence between the notes.

In practice, finger movements are neither instantaneous nor simultaneous, and it takes a finite time to establish a new standing wave in the instrument bore. Slurred transitions involving the motion of only a single finger can produce transients that result from the finite speed of the finger that pushes a key in one direction or of the spring that returns it to its rest position. For transitions involving the motion of two or more fingers, there can be an additional transient time due to the time differences between the movements of each finger, which invites the question: How close to simultaneous can flutist finger movements be, and are there preferred finger orders in particular note transitions?

Although previous studies have monitored finger motion on the flute, they have been concerned with the flute as a controller for electronic instruments. The musical instrument digital interface (MIDI) flute developed at IRCAM initially used optical sensors, but the final version used Hall effect sensors with magnets attached to the keys (Miranda and Wanderley, 2006). The "virtually real flute" (Ystad and Voinier, 2001) used linear Hall effect sensors and could detect the speed of key transitions. The "hyper flute" (Palacio-Quintin, 2003) employed a large number of sensors, but only two keys had linear Hall effect sensors. Palmer *et al.* (2007) used infrared tracers attached to a player's fingernails and recorded their motion with a video camera. Although suitable for detecting the broad gestures of a player, this approach does not provide sufficient resolution (in either space or time) to monitor the detailed fingering behaviors occurring in note transitions.

This paper reports explicit measurements of the times taken for keys to open and to close under the action of fingers and springs and determines the key order and relative timing in transitions involving multiple fingers. The flute was chosen partly because of the similarity in size and construction of most of its keys, which means that similar sensors could be used for each. These sensors monitored the position of each key using reflected, modulated infrared radiation and had the advantage that they did not alter the mass of keys nor affect their motion. The flute has the further advantage that measured acoustic impedance spectra are available for all standard fingerings (Wolfe *et al.*, 2001), in addition to acoustical models of all possible fingerings (Botros *et al.*, 2002).

II. SOME BACKGROUND IN FLUTE ACOUSTICS

In most woodwind instruments, the played note is determined in part by the combination of open and closed holes in the side of its bore, which is called a fingering. Each fingering produces a number of resonances (corresponding to extrema in the input impedance), one or more of which can be excited by a vibrating reed or air jet. On many modern woodwinds, there are more holes in the instrument than fingers on two hands. Consequently, some keys operate more than one tone hole, often using a system of clutches, and some fingers are required to operate more than one key.

[a]Author to whom correspondence should be addressed. Electronic mail: aalmeida@phys.unsw.edu.au

The acoustical impedance spectrum of the flute for a particular fingering can be predicted by acoustical models, and important details of its behavior can be deduced from this. The "virtual flute" is a web service using such an acoustical model to predict the pitch, timbre, and ease of playing (Botros *et al.*, 2006). This service, however, does not yet give indications on the playing difficulties associated with transitions between two different fingerings.

The embouchure of the flute is open to the air, and so the instrument plays approximately at the minima in the input impedance $Z(f)$ at the embouchure. The player selects between possible minima by adjusting the speed of the air jet (Coltman, 1976), and consequently a periodic vibration regime is established with fundamental frequency close to that of a particular impedance minimum or resonance. Fine tuning is achieved by adjusting the jet speed or rotating the instrument slightly, which has the effect of changing the jet length, the occlusion of the embouchure hole, and thus the solid angle available for radiation, thereby modifying the acoustical end effect. Changing from one fingering to another usually changes most of the frequencies of the bore resonances and consequently also the note played. de la Cuadra *et al.* (2005) discussed flute control parameters in detail.

A simple fingering is one in which all of the tone holes above a point are closed and all (or most) of those below are open. For low notes and instruments with large tone holes, the acoustic wave is rather weaker downstream from the first open tone hole, so the length from the embouchure hole to this first open hole determines approximately the effective length of the bore. Simple fingerings usually have several strong resonances whose frequencies are in approximately harmonic ratios. For low notes, these nearly harmonic resonances are excited by the nonlinearity in the air jet. For complex fingerings, including some of those that arise briefly in rapid transitions between notes, the resonances are often weaker, and their frequencies are not in simple ratios. These issues are discussed in greater detail by Wolfe and Smith (2003).

A. The transitions between notes

In some cases, no fingering changes are required: The player can select among different resonances by adjusting the speed and length of the jet at the embouchure. Thus, the standard fingering for F4 is used by players to play the notes F4 and F5, (and can also play C6, F6, A6, and C7).

Many of the transitions between two notes separated by one or two semitones involve moving only a single finger. Provided that fingers or springs move the key or keys sufficiently quickly, one would expect no strong transients when slurring such transitions. Small transient effects can always arise because of the fact that the strength of the resonances in the bore of the flute is somewhat reduced when a key is almost but not completely closed (an example is given in Fig. 4).

Several mechanisms can produce undesirable transients in note transitions. One of particular interest may occur when a slurred transition involves the motion of two or more fingers. The speed of moving keys is limited by the speed of fingers in one direction and of the key springs that move them in the opposite direction. Further, the acoustic effects produced by the motion of a key are not linear functions of key displacement, so it is difficult to define simultaneous motion, particularly for keys moving in opposite directions. In practice, fingers will always move at slightly different times and with different speeds (how different are these times is one of the questions addressed here). This means that there are several possible intermediate discrete key configurations as well as continuous variations during transitions between two notes that involve the motion of more than one finger. Furthermore, these different intermediate states may have quite different acoustic properties, which are not necessarily intermediate between those of the initial and final key configurations.

B. Safe and unsafe transitions

Intermediate fingerings may be divided into four categories.

Safe. If the relevant minimum of the input impedance $Z(f)$ lies at a frequency close to (or intermediate between) those of the initial and final states and has similar magnitudes, then a steady oscillation of the jet can be maintained during a slurred transition. A transition that passes transiently through such a fingering can be called a safe transition and is illustrated schematically in Fig. 1. This case is discussed in more detail in Sec. IV B.

Unsafe (detour). If one of the intermediate states exhibits a minimum in $Z(f)$ at a frequency unrelated to both notes in the desired transition, it may cause an undesired note to sound briefly during the transition—see Fig. 1. Although $Z(f)$ of the flute is only valid for stationary geometric configurations, one can suppose that a transient geometric configuration is quasi-steady if the intermediate state is long enough. If the transition time is faster than a few periods of the initial or final steady states, the quasi-steady approach is, of course, not valid. For the flute, the period lies between 0.4 and 4 ms.

Unsafe (dropout). If there are no deep minima in $Z(f)$ at frequencies close to those of the initial and final states, the jet oscillation may not be maintained during a slurred transition because the jet length and speed are inappropriate for the frequency of the nearest minimum. In this case, the intensity of the tone will decrease substantially. Figure 2 shows an example of a note transition for which one of the intermediate fingerings has a weak resonance.

Unsafe (trapped). If the second fingering has, in addition to the desired resonance, a strong resonance at a frequency close to that of the first note, the latter may "trap" the jet. Botros *et al.* (2003) gave examples of this situation.

In some cases, such as the C6 to D6 transition (also analyzed later), there is no safe intermediate fingering so, unless fingers move nearly simultaneously, audible transients are expected. Of course, even for this definition of safe transitions, transients are expected in the flute sound: It takes time for a wave to travel down the bore, to reflect at an open tone hole, and to return, and several such reflections may be

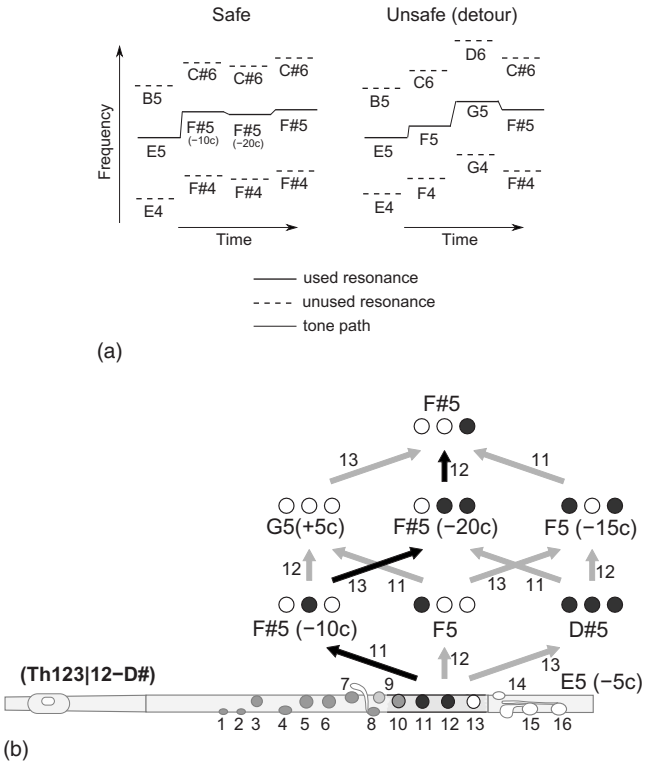


FIG. 1. A schematic example of a safe transition and an unsafe (detour) transition from E5 to F♯5. In the safe transition, all intermediate fingerings produce notes with a pitch very close to that of the initial or final note. In the unsafe (detour) transition, the tone is not interrupted, but the transitory notes are not close to E5 or F♯5. The lower frame shows possible intermediate key states and transitions in that transition. The safe paths are shown with dark arrows and labeled with number of the key that moves. White circles indicate open tone holes, black indicates holes closed by a key directly under the finger, and gray shows those closed indirectly by the mechanism. The legends in parentheses show a common notation for these fingerings (the pitch presented with each fingering is estimated from the measured impedance spectrum of the flute).

required to establish a standing wave with a new frequency. Finally, it should be remembered that some transients are an important part of the timbre of wind instruments and may be expected by musicians and listeners.

III. MATERIALS AND METHODS

A. Monitoring the key positions

An optical method was chosen because, unlike magnetic systems, there was no need to attach magnets or other materials to the keys and thus alter their mass or feeling under the fingers. Mechanical contacts suffer reliability problems and exhibit bounce and/or hysteresis.

A reflective photosensor was glued below the edge of each key so that the intensity of light reflected from the edge of the key increased monotonically as the key was closed (see Fig. 3). The chosen sensors (Kodenshi SG-2BC) were small (4 mm diameter), had low mass (160 mg), and combined an infrared light emitting diode (LED) with a high-sensitivity phototransistor (peak sensitivity at 940 nm). The sensor signal was distinguished from the background illumination by modulating the output of the LED in the sensor with a 5 kHz sine wave. The phototransistor in the sensor was connected as an emitter follower with a filter to remove

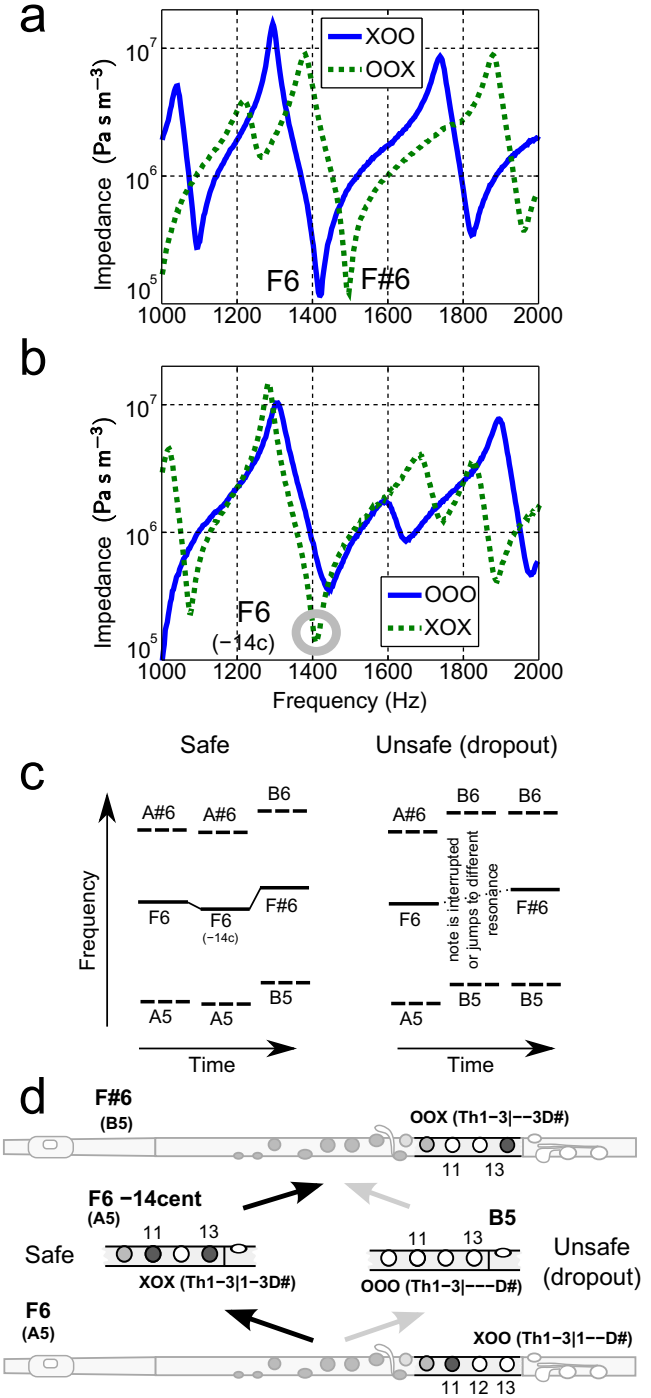


FIG. 2. (Color online) Schematics illustrating a safe and an unsafe (dropout) transition from F6 to F♯6. (a) The relevant region of the input impedance spectra measured for F6 and F♯6. (b) Measured spectra of the two possible intermediate fingerings in the same region, with the nearby playable minima circled. (c) Possible transition paths via these intermediate fingerings are shown in a schematic graph of frequency vs time. (d) Fingerings involved. The keys controlled by the right hand are emphasized, and the rest of the flute is shown pale. The numbering scheme used for the keys is shown in the bottom drawing.

dc variations due to changes in lighting conditions. Because the background illumination and degree of shading can vary for each experiment, the dc bias was adjustable to provide adequate dynamic range for the 5 kHz signal without clipping. This was set using a separate eight element bar LED display for each key. This procedure removes most, but not

FIG. 3. An author (AA) demonstrates the modified flute used for this work. The 50 wire ribbon cable that connects the flute to the sensor electronics is visible below the flute. The microphone can be seen in the lower left corner. The inset shows how a photosensor is mounted below a key.

all, of the dependence on background illumination: A small component remains because of nonlinearities. The sensor signals from 16 keys and the sound were recorded on a computer using two MOTU 828 firewire audio interfaces (24 bit at 44.1 kHz). Because the same hardware was used to sample both the sound and the output of the key sensors, the authors can be certain that any delays in sampling due to latencies and buffering will be identical and, consequently, will cancel exactly when timing differences are calculated.

The sensor output as a function of position was measured in experiments in which the sensor output was recorded, while a key was slowly closed using a micrometer screw. These showed that the amplitude of the modulated output from the sensor was a monotonic but nonlinear function of key position. In a further series of experiments, the flute was played by a blowing machine, while a key was slowly closed by the micrometer. The frequency and sound level are plotted as a function of sensor output in Fig. 4. The

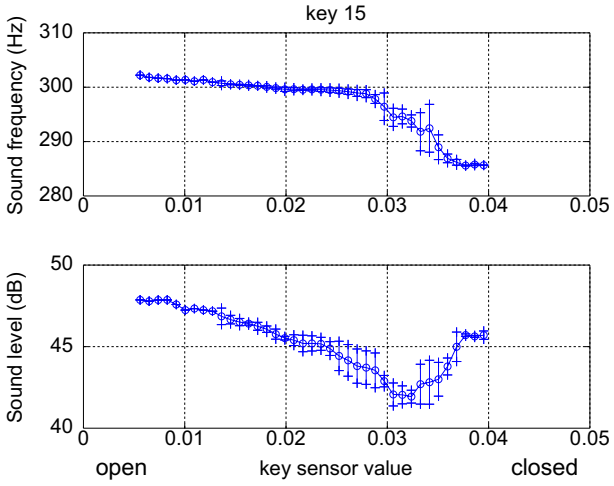


FIG. 4. (Color online) The variation in sound frequency and level produced by a flute when a key is slowly closed. The air jet was provided by a blowing machine. The saturation of the frequency (top graphic) with increasing sensor value at the right hand side of the curve is the result of continuing compression of the key pad with no further acoustic effect. The bottom graphic shows how the intensity is reduced when the pad is almost but not completely closed. Error bars indicate the standard deviation for six trials.

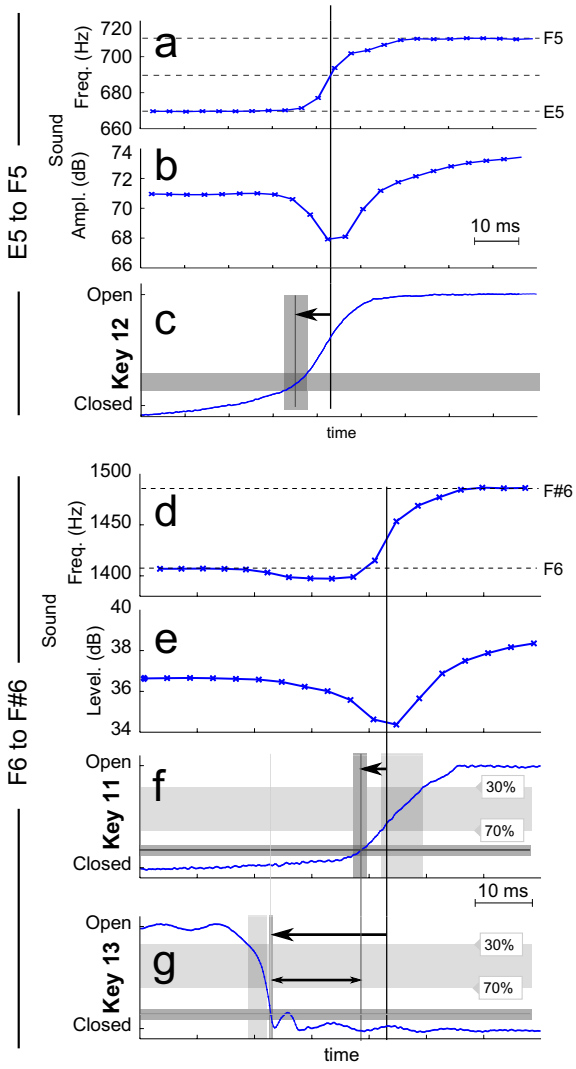


FIG. 5. (Color online) Examples of parameter extraction from the measured sound and key sensor output. Typical measurements of the (a) frequency, (b) sound level, and (c) key sensor output are extracted for a single finger transition from E5 to F5, corresponding to the movement of a single key. The horizontal dark shaded band shows the uncertainty in the key transition value, and consequently the vertical shaded band shows the uncertainty in the time of transition. The horizontal arrow shows the offset between effective key opening and the midpoint of the frequency transition. A two finger transition from F6 to F♯6 is shown in (d)–(g). Their sensor values are shown in (f) and (g). The dark shaded bands again show the uncertainty in the key transition value and time. The pale shading shows the time between 30% and 70% of the key sensor value, which is discussed in the text. Again, the arrows show the interval between note transition and effective key opening/closing. The difference between these arrows gives the delay between the two keys (shown with smaller headed arrows), here about 15 ms.

shape of this plot is explained by the compressibility of the key pad. Once the pad seals on the rim of the tone hole chimney, the playing frequency reaches its lowest value and remains unchanged as the pad is compressed while the key is further depressed. The sound level is high when the key is fully open and also when the key is closed with the pad compressed. The sound level is lower, however, when the hole is incompletely closed by the uncompressed pad, which is presumably a consequence of leaks between the pad and the rim; this can be seen in Fig. 5. This can be related to the work of Guillemain and Terroir (2006), who found a minimum in perceived loudness at the region of maximum pitch variation.

Measurements such as those shown in Fig. 4 cannot be simply used to calibrate measurements made with players rather than blowing machines. The main reason is that for some players, the tip of the finger occasionally overhangs the key and contributes a small component to the reflection measured by the sensor.

B. Recording and analyzing the sound

The sound produced by the flute was recorded using a Rhode NT3 microphone placed on a fixed stand and digitized using one input of the MOTU audio interfaces. The midpoint of the flute was approximately 50 cm away from the microphone. The frequency was calculated from the recorded sound file using PRAAT sound analysis software (Boersma, 1993), using autocorrelation with an analysis window of 15 ms. When studying a synthesized semitone discontinuity in a sinusoidal signal (between 1000 and 1059 Hz), the frequency transition was detected approximately 8 ms after the actual transition. The intensity and sound level were also extracted using PRAAT.

C. Characterization of open/closed states and note transitions

One of the aims of this study is to measure the relative timing of the open/closed transitions, so it is necessary to relate a defined value of sensor output to the effective opening/closing of the associated key. Most of the variation in sound frequency occurs close to the point of key closure, so the saturation point in Fig. 4 was considered as a possible choice. In practice, because of the variations described earlier, this value would be somewhat different for each flutist, key, and level of background illumination. Instead, guided by curves such as those shown in Fig. 4, the effective opening value for a key transition was set between 70% and 85% of the total variation in sensor output, the exact value depending on the key and situation (see Fig. 4).

Determining the duration of effective key opening and closing is also complicated by the saturation of sensor output described above. After examination of a range of traces under different conditions, it was chosen to measure the time taken between sensor signals of 30% and 70% of the maximum range of the sensor output. This rate of change was then multiplied by a factor of 100/40 to produce a measurement of the effective key closing or opening time. Examples are shown in Fig. 5.

An automated software routine was used to detect and to characterize the key movements in each recording. First, it detected each time the output of a key sensor passed through a value midway between neighboring fully closed and fully open states. These then served as initial starting points to find the nearest times when a key was effectively open or closed. These allowed calculation of the duration of each open/closed or closed/open transition. An uncertainty in each individual measurement is estimated by determining how long it took each key sensor output to vary by $\pm 5\%$ around the effective open or closing value (as defined in Sec. III C). This value was, on average, 1.4 ± 1.4 ms ($n=2069$) for a closing key and 4.1 ± 2.6 ms ($n=1639$) for an opening key.

A single key may be associated with several different note transitions, so the key movements, detected as described above, need to be associated with the note transitions of interest. In order to find note transitions automatically, the detected pitch was quantized to the set of notes used in the exercise. These quantized data were smoothed using a filter which calculates the median value within a window of 50 ms so that only sufficiently long values corresponding to note transitions are detected. For each note transition thus detected, the corresponding nearest transition reference times for key motion are found. The pitch and key position detection is shown for two particular exercises in Fig. 5.

The following parameters were calculated for each key transition associated with a given note transition: the effective duration of the key transition, the offset in time between the key transition and the pitch transition (see Fig. 5), and an estimate of the uncertainty in the key closing or opening time.

D. Subjects

The 11 players participating in this experiment were divided into three categories according to experience. Professionals (players P1–P7) are those with more than 8 years experience and for whom flute playing is a significant part of their current professional life, either as performers or as teachers. Amateurs (A1–A3) have between 3 and 8 years of flute playing experience and regularly play the flute as a non-professional activity. Beginners (B1) have less than 3 years of experience.

E. Experimental protocol

These experiments were conducted in a room with low reverberance that was significantly isolated from external noise. All measurements were performed on a specific flute from the laboratory, a C-foot Pearl flute fitted with a sensor near each key. The flute is a plateau or closed key model; i.e., the keys do not have holes that must be covered by the fingers, and it has a split E mechanism, which means that there is only one hole functioning as a register hole in the fingering for E6. Players could use their own head joint if desired.

A typical session took about 75 min. Each subject was asked to warm up freely for about 15 min in order to become accustomed to the experimental flute, the change in balance, and some awkwardness caused by the cables. They also used this time to rehearse the particular exercises in the experimental protocol. The musical exercises, written in standard musical notation, were delivered to the subjects approximately 1 week before the recording session. Some players did not complete all exercises.

F. Experimental exercises

The players were recorded performing a selection of note transitions, scales, arpeggios, and musical excerpts. Except for the musical excerpts, each written exercise was performed at least twice at two different tempi: players were asked to play "fast" (explained to them thus: as fast as possible while still feeling comfortable and sure that all the notes in the exercise were present) and "slow" (in a slow

TABLE I. Durations (average ± standard deviation and number of transitions used for statistics in brackets) of different key movements.

Key	Finger	Press time (ms) (n)	Release time (ms) (n)
3	L index	11.3 ± 4.8 (100)	15.9 ± 4.9 (100)
4	L thumb	8.7 ± 1.1 (204)	9.2 ± 2.1 (205)
6	L medium	15.2 ± 6.0 (101)	16.9 ± 3.7 (304)
7	L ring	17.0 ± 9.5 (273)	22.2 ± 9.4 (480)
11	R index	11.2 ± 5.1 (160)	15.6 ± 3.3 (185)
12	R medium	8.9 ± 11.1 (160)	15.3 ± 3.1 (181)
13	R ring	8.3 ± 2.5 (532)	16.7 ± 9.4 (524)
14	R little	12.1 ± 2.6 (179)	12.9 ± 9.3 (172)

tempo but still comfortable to perform the exercise once in a single breath). In the case of the fast performance, the musician was asked to repeat the exercise as many times as possible (but still comfortable) in a single breath.

IV. RESULTS AND DISCUSSION

The times taken for a key to be effectively depressed and released (i.e., to make the relevant key open or close depending upon its mechanism) are shown for some keys in Table I. In all cases, pressing times are quicker perhaps because the finger can transfer momentum gained in a motion that may begin before contact with the key, whereas a released key has to be accelerated from rest by a spring. The large variation among the durations for finger activated motion may include the variations in strength and speed of the fingers in the somewhat different positions in which they are used. There is rather less variation among the mechanical release times. However, some keys differ noticeably from the others. Large variation in the latter involves the key mechanism: Some keys move alone, others in groups of two or three, because of the clutches that couple their motion. In particular, the left thumb key and the right little finger (D♯ key) have stiffer springs, so their release movement is significantly faster ($p<0.001$) than for other keys. These keys have relatively important roles in supporting the flute. Overall, slow tempi produce significantly slower ($p=0.03$) key press times.

A. Single finger transitions

When only one key is involved in a note transition, the pitch change is a direct consequence of the motion of that key. As explained above, the transition from one note to another is not discrete. The frequency of the minimum in $Z(f)$ corresponding to the fundamental of one note is shifted gradually as the opening cross-section of the hole is increased. A relatively small range of key positions, near the fully closed limit, is associated with most of the changes in pitch (Fig. 4): Variations in position near the fully open state have much less effect. The delay between detected key motion and frequency change was 1.9 ± 3.8 ms ($n=1303$), which is less than the uncertainties in the measurements: The uncertainty in frequency change is several milliseconds, and the experimental uncertainty in key motion is a few milliseconds (Fig. 5). (The time for radiated sound from any part of the flute to reach the microphone was only about 1 or 2 ms.)

B. Two finger transitions

When two keys are involved in a note transition, there are two possible intermediate key configurations due to the non-simultaneous movement of the fingers. Examples involving the index and ring fingers of the right hand moving in opposite directions are the transitions F4 to F♯4, F5 to F♯5, and F6 to F♯6, but only F6/F♯6 involves an unsafe intermediate. Using X to indicate depressed and O to indicate unpressed and only indicating the index, middle, and ring fingers of the right hand, this transition goes from XOO to OOX, with the possible intermediates being XOX and OOO, as shown in Fig. 2. The fingering with both keys depressed (XOX) plays a note 14 cents flatter than F6. The fingering with neither key depressed (OOO) is more complicated. If the speed of the air jet is well adjusted to play F6 or F♯6, then this fingering does not play a clear note (see Fig. 6). If the speed of the air jet is somewhat slower than would normally be used to play these notes, then it will play a note near B5. So, apart from the ideal, unrealizable, "simultaneous" finger movement, there can only be one safe path for the transition XOO to OOX, and that goes via the fingering XOX (in which both keys are briefly depressed): The slight perturbation in pitch cannot be detected in a rapid transition.

The authors have also sought to compare intermediate states used during different exercises involving the same transition. Players were asked to alternate rapidly between the two notes as well as play them in the context of a scale. The exercise of rapidly alternating between XOO and OOX fingerings is an artificial exercise. For such rapid alternations, players often use special trill fingerings, in which intonation and/or timbre are sacrificed in return for ease of rapid performance. To perform this trill, a flutist would normally alternate the XOX and OOX fingerings, i.e., transform it into a single finger transition using the index finger only. Thus, the alternation exercise is one that flutists will not have rehearsed before this study. By contrast, the same key transition in the context of a scale (here the B♭ minor scale) is one which experienced flutists will have practiced over years.

Considerable differences were found between slow and fast trials (data not shown). The results for all players are presented in Fig. 7. Figure 7 shows that the descending transition (F♯6 to F6) has a relatively consistent behavior. For note alternations, professional musicians used a safe finger order 72% of the times (that is, transiting through the XOX state where both keys are depressed). Although they sometimes (48%) use the unsafe finger order in the scale context, $t_{13}-t_{11}$ was in average 3.8 ± 9.6 ms ($n=210$), so that this transition is close to simultaneous.

In the ascending case (F6 to F♯6) of the alternation exercise, professionals used safe transitions in 57% of cases although the behavior was less homogeneous among players ($p<0.001$ in the F6/F♯6; $p=0.02$ in F♯6/F6), but in the scale context all musicians used safe transitions 97% of the time [$\langle t_{13}-t_{11}\rangle=-26\pm17$ ms ($n=33$)].

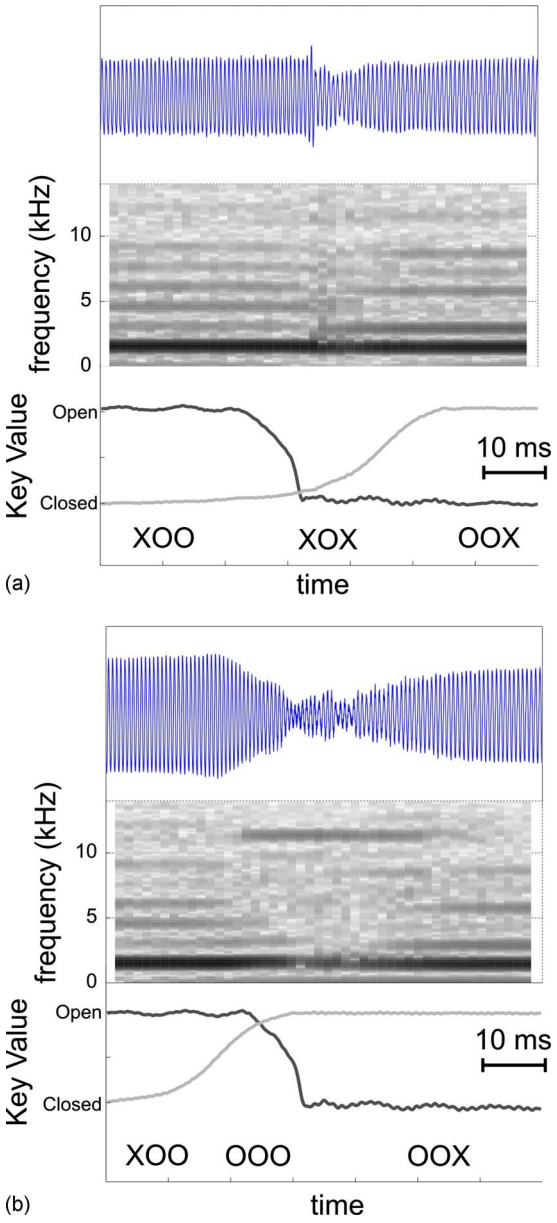


FIG. 6. (Color online) Oscillograms and spectrograms of the sound produced by the same player in two examples taken from the same trial for F6 to F♯6 transitions (nominally 1397–1480 Hz), with key sensor signals shown below. Spectra were taken with windows of 1024 samples (23 ms), separated by 256 samples (6 ms); gray levels are proportional to the logarithm of the amplitude in each frequency bin. The example shown on top with both keys closed during the transient is a safe transition (see text), and that on the bottom with both keys open during the transient is an unsafe one (dropout). The former shows a brief and relatively continuous transient. In the latter, the harmonic components of the sound are interrupted for tens of milliseconds. During the unsafe transient, higher frequency modes are excited in the absence of a suitable resonance.

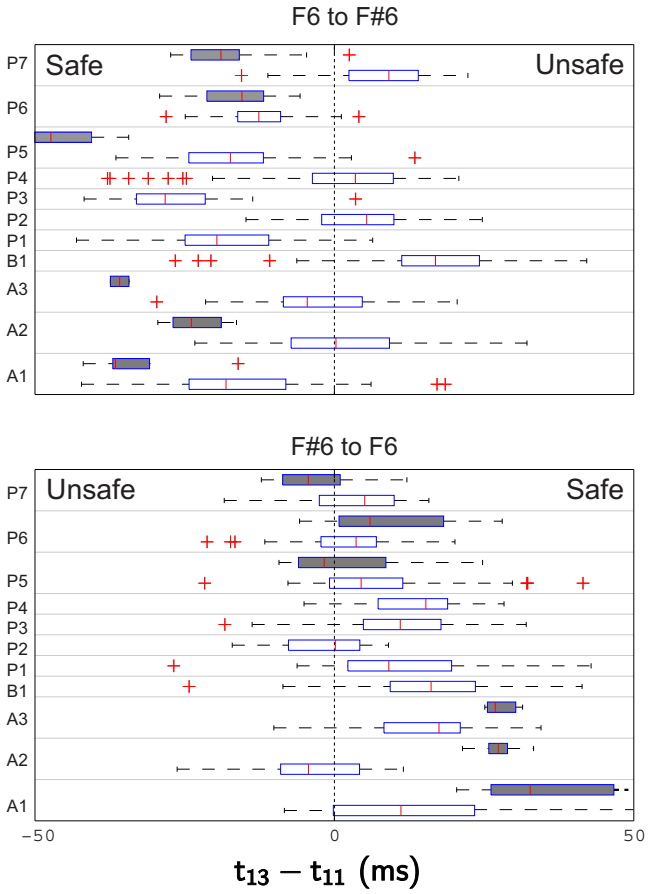


FIG. 7. (Color online) Box plot representation of time differences between the movement of keys 13 and 11 for transitions between F6 and F♯6 in alternations (hollow rectangles, typically $n \sim 20$) and scale exercises (filled rectangles, typically $n \sim 4$) for different players. Rectangles represent the range between the first and third quartiles with a central line representing the median. Error bars extend to the range of data points not considered as outliers. Outliers are represented as crosses.

Although the transition from F4 to F♯4 uses the same finger movements as F6 to F♯6, in this case all transition pathways are safe. Interestingly, most of the professionals tend to use the finger order, which would be safe for F6 to F♯6 with similar time differences between keys ($p=0.02$ for F/F♯ and $p=0.13$ for F♯/F), even though there is no unsafe intermediate for F4 to F♯4. For one professional (P5) and most amateurs, the time differences did change significantly but with no consistent direction (either becoming more or less simultaneous). The finger order remained the same as for the other subjects.

C. Three finger transitions

The example that will be used is the E5 to F♯5 transition, which involves three different keys (11, 12, and 13), moved respectively by the index, middle, and ring fingers of the right hand, with the little finger remaining depressed throughout. Using the notation described above and neglecting other keys, E5 is played using XXO (Th123|12-D♯) and F♯5 is played using OOX (Th123|--3D♯). Thus, the fingers are lifted from keys 11 (index) and 12 (middle), and key 13 (ring) is depressed.

Discounting the idealized simultaneous movement of the fingers, there are six possible pathways involving discrete transients. From an acoustical point of view, only one of these is safe: Key 11 moves first, then key 13, and then key 12 (i.e., XXO, OXO, OXX, OOX). This path is safe because OXO and OXX both play slightly flat versions of F♯5. Conversely, when descending from F♯5 to E5, the only safe transition is 12, 13, and then 11 (OOX, OXX, OXO, XXO). Any other path involves a fingering that produces a note near

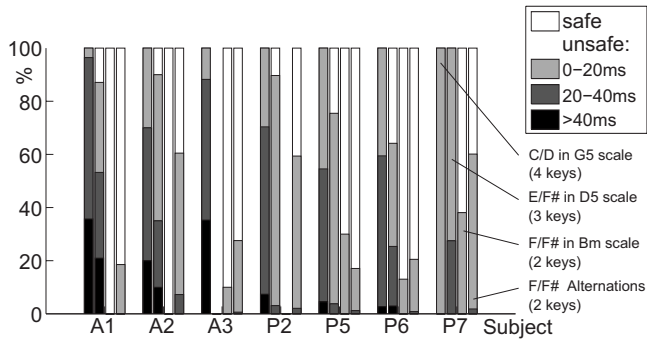


FIG. 8. The percentage of safe and unsafe note transitions. The different shaded sections of each bar show the data for three amateurs (left) and four professionals (right). For each player, the group of four vertical bars represents transitions involving contrary motion of fingers: C6–D6 in a G scale (four fingers), E5–F♯5 in a D scale (three fingers), and F6–F♯6 in a B minor scale and also in an alternation exercise (two fingers).

G5 (OOO), F5 (XOX or XOO), or D♯5 (XXX), which may or may not be noticeable depending on the duration of the intermediate. The paths are shown in Fig. 1.

The results for seven flutists for this transition are summarized in Fig. 8. For this transition, most professionals are nearly safe, passing through states that are unsafe for only 20 ms. The durations in the intermediate states vary substantially among players. Even though this context (a D major scale) is one that flutists would have practiced many times, the delay times are substantially longer here than for the two fingers, contrary motion example shown above, which uses two of the same fingers. From examining the average and standard deviations of the inter-key time difference, it was also observed that some of the subjects (P2 and P5) have significantly smaller time differences (data not shown). To summarize, professionals spend an average of 13 ± 9 ms ($n=225$) in unsafe transitions, whereas amateurs spend 25 ± 16 ms ($n=88$), which are significantly different.

D. Four finger transitions

The thumb and the three long fingers of the left hand move in the transition from C6 (OXOO or 1--|---D♯) to D6 (XOXX or Th-23|---D♯), with the index finger releasing a key and the others depressing keys. Here, there is no completely safe path of transient fingerings because the intermediate states involving a change in position of any two fingers all sound a note different from C6 and D6. There are partially safe paths: For example, starting with C6 and moving first either the middle or ring finger still produces a note very close to C6.

Average times spent in unsafe transitions, measured in the context of a G major scale, are shown in Fig. 8. As with the previous example, most flutists exhibited some preferred paths, but they were not consistent among players, and sometimes the same player may use different finger orders while playing quickly or slowly.

To summarize, all professional players spend less time in unsafe configurations than do the amateurs: Unsafe intermediate states last for an average of 34 ± 14 ms ($n=65$) for amateurs and 21 ± 11 ms ($n=133$) for professionals. The difference is significant ($p<0.001$). Thus, for both professionals and amateurs, the time spent in unsafe configurations is larger when four fingers rather than three are involved.

E. Summary of multi-finger transitions

Typical values of delays between fingers are about 10–20 ms for transitions that involve the motion of two or more fingers and significantly longer for amateurs than for professionals. In multi-finger transitions, the most frequent finger orders are often those that avoid or minimize the use of transient fingerings that are unsafe, as defined above. This is particularly true for transitions involving two fingers but less evident in more complicated transitions.

Finger motion delays for the four-, three-, and two-finger changes discussed above can be compared. (For one finger, the delay by definition is zero, as it does not include the time for key motion.) This is shown in Fig. 8, which summarizes the results obtained in the examples studied in this article.

These results can be related to similar studies on repetitive tapping movements in non-musical exercises (Aoki *et al.*, 2003). In these, single finger movements show a variability in inter-tap intervals of about 30 ms, increasing to 60 ms in the most agile fingers when two fingers are involved. When ring and little fingers are involved, this value is increased to 120 ms. These high values suggest that for non-musicians the inaccuracies in multi-tap intervals are mostly due to the duration of the finger motion rather than to synchronization between the motion of two fingers, but no references were found for measurements of these values. Trained musicians, of course, may yield different results.

Informal tests on the subjects of this experiment show that when no sound output is involved, the standard deviation in delays between keys can increase from approximately 20–40 ms, independently of the proficiency. These results suggest that for musicians the context is important in determining finger coordination.

Finally, delays between fingers are unimportant if their effect on sound cannot be detected. Gordon (1987) studied perceptual attack times in different pairs of instruments. These have variations that range from 6 to 25 ms, but in the flute they are about 10 ms. Minimum durations needed to identify pitch are typically four periods (less than 10 ms for flute notes) (Patterson *et al.*, 1983).

V. CONCLUSIONS

For single key transitions, the transition time is determined by the time for a finger to push a key in one direction, typically 10 ms, or for a spring to push it in the other, typically 16 ms (for the springs on this particular flute). When more than one finger is involved, the delay times between individual key movements must be added to this. For a transition involving only two fingers and thus only two pathways, players in general coordinate their fingers so that an unsafe transition is avoided. For some transitions, there is no safe path. Professionals, unsurprisingly, are more nearly simultaneous than amateurs. For both amateurs and professionals, total delay increases with the number of fingers in contrary motion.

ACKNOWLEDGMENTS

We thank our experimental subjects and The WoodWind Group. This research was supported by the Australian Research Council.

Aoki, T., Francis, P. R., and Kinoshita, H. (**2003**). "Differences in the abilities of individual fingers during the performance of fast, repetitive tapping movements," Exp. Brain Res. **152**, 270–280.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Proc. Inst. Phonetic Sci., **17**, 97–110.

Botros, A., Smith, J., and Wolfe, J. (**2002**). "The virtual Boehm flute—A web service that predicts multiphonics, microtones and alternative fingerings," Acoust. Aust. **30**, 61–66.

Botros, A., Smith, J., and Wolfe, J. (**2003**). "The 'virtual flute': Acoustics modelling at the service of players and composers," in Proceedings of the Stockholm Music Acoustics Conference (SMAC 03), Stockholm, Sweeden, pp. 247–250.

Botros, A., Smith, J., and Wolfe, J. (**2006**). "The virtual flute: An advanced fingering guide generated via machine intelligence," J. New Music Res. **35**, 183–196.

Coltman, J. W. (**1976**). "Jet drive mechanisms in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733.

de la Cuadra, P., Fabre, B., Montgermont, N., and Ryck, L. D. (**2005**). "Analysis of flute control parameters: A comparison between a novice and an experienced flautist," in Proceedings of the Forum Acusticum 2005, Budapest, Budapest, Hungary.

Gordon, J. W. (**1987**). "The perceptual attack time of musical tones," J. Acoust. Soc. Am. **82**, 88–105.

Guillemain, P. and Terroir, J., *Dynamic Simulation of Note Transitions in Reed Instruments: Application to the Clarinet and the Saxophone* (Springer-Verlag, Berlin, **2006**), pp. 1–23.

Miranda, E. R., and Wanderley, M. (**2006**). *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, 1st ed. (A-R Editions, Middleton, WI).

Palacio-Quintin, C. (**2003**). "The hyper-flute," in Proceedings 2003 Conference on New Interfaces for Musical Expression, National University of Singapore, pp. 206–207.

Palmer, C., Carter, C., Koopmans, E., and Loehr, J. D. (**2007**). "Movement, planning, and music: Motion coordinates of skilled performance," in Proceedngs of the Inaugural Conference on Music Communication Science, Sydney, Australia.

Patterson, R. D., Peters, R. W., and Milroy, R. (**1983**). "Threshold duration for melodic pitch," in *Hearing: Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, Berlin), pp. 321–325.

Wolfe, J., and Smith, J. (**2003**). "Cutoff frequencies and cross fingerings in baroque, classical, and modern flutes," J. Acoust. Soc. Am. **114**, 2263–2272.

Wolfe, J., Smith, J., Tann, J., and Fletcher, N. H. (**2001**). "Acoustic impedance of classical and modern flutes," J. Sound Vib. **243**, 127–144.

Ystad, S., and Voinier, T. (**2001**). "A virtually real flute," Comput. Music J. **25**, 13–24.

Modeling source-filter interaction in belting and high-pitched operatic male singing

Ingo R. Titze
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204 and Department of Communication Sciences and Disorders, The University of Iowa, Iowa City, Iowa 52242

Albert S. Worley
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

(Received 7 June 2008; revised 14 May 2009; accepted 9 June 2009)

Nonlinear source-filter theory is applied to explain some acoustic differences between two contrasting male singing productions at high pitches: operatic style versus jazz belt or theater belt. Several stylized vocal tract shapes (caricatures) are discussed that form the bases of these styles. It is hypothesized that operatic singing uses vowels that are modified toward an inverted megaphone mouth shape for transitioning into the high-pitch range. This allows all the harmonics except the fundamental to be "lifted" over the first formant. Belting, on the other hand, uses vowels that are consistently modified toward the megaphone (trumpet-like) mouth shape. Both the fundamental and the second harmonic are then kept below the first formant. The vocal tract shapes provide collective reinforcement to multiple harmonics in the form of inertive supraglottal reactance and compliant subglottal reactance. Examples of lip openings from four well-known artists are used to infer vocal tract area functions and the corresponding reactances.
 [DOI: 10.1121/1.3160296]

PACS number(s): 43.75.Rs [NHF] Pages: 1530–1540

I. INTRODUCTION

Many of the pedagogical approaches to teaching singing styles are based on the concept that there are preferred vowel configurations for a given pitch (Appelman, 1967; Vennard, 1967; Miller, 1986, 2008). Speaking vowels are modified and adjusted not only to create a variety of timbres, but also to support the sound source in self-sustained oscillation by providing favorable acoustic reactance (Titze, 1988a; Fletcher, 1993). Because no source-vocal tract interaction is claimed in linear coupling, linear source-filter theory as traditionally applied to speech cannot account for source strengthening by vocal tract coupling, nor can it account for source instabilities and bifurcations in vocal fold oscillation related to vowel selection. In two recent investigations (Titze *et al.*, 2008; Titze, 2008a) it has been shown that source energy in phonation (vocal fold vibration and the associated glottal airflow) can be significantly increased by vocal tract interaction. However, when any harmonic that carries a significant portion of the source energy passes through a formant (a vocal tract resonance), vocal fold vibration can also be destabilized. Pitch jumps, subharmonics, chaotic vocal fold vibration, and other bifurcations can occur that are (in part) attributable to acoustic loading by the vocal tract. Hence, it appears that a singer of harmonically-based singing styles may seek to obtain both stability and uniform reinforcement of the harmonics by carefully selecting a favorable vocal tract configuration.

An insightful exposition to contrasting styles was given by Schutte and Miller (1993). Focusing on the female voice in middle to high-pitch ranges, the authors observed that belters use vocal tract resonances (formants) differently from classically-trained (opera and art song) singers. In particular, the second harmonic was found to receive strong reinforcement by the first formant in belting, much more so than in the classically-trained style. Schutte and Miller (1993) went so far as to say that the entire characteristic of a belt is based on a strong second harmonic, combined with a high degree of glottal closure during vocal fold vibration. In a later investigation, Miller and Schutte (2005) demonstrated that successful bridging of registers in singing (perceptual discontinuities in the timbre of the sung tone) "may be more a consequence of skillful use of resonance than of muscular adjustments in the glottal voice source." In the same year, Schutte *et al.* (2005) showed that some famous operatic tenors reinforce the third harmonic on a high B^{b}_{4} in a well-known operatic aria, "Celeste Aida." They suggested that this is accomplished by elevating the second formant (F_2).

Extending the investigations to a greater variety of male voices (basses, baritones, and tenors), Neumann *et al.* (2005) showed that in the male modal register (Hollien, 1974, 1983; Titze, 2000), the second and fourth harmonics dominate, one being resonated by the first formant and the other by the second formant. As the male singers transit through the *primo passagio* (a passage around F_4 where a change occurs from modal register to a mixture of modal and falsetto registers), the third harmonic gains strength from the second formant, while the second harmonic loses energy. Neumann *et al.* (2005) also stated that supraglottal resonances play a greater role in register discrimination than subglottal reso-

nances, reversing a former hypothesis by one of the current authors (Titze, 1988b).

The overall source spectrum distribution was studied by Stone *et al.* (2003). Studying a female that could sing several styles, they showed that the Broadway style (which often incorporates belt) has the greatest proportion of high-frequency energy, followed by the operatic style, and then by the normal speech of the subject. The subglottal pressure was higher in the Broadway style than in the operatic style, and the open quotient in the glottis was smaller. Overall, the formant frequencies were higher in Broadway style than operatic style.

What is not yet exposed in the above investigations is the interaction between source characteristics and vocal tract resonance. Traditional analysis has been guided by the long-standing linear source-filter theory (Fant, 1960), which assumes that the source and the filter operate independently, even though an explicit "correction" is given to the glottal waveform that carries vocal tract loading effects in the form of pulse skewing and formant ripple (Flanagan, 1968; Rothenberg, 1981; Fant, 1986, Fant and Lin, 1987). With such a flow source correction, source and filter can be combined or recombined (as in analysis—synthesis) spectrally to produce the mouth output, but interactions that increase the amplitude of vocal fold vibration or destabilize the source (i.e., major bifurcations in tissue movement) cannot be treated easily with such corrections.

Registers have generally been described in the domain of the sound source (for an up-to-date review, see Henrich, 2006), while voice quality and singing style have more often been described in the domain of vocal tract resonance (Estill, 1988; Yanagisawa *et al.*, 1991; Miller, 2008; Story *et al.*, 2001; Bergan *et al.*, 2004). That variations in the source and the filter co-exist in the singing styles have clearly been recognized, but how they feed off each other (constructively and destructively) has only been described recently (for a popular review, see Titze, 2008b).

The current nonlinear source-filter theory for singing is based on the assumption that stored energy in the vocal tract can assist in vocal fold vibration through feedback. The stored energy is quantified in terms of acoustic reactance of the air column above or below the vocal folds. Thus, for certain singing styles, there can be a much closer analogy to wind instrument acoustics (Fletcher, 1993; Fletcher and Rossing, 1998) than has traditionally been claimed for speech. In fact, the analogy between lip vibration in brass acoustics and vocal fold vibration in vocal tract acoustics for singing is remarkable (Adachi and Sato, 1996; Ayers, 1998). Yet there is a major difference. In singing, multiple resonances of the vocal tract are not generally "tuned" to the harmonics of the source. Two factors prevent this: (1) the shortness of the tube (15–20 cm for a supraglottal vocal tract and 12–16 cm for a subglottal tract) and (2) the desire to communicate a verbal message with vowels and consonants along with the musical message.

Instead of formant-harmonic "tuning," it is hypothesized that the singer learns to utilize supraglottal inertive reactance (and occasionally subglottal compliant reactance) to reinforce vocal fold vibration by choosing pitch-vowel combinations that keep several harmonics in favorable reactance regions simultaneously (Titze, 2008a), but not necessarily tuned to the formants. While ascending or descending in pitch, it appears that singers who want to maintain a stable harmonic spectrum learn to "lift" their harmonics over unfavorable reactance regions by adjusting formant frequencies. A weak voice and unstable nonlinear effects can thereby be avoided, such as excessive subharmonics or irregular vocal fold vibration. Although formant tuning to harmonics has been claimed in earlier work by Sundberg (1977) and later by Schutte and Miller (1993) and Neumann *et al.* (2005), their published data suggest that while harmonics are often near the center of the formant, they are not generally in the center. When formants are measured with an independent sound source during singing (Joliveau *et al.*, 2004), the case for exact formant-harmonic tuning is also weak, even though harmonics and formants can move up and down together in close proximity.

The purpose of this paper is to contrast two vocal tract shapes in terms of source-filter interaction. These shapes resemble vowels modified for singing in the two contrasting styles mentioned. Specifically, the following questions are of interest: (1) Does the inverted megaphone mouth shape often used by singers of Western opera and art songs reinforce harmonics above the fundamental in favorable reactance regions *above* the first formant? (2) In contrast, does the megaphone mouth shape often used by belters reinforce both the fundamental and the second harmonic with favorable inertive supraglottal reactance *below* the first formant? Given the large number of possible pitch-vowel interactions and differences in male-female anatomy, the authors limit theirselves to a few male high-pitch productions. The female voice with the same stylistic differences will be discussed in a follow-up paper. In order to make this paper useful to a broad audience that includes singing pedagogues, some tutorial material on nonlinear source-filter interaction is included that leads to the case presentations.

II. PITCH-VOWEL INTERACTION

The degree of interaction between the source of sound (vocal fold vibration with its accompanying glottal flow) and the vocal tract filter depends on the relation between the source impedance and the vocal tract input impedance. As in electric circuit theory (Skilling, 1966), the source impedance is large compared to the vocal tract impedance, little interaction will occur. If the impedances are comparable, much interaction will occur. The underlying hypothesis is that reactive impedance, above and below the glottis, can store energy and feed it back to the source with delayed or advanced phase, thereby interfering (either constructively or destructively) with vocal fold vibration.

A. Interaction with wave-reflection algorithms

It has been customary to simulate vocal tract acoustics with a wave equation that is modified to include wall vibration, kinetic loss, viscous loss, and lip radiation (Lilljencrants, 1985; Story, 1995). The vocal tract is subdivided into many cylindrical sections, typically about 36 for the subglot-

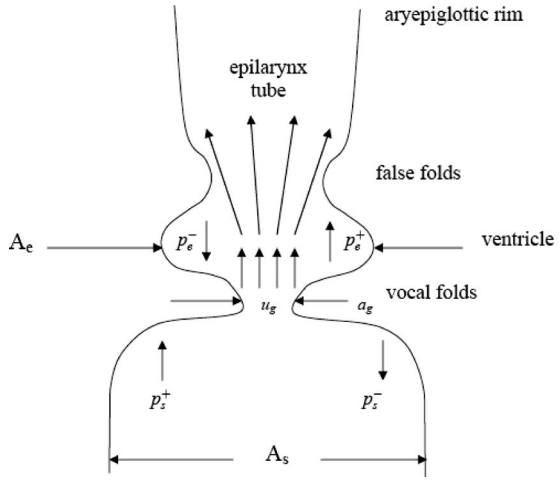


FIG. 1. Diagram of vocal folds and lower vocal tract to illustrate source-filter interaction.

tal (tracheal) system and 44 for the supraglottal system for males. Reflection coefficients are computed at the boundaries of abutting sections and the loss terms are added as "corrections" to the basic scattering equations at the boundaries (Lilljencrants, 1985).

To account for source-tract interaction (Fig. 1), an analytical closed-form solution for glottal flow can be used (Titze, 1984), which has the following form:

$$u_g = \frac{a_g c}{k_t}\left\{-\left(\frac{a_g}{A^*}\right) \pm \left[\left(\frac{a_g}{A^*}\right)^2 + \frac{4k_t}{\rho c^2}(r_3 p_s^+ - r_e p_e^-)\right]^{1/2}\right\}. \quad (1)$$

Here u_g is the interactive glottal flow, a_g is the time-varying glottal area (computed where flow detachment occurs in the glottis and a jet is formed; see Fig. 1), c is the sound velocity, and k_t is a transglottal pressure coefficient for modified Bernoulli flow through this glottis. Further, A^* is an effective vocal tract area defined as

$$A^* = \frac{A_s A_e}{A_s + A_e}, \quad (2)$$

where A_s is the subglottal area and A_e is the epilaryngeal (supraglottal) area, which begins at the laryngeal ventricle (Fig. 1). Two reflection coefficients in Eq. (1) are defined as follows:

$$r_s = \frac{A_s - a_g}{A_s + a_g}, \quad (3)$$

$$r_e = \frac{A_e - a_g}{A_e + a_g}. \quad (4)$$

Finally, p_s^+ is the forward traveling acoustic wave pressure from the subglottis while p_e^- is the backward traveling wave pressure from the supraglottis. (When added together, forward and backward traveling waves form the total acoustic pressure in any section of the vocal tract.)

To complete the analytical calculation, the departing partial pressure waves from the subglottis and supraglottis, respectively, are

$$p_s^- = \frac{\rho c}{A_s}(-u_g) + r_s p_s^+, \quad (5)$$

$$p_e^+ = \frac{\rho c}{A_e}(+u_g) + r_e p_e^-, \quad (6)$$

and the total subglottal and supraglottal acoustic pressures are

$$p_s = p_s^+ + p_s^-, \quad (7)$$

$$p_e = p_e^+ + p_e^-. \quad (8)$$

The only advancement in the above formulation over what was published previously (Titze, 1984; Titze *et al.*, 2008) is the addition of the reflection coefficients r_s and r_e, which were previously set to 1.0 but are now time-varying because of the time-varying glottal area a_g. This refinement in the equations produces wave transmission losses through the glottis in both directions.

B. Special vocal tract shapes and their impedances

The acoustic input impedance of the vocal tract is frequency-dependent due to standing waves in the vocal tract (Fant, 1960; Flanagan, 1972), but the *characteristic impedance* is not frequency-dependent. Its value is $\rho c/A_e$, where ρ is the density of air, c is the sound velocity, and A_e is the entry area into the supraglottal vocal tract, known as the epilarynx tube area. [The epilarynx tube, which includes the laryngeal ventricle, the space between the ventricular (false folds), and the laryngeal vestibule, makes up the first 2–3 cm of the vocal tract above the vocal folds, terminated by the aryepiglottic rim, where the aryepiglottic folds are located.] If the vocal tract were infinitely long and of constant cross section A_e, no reflections would take place to create standing waves, and the input impedance would be the constant value $\rho c/A_e$, an acoustic resistance. An average value of A_e for speakers (Story, 2005) is about 0.5 cm^2, which makes the characteristic impedance about 7.0 kPa per l/s. This is in the middle of the range of glottal impedances gleaned from pressure-flow data in the literature (Holmberg *et al.*, 1988; Dromey *et al.*, 1992; Sundberg, 1995; Alipour *et al.*, 1997; Stathopoulos and Sapienza, 1993, 1997; Sundberg *et al.*, 2004). Figure 2 shows a bar graph of glottal resistances for pressed voice, male modal voice, female modal (mixed register) voice, and falsetto voice. The glottal resistances are shown with clear bars. Also shown are three solid bars for characteristic vocal tract impedances for A_e=0.3, 0.5, and 1.0 cm^2. Note that there are many options for impedance matching and mismatching. For example, a 1.0 cm^2 epilarynx tube matches well with falsetto voice, a 0.5 cm^2 epilarynx tube matches well with male modal voice (and to a slightly lesser degree with mixed or female modal voice), and a 0.3 cm^2 epilarynx tube matches well with pressed voice.

But the complete vocal tract is nonuniform in cross section and finite in length, which means that the characteristic tube impedance becomes only a scale factor for the frequency-dependent impedance (Fant, 1960; Flanagan,

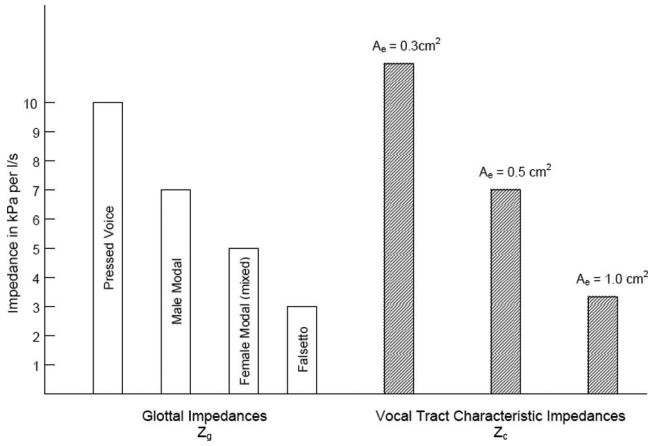


FIG. 2. Glottal impedances for different phonatory control (clear bars), and characteristic vocal tract input impedance Z_c (cross hatched bars) for different cross sectional areas A_e of the epilarynx tube.

1972; Stevens, 1999). Impedance values are complex, having both a real and an imaginary component, and can range from much smaller to much larger than the characteristic impedance at different frequencies.

Figure 3 shows the vocal tract input impedance for a collection of artificial vocal tract shapes that the authors consider the beginning caricatures (not measured on humans) of some of the singing styles discussed later in this paper. The vocal tract shapes are shown in the left panel and the corresponding supraglottal impedances are shown on the right panel. Because acoustic impedance is a complex quantity, as mentioned, the right panel shows the resistive component (real part of the impedance) in thin lines and the reactance (imaginary part of the impedance) in thick lines. Characteristic impedances Z_c are shown with short horizontal lines on the vertical axis (above or below the 10 kPa per l/s tic mark). The complex impedances were computed with transmission line theory (cascade matrices for variable cross sections; Story *et al.*, 2000) and include the radiation impedance, as well as viscous losses and wall vibration losses in all sections of the vocal tract (Sondhi and Schroeter, 1987; Story *et al.*, 2000). For all cases except the uniform tube, the epilarynx tube was chosen to be 0.5 cm^2. Note that impedance maxima can be as high as 50 kPa per l/s for the narrow megaphone shape (at the bottom). For the inverted megaphone shape (middle), the impedance maxima are less than 20 kPa per l/s, and for the uniform tube (top) they reach only about 10 kPa per l/s. The characteristic tube impedance Z_c is 7 kPa per l/s for all shapes except the uniform tube, which does not have the narrowed epilarynx tube. Z_c for the uniform tube is only 1.3 kPa per l/s.

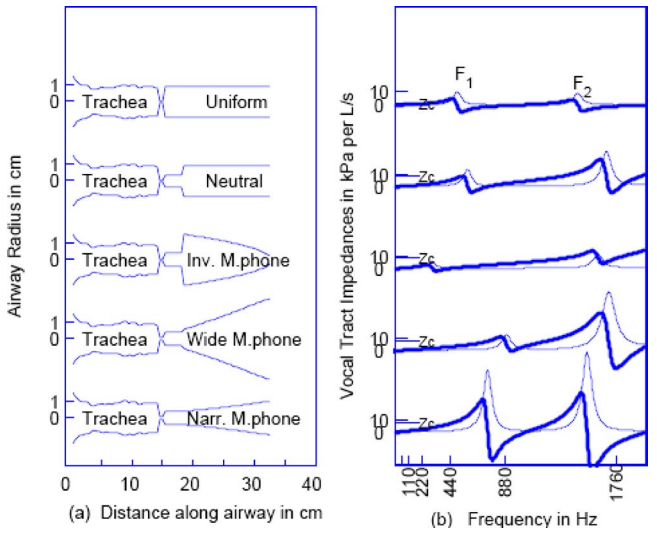


FIG. 3. (Color online) (a) Vocal tract caricatures and (b) corresponding input impedances as a function of frequency; thick lines are reactances and thin lines are resistances.

The authors reason that, in terms of an average impedance level, the inverted megaphone and neutral shapes may match well with moderate glottal adduction, whereas a narrow megaphone shape may match well with a pressed glottal adduction, as in a shout or a belt. The uniform tube, which has an extremely low input impedance, is an unlikely configuration for a human vocal tract because it is difficult for anyone to widen their epilarynx tube to the same diameter as the pharynx. It would produce an impedance mismatch with anything but a very wide glottis, as perhaps in very breathy voice. Nevertheless, the uniform tube is shown as a reference configuration because it is so widely discussed in speech science. In fact, it becomes the asymptote for linear source-filter coupling, for which the vocal tract input impedance must by definition be much lower than the glottal impedance (Titze, 2008a).

Whereas vocal tract resistance is always positive, reactance can be both positive and negative, as Fig. 3 shows. Two formant (resonant) frequencies, F_1 and F_2, are identified on the top impedance curve. Frequency ranges linearly from 0 to 2000 Hz and standard frequencies for musical pitches A_2-A_6 are labeled at the bottom. The musical pitches are spaced logarithmically on the linear frequency scale. Formant frequencies are located where the resistance has a local peak. For the 17.5 cm long uniform tract, these formants are located at 500 and 1500 Hz. Positive (inertive) and negative (compliant) supraglottal reactances alternate to the left and right of the formants, respectively. Positive supraglottal reactance has been shown to assist in self-sustained vocal fold oscillation, whereas negative supraglottal reactance hinders self-sustained oscillation (Titze, 1988a; Fletcher, 1993; Titze, 2008a). In the subglottal system, the effect is reversed. Negative (compliant) subglottal reactance helps vocal fold vibration whereas positive (inertive) subglottal reactance hinders vocal fold vibration. A computer simulation of this interaction will now be given.

C. Effect of vocal tract shape on vocal fold vibration

Given the many possible impedance curves with different vocal tract shapes, only a few special shapes can be chosen here to demonstrate source-tract interaction. Figure 4 shows simulations of glottal airflow with a well-described body-cover model of the vocal folds that self-sustains oscillation when a vocal tract is attached (Story and Titze, 1995; Titze and Story, 2002). The interaction is calculated with Eq. (1), in combination with a wave-reflection simulation of vocal tract acoustic pressures (Lilljencrants, 1985; Story, 1995).

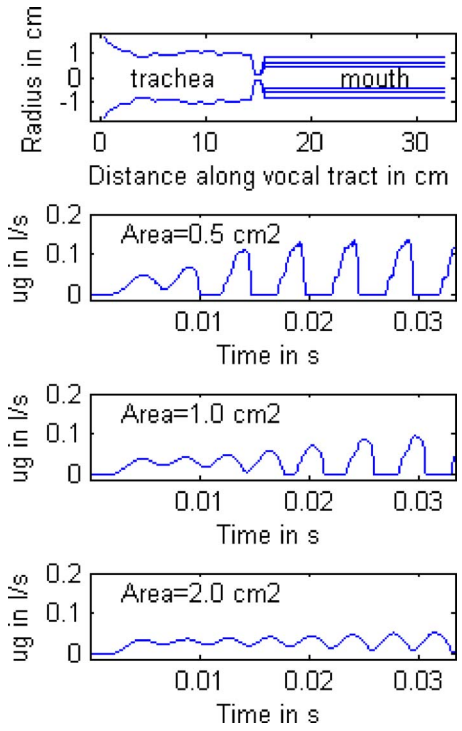


FIG. 4. (Color online) Computer simulation of glottal airflow with a self-sustained oscillation vocal fold model that interacts with three uniform tubes as shown in the top graph.

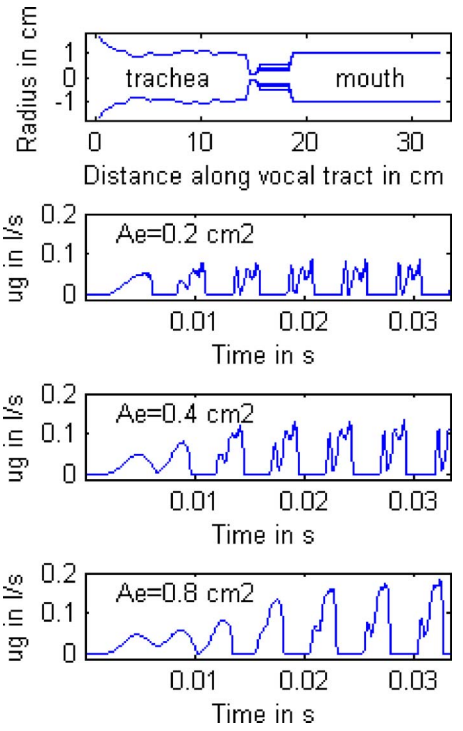


FIG. 5. (Color online) Computer simulations of glottal airflow (bottom three panels) with a self-sustained oscillation vocal fold model that interacts with a neutral tube and three epilarynx areas A_e.

The top of Fig. 4 shows three uniform supraglottal tubes with different cross sectional areas (0.5, 1.0, and 2.0 cm^2). All else in the model was kept identical; hence, the details of all other parameters will not be repeated here. The fundamental frequency was about 200 Hz, but varied slightly with vocal tract load. Note that the widest tube (2.0 cm^2) resulted in oscillation barely above threshold (bottom curve). As the tube narrowed, the onset of vibration was quicker, pulse height was greater, and the flow declination prior to closure was more abrupt.

Figure 5 shows similar curves, but in this case the cross sectional area of the epilarynx tube was varied. There was more formant ripple on the glottal flow waveform. The flow amplitude decreased with a narrower epilarynx tube, but the maximum flow declination prior to glottal closure still increased. Oscillation onset was again fastest with the narrowest tube.

These two examples point out that vocal tract configuration can have a profound effect on the source. Singers may widen or narrow their vocal tracts for different styles, even in the presence of specific vowels. The authors suspect that they also learn to control the cross sectional area of the epilarynx tube, although the musculature used for this control is not clearly understood. Favorable or unfavorable source-filter interaction is likely to dictate which vocal tract shape works with which singing style.

D. An inertogram for frequency-dependent interaction

To view the F_0-vowel interaction over large frequency ranges, it is useful to plot supraglottal vocal tract *inertance* (inertive reactance divided by the angular frequency $\omega = 2\pi F$). Inertance is a more evenly scaled quantity over a wide frequency range. Also, a logarithmic frequency scale is more suitable for matching frequency to keyboard pitches. Figure 6 shows a set of *inertograms* (supraglottal inertance versus frequency) for the six simple configurations chosen in the simulations of Figs. 4 and 5. The vocal tract shapes (three uniform tubes and three neutral tubes with different epilarynx diameters) are shown on the left, and inertance is shown in solid horizontal bars on the right. Negative supraglottal inertance (which would be compliance) is set to zero (producing only a baseline). Thus, whenever the supraglottal re-

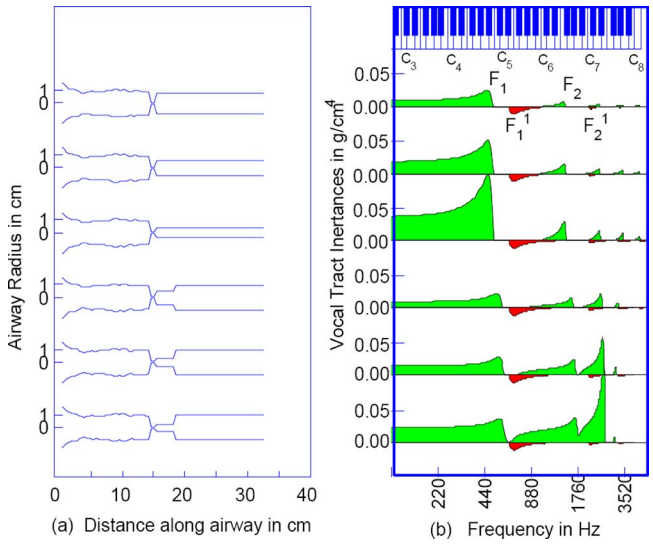


FIG. 6. (Color online) Six tube shapes (left) and their corresponding inertograms (right).

actance as calculated in Fig. 3 goes negative, the inertance merges into a single line, its value being set to zero to simplify the graphs.

Below the baselines one observes some small "teardrops" that represent subglottal compliance, which for the constant tracheal configurations shown in Fig. 6 exists in the 600–800 Hz region, and to a much lesser extent in the 2000–2500 Hz region. This subglottal compliance may also be useful for the singer to reinforce a harmonic, but further discussion is beyond the scope of this paper.

The mks units of inertance are kg/m^4, but the authors prefer to plot inertance in g/cm^4, which agrees more with the dimensions of the system ($1.0\ g/cm^4 = 10^5\ kg/m^4$). Inertance can be thought of as *density of an air column per unit length*. Oscillation threshold pressures (Titze, 1988a; Chan and Titze, 2006; Jiang and Tao, 2007) and glottal flow pulse skewing (Rothenberg, 1981; Titze, 2006) have previously been quantified in terms of vocal tract inertance. The vertical tic marks in Fig. 6 (right panel) indicate that the low-frequency inertances range between 0.01 and 0.04 g/cm^4 for the collection of tubes. Conceptually, this means that the vocal tract air columns, with an air density of about $0.001\ g/cm^3$, have effective acoustic lengths of 10–40 cm, even though the actual vocal tract length is a constant value of 17.5 cm. The narrowed epilarynx tube increases the acoustic length. At selected frequencies, peak inertance can reach above $0.1\ g/cm^4$, as shown in the inertograms.

The supraglottal formants (resonances of the vocal tract) are identified as the locations where the inertance bars suddenly collapse to the baseline. Similarly, subglottal formants are at the beginning of the sudden downward trend of the tear-drops (see labels on top of inertogram). The first subglottal formant (F_1^1) occurs at about 600 Hz (near E_5) and the second subglottal formant (F_2^1) occurs at about 1900 Hz (barely visible). Five supraglottal formants can be identified for the uniform tubes and four for the neutral tubes with a narrowed epilarynx tube. Note that changing the diameter of the uniform tube does not change the locations of the formants, but narrowing the epilarynx tube does. F_1 and F_2 are raised slightly, F_3 stays about the same, and F_4 is lowered slightly with epilarynx narrowing. The slight clustering together of F_3 and F_4 is known as *singer's formant clustering* (Sundberg, 1974; Titze and Story, 1997). The clustering may also include F_5, which is not seen here on the lower three inertograms.

The most important contrast between changing the entire tube diameter and changing only the epilarynx diameter is reflected at frequencies between 2000 and 3000 Hz. Note that inertance decreases monotonically with higher formants for the uniform shapes, but increases dramatically in the F_2 to F_3 region for the neutral tubes with a narrow epilarynx tube. This means that harmonics in the 2000–3000 Hz range can be strengthened with a narrowed epilarynx tube. The increased formant ripple in the previously discussed simulations of Fig. 5 (bottom to second from top) is evidence of this effect.

One way to quantify the acoustic benefit of source-tract interaction is to compute the maximum flow declination rate (MFDR) and compare it to the value it would have if the flow were sinusoidal. MFDR is known to correlate well with vocal intensity (Holmberg *et al.*, 1988; Gauffin and Sundberg, 1989). The authors define the normalized MFDR as

$$\mathrm{MFDR}_n = \frac{\dot{U}_m}{\omega U_m}, \tag{9}$$

where $\dot{U}_m$ is the MFDR (the maximum negative derivative of the flow), ω is the angular frequency ($2\pi F_0$), and U_m is the peak glottal flow. For a sinusoid, this ratio is 1.0. Figure 7 shows a diagram of this ratio computed from the waveforms of Figs. 4 and 5. It is seen that epilarynx narrowing is generally more effective than overall tube narrowing in increasing MFDR_n. The lowest value of MFDR_n is obtained for the $3.0\ cm^2$ uniform tube. Recall that the waveform for this case is nearly sinusoidal (Fig. 4, bottom), so a value near 1.0 for MFDR_n is expected. A value of 20 is obtained for the tube with the narrowest epilarynx ($0.2\ cm^2$). The corresponding waveform was the least sinusoidal (Fig. 5, second from top). This confirms the earlier claim that nonlinear source-filter coupling increases the *source strength*, measured by MFDR, not simply the energy transfer through the vocal tract at selected frequencies.

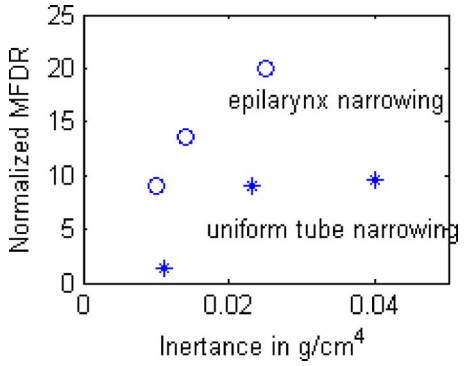


FIG. 7. (Color online) Normalized maximum flow declination rate (MFDR_n) for the six simulations of Figs. 4 and 5.

Note that for all the shapes shown in the inertogram of Fig. 6, there is a "dead" spot just above F_1. This occurs around 500–600 Hz. Singers experience difficulties when either F_0 or $2F_0$ is in this region. The authors will now show how singers may manage the avoidance of this dead spot.

III. VOCAL TRACT SHAPES DERIVED FROM MALE SINGERS

Magnetic resonance images (MRIs) of vocal tract shapes of a lyric baritone were obtained from Dr. Brad Story at the University of Arizona. The procedure followed work reported earlier (Story *et al.*, 1996). The singer produced several vowels and consonants in both a speaking mode and a singing mode. Figure 8 (top two rows on the left) shows the measured vocal tract area functions for the spoken /ɑ/ vowel and the sung /ɑ/ vowel for this baritone. The corresponding inertograms are to the right. Vertical lines are harmonics of the A_4 pitch, to be discussed later. The two shapes in the lower half of Fig. 8 will also be explained later.

The lyric baritone was neither a belter nor an operatic singer. His professional singing styles were lieder, early mu-

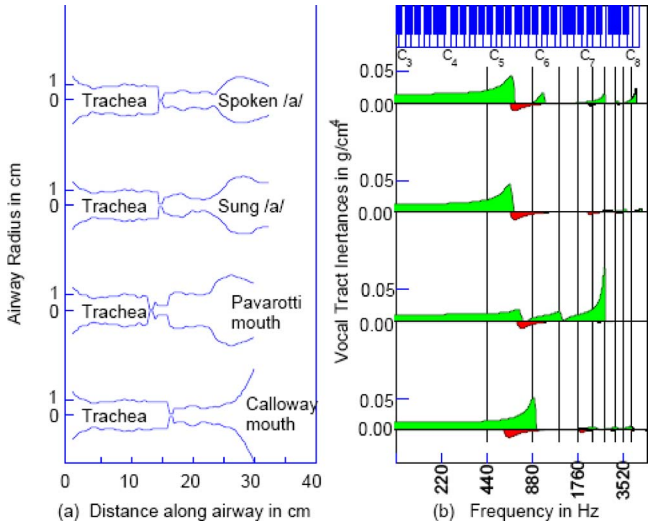


FIG. 8. (Color online) (Left) Vocal tract shapes derived from MRI data of a lyric baritone singer with various shape modifications, and (b) corresponding inertograms.

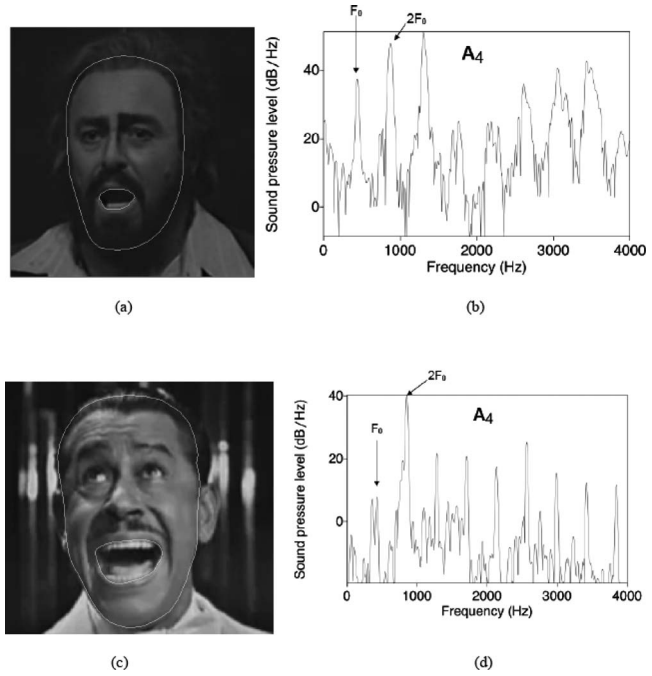


FIG. 9. (a) Mouth area and head area for Luciano Pavarotti singing A_4 on an /ɔ/ vowel. (b) Corresponding frequency spectrum. (c) Mouth area and head area for Cab Calloway singing A_4 on an /a/ vowel, and (d) corresponding frequency spectrum.

sic, and chanting. The main differences between his speaking and singing vocal tract were a wider mouth opening and a wider throat opening for the sung /ɑ/. He did not produce much vocal ring, which is evidenced by the fact the he maintained a relatively wide epilarynx tube (0.8–1.0 cm^2), even more so in singing than in speaking. So, one certainly needs to question whether his vocal tract shape is representative of an operatic singer.

A side-note about access to human subjects for singing research is in order. To the authors' knowledge, premier opera singers and musical theater belters have not made themselves available for detailed three-dimensional (3D) MRI studies, which requires several hours of phonation in a supine position. The best singers have too busy a schedule in their prime years, and their agents prefer not to see them engaged in such intensive research activities. While amateur or low-rank professionals are available, their techniques are sometimes less convincing. Hence, the authors opted to combine some data from their semi-professional baritone with mouth shapes from star-quality professionals. Mouth shapes are relatively easy to obtain from artists on public access video and audio recordings. While two-dimensional (2D) imaging could have been used with professional singers, the vocal tract area functions derived from 2D images require assumptions about cross-dimensions that are no easier to justify than "morphing" mouth shapes to known 3D images.

Several video recordings were chosen to provide examples for analysis. For male operatic singing, the tenor Luciano Pavarotti singing the aria "Vesti La Giubba" from Leoncavallo's opera *I Pagliacci* was analyzed. The video, a 1994 performance at the Metropolitan Opera, is freely available on YouTube as a high quality MP4 recording (Pavarotti, 1994). Because it was a live performance, the audio and video are synchronized; there is only one signal source. There is a negligible amount of background noise in the recording and, although the performance is accompanied by full orchestra, there is a brief unaccompanied segment of Pavarotti singing an A_4 at 0:43 s into the recording. As a second example of operatic production, Roberto Alagna singing and A_4 in the aria "E lucevan le stele" from Puccini's opera *Tosca* was analyzed. The video, freely available on YouTube (Alagna, 2000), is from a filmed performance in the year 2000.

For one example of male belt production, simultaneous audio and video recordings of the jazz singer Cab Calloway singing "St. James Infirmary" were analyzed. This video, recorded live in 1964 on the Ed Sullivan Show, is also freely available on YouTube (Calloway, 1964). Here, Cab Calloway sings an unaccompanied A_4 during a scalar run near the end of the song, 2:10 to 2:20 min into the recording. The A_4 segment analyzed for this paper occurs at 2:13 of the performance. For the second example, the musical theater singer Tony Vincent is shown in a live performance in Beijing singing "Love changes everything" from Andrew Lloyd Webber's musical "Aspects of Love." The performance is from the 2008 Summer Olympics and is freely available on YouTube (Vincent, 2008).

Figure 9(a) shows a video frame of Pavarotti singing an /ɔ/ vowel on the pitch A_4 from the phrase "sei tu forse un uom?" four bars before the beginning of "Vesti La Giubba." The vowel is from the word *uom*, briefly sung while unaccompanied by the orchestra, from which the A_4 is taken. The head shape and the mouth shape are highlighted with white lines. The images were processed using a MATLAB script that found the ratio of mouth area to frontally-projected head area by defining two polygons. From these, the absolute area of the mouth was estimated from mean head size measurements of ordinary individuals. Results are shown in Table I for this note and several other notes in the aria. For the A_4 shown in the figure, the mouth/head area ratio is 0.0291 or about 3%.

TABLE I. Mouth-to-head area ratios.

Note	Ratio	Vowel
	Male operatic (Luciano Pavarotti)	
$D^{\#}_4$	0.0137	/e/
E_4	0.0205	/ɑ/
$F^{\#}_4$	0.0288	/ʊ/
G_4	0.0290	/ɑ/
A_4	0.0291	/ɔ/
	Male belt (Cab Calloway)	
$D^{\#}_4$	0.0170	/u/
E_4	0.0364	/o/
$F^{\#}_4$	0.0614	/a/-/o/ (diphthong)
G_4	0.0662	/a/
A_4	0.0840	/a/

Estimating a 350 cm^2 head area for a slightly larger than normal male, the absolute mouth area for A_4 is about 10 cm^2. (Precision in this estimate is not important because the differences between the examples described here are very large).

The authors know little about the rest of the vocal tract of Pavarotti, other than he was a large man with a wide neck. Assuming his supraglottal vocal tract length to be about the same as that of the baritone (Pavarotti was a tenor, but larger than most), assuming a wider pharynx (about 4 cm^2) and assuming a 0.3 cm^2 narrowed epilarynx tube (because of a strong ring in his voice), the approximate 10 cm^2 mouth area can be extrapolated backward from the general MRI shape of the lyric baritone. For results the authors return to Fig. 8, third row. This is obviously at best an intelligent guess, but it serves to produce one caricature of a classical male operatic singing shape, the inverted megaphone mouth shape. As an additional vocal tract modification for operatic singing, a slight larynx lowering (often taught in vocal studios) was included by shortening the trachea by 1.5 cm.

For the jazz singer Calloway, Fig. 9(c) shows the mouth shape on the same pitch and vowel. The mouth/head area ratio is 0.084 (see also Table I), nearly three times larger than that for Pavarotti. With a 30 cm^2 mouth opening, a backward extrapolation from this mouth shape to a speech-like pharynx and epilarynx tube is shown in Fig. 8, bottom row. A slight larynx raising is part of a belt production, which was simulated by lengthening the trachea by 1.2 cm and shortening the supraglottal tract proportionately. The supraglottal tract was shortened because mouth-corner retraction is also part of belting.

Consider now the inertograms of Fig. 8 (right panel). Vertical lines are drawn for the pitch A_4 (440 Hz) and eight higher harmonics. Note that F_0 is safely in the inertance region below F_1 for all four configurations. However, the second harmonic (880 Hz) is above F_1 for all but the megaphone (Calloway) mouth shape. For the inverted megaphone mouth shape (the shape extrapolated from Pavarotti's mouth), $2F_0$ is in both the subglottal compliance region and the supraglottal inertance region (below F_2). Because the trachea is slightly shorter than for the speech vowel, the first subglottal resonance (F_1^1) overlaps with the second supraglottal resonance (F_2) to offer combined reinforcement to $2F_0$. The third harmonic ($3F_0$) benefits from being near the highest inertance point on the upskirt of F_2. In addition, the overall inertance in the 2500 Hz region is increased (relative to the original baritone inertograms) because of epilarynx narrowing. The sixth harmonic would be predicted to be strong.

For the Calloway mouth shape, $2F_0$ should receive an exceptionally large boost from the supraglottal inertance just below F_1. The third harmonic is not expected to be strong with the megaphone mouth shape at this pitch.

The measured spectra from the two singers [Figs. 9(b) and 9(d)] confirm some of the predictions of the inertograms. All audio samples were in the AAC format, a high quality audio encoding scheme with a sampling rate of 44.1 kHz. These were analyzed with PRAAT (Boersma and Weenick, 2009) using a narrow band fast Fourier transform (FFT) with Gaussian windowing. The window length was set to 0.06 s, resulting in a 21.64 Hz bandwidth (narrow band) analysis. The dB threshold was set to 60 dB. Objections may be raised about performing spectral analyses on highly compressed and processed YouTube recordings. While these objections are generally valid, they do not affect the general conclusions reached here. The authors have uploaded male high-pitched singing sounds to YouTube and analyzed their spectral content pre- and postuploading, and then again after downloading. The major harmonic amplitudes differed only by 1–2 dB. Also, independently-extracted spectra from non-compressed original recordings by Schutte *et al.* (2005) confirm the spectra for one of the artists, Pavarotti.

The magnitude spectrum for Pavarotti in Fig. 9(b) shows that F_0, $2F_0$, and $3F_0$ are all strong, particularly $2F_0$ and $3F_0$. Is this spectrum predicted by the inertograms of Fig. 8 (second from bottom)? As discussed above, the lower three harmonics are predicted to be reinforced by favorable supraglottal inertance (and subglottal compliance in the case of $2F_0$). But harmonics $6F_0$, $7F_0$, and $8F_0$ are also collectively strong in the recording. From Fig. 8, only $6F_0$ is predicted to be strong. Since the authors do not know the precise epilaryngeal dimensions of Pavarotti (length and diameter), it is possible that $7F_0$ and $8F_0$ may also be reinforced by the narrowed epilarynx tube and the clustering of F_3, and F_4 that produces the operatic ring (Bartholomew, 1934), but further exploration would be needed.

The spectrum of Calloway [Fig. 9(d)] shows an exceptionally strong second harmonic $2F_0$. It rises 30 dB above the energy of F_0 and 20 dB above the energy of $3F_0$. (For Pavarotti, the energies in $2F_0$ and $3F_0$ were about 10 dB above the energy in F_0.) Figure 8 (bottom right) predicts this strong second harmonic on the basis of a high larynx and a megaphone mouth shape.

Figure 10 shows two more examples of males singing high pitches, opera singer Roberto Alagna, and musical theater singer Tony Vincent. Mouth-to-head area ratios were 0.0563 and 0.0289, respectively, an approximate 2:1 difference. The spectrum for Vincent is again characteristic of a strong second harmonic, little fundamental (basically in the noise), and only a moderate amount of energy in the singer's formant cluster (harmonics 6–8). In contrast, Alagna has

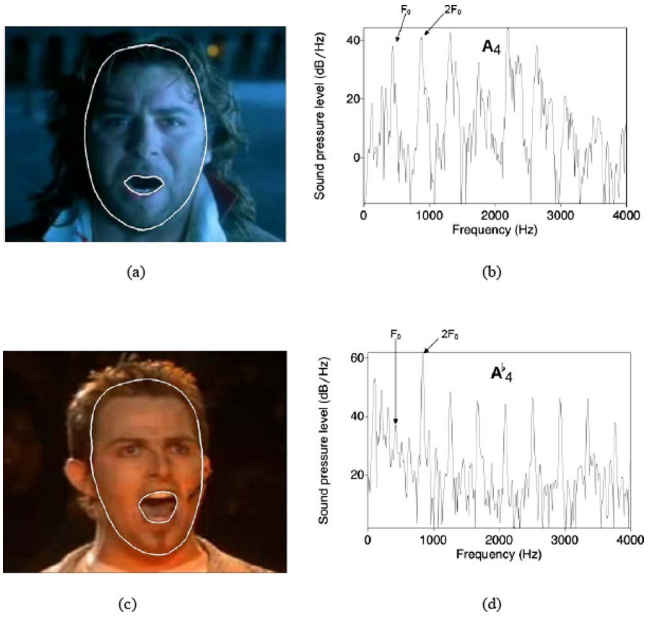


FIG. 10. (Color online) (a) Mouth area and head area for Roberto Alagna singing A_4 on an /ɔ/ vowel. (b) Corresponding frequency spectrum. (c) Mouth area and head area for Tony Vincent singing A-flat$_4$ on an /a/ vowel, and (d) corresponding frequency spectrum.

more balanced energy in the lower harmonics and a strong singer's formant cluster to boost $5F_0$ and $6F_0$. Some difference in the harmonic energy distribution is not surprising on the basis of the inertogram of Fig. 8 because the inertance regions are broad and precise tuning of harmonics to formants is not necessary.

IV. HARMONIC LIFTING OVER FORMANTS

The authors return to some pedagogical issues. It is hypothesized that dealing with harmonic-formant interaction is an essential component of vocal technique. In classical voice pedagogy, the "lifting" of a harmonic over the formant when pitch is changed is part of managing the *passaggi* in the voice, "covering" the high notes, or "vowel modification" (Miller, 1986). Vocal instabilities (pitch jumps, subharmonics, or occasionally aperiodic vibration) can occur when a harmonic passes through a formant on a pitch change (Titze *et al.*, 2008). As was seen in the data, the inertance changes quickly near the formants. If vocal fold vibration is highly facilitated by supraglottal inertance, a sudden change can destabilize the modes of vibration. Thus, a vocalist who relies on source-vocal tract interaction to boost the power of his voice must learn to modify the vowel to seek out as much reinforcement as possible for each harmonic.

For the cases studied here, the narrowed epilarynx tube for the operatic shape in Fig. 8 (second from bottom) has the effect of increasing supraglottal inertance over the entire frequency range. This gives the singer the opportunity to reinforce many harmonics on overlapping skirts between the formants. There are no wide "dead spots," only a few small dips above the formants (recall Fig. 8, third inertogram versus the top inertogram). There is an expected asymmetry between the strength of a harmonic directly below a formant and one directly above a formant. Examination of 47 spectra of singers displayed in Miller (2008) reveals that 37 spectra show this asymmetry and only 10 show approximate symmetry. This is a strong verification of the nonlinear source-filter theory. In linear source-filter coupling, symmetry in harmonic energy around the formants is predicted because vocal tract reactance does not affect the source and the vocal tract transfer function is symmetric around the formant. Thus, whether reactance is positive or negative should have no effect on the strength of the harmonic. (A small asymmetry does exist because of the gradual spectral decay at the source, but that was taken into account when adding up the profound asymmetries in the above-mentioned 47 spectrograms.)

For jazz and theater belt productions, the second harmonic, which is characteristic of the male quality (and female belt quality) according to Schutte and Miller (1993) and Neumann *et al.* (2005), needs to be carefully managed by male singers in their high-pitch ranges. The vibration regime of the vocal folds could easily be destabilized by a sharp change in $2F_0$ reinforcement. The register could easily flip from modal to falsetto without second harmonic reinforcement (Titze, 2008a). Classically-trained singers prevent this possible destabilization by covering or modifying any vowels that would have a wide-open mouth shape (Appelman, 1967). Centralized vowels such as /ɔ/, /ʊ/, /ɛ/, or /ɪ/ keep $2F_0$ in positive inertance territory below F_2. An exercise used by Enrico Caruso, a famous tenor of the first half of the 20th century, is based on a gradual change from the /ɑ/ vowel to the /ɔ/ vowel for high notes (Coffin, 1987). Some vocal pedagogues have gone on record to describe "highly favored" vowels for classically-trained baritones and tenors as they transit into their highest pitches (Coffin, 1987).

Male belters, on the other hand, purposely do not modify toward these centralized vowels. With higher F_0, they open the mouth ever further than for the speech /ɑ/, all the while raising the larynx. The combined action raises F_1 (Bjorkner, 2008), thereby keeping $2F_0$ below F_1. There is an upper limit to this strategy, however. Belters generally break into falsetto register when F_1 can no longer be raised in modal register, which by nature of its characteristic airflow requires a strong second harmonic (Sundberg *et al.*, 1993). If the $2F_0$ interaction with F_1 has not been smoothed out with much practice, a noticeable timbre change will occur.

By lowering the larynx, the tracheal compliance region can be raised in frequency to maintain a chest voice all the way to C_5 (the trachea will be shortened). In Fig. 8, if the subglottal compliance tear-drop were to shift upward by about 200 Hz, $2F_0$ would benefit from tracheal reinforcement. Some very robust tenors sing their top notes with a lowered larynx and may make use of subglottal (chest) reinforcement. However, the detailed acoustic analysis of tracheal resonance in singing is left for a future study.

V. A HISTORICAL NOTE

In the year 1831, a revolution took place in the male singing voice in Italy. The French tenor Gilbert Duprez (1806–1896) presented a C_5 in "chest" voice in Rossini's opera *William Tell*. It was referred to as *Do di petto*, C in

chest. Repeat performances in his own country with this production brought about much critique in the media, and legend has it that it ultimately led to the suicide of one of his rival tenors, Adolphe Nouritt, who could never produce this "chest" sound (Walker and Hibbard, 1992). Prior to this, male high voice was likely to be produced in a much lighter register, resembling more of what today would call a *leggiero* sound or even a *tenorino* production. Rossini did not care for Duprez's sound, having himself led vocal pedagogy through the *bel canto* era. He referred to it as "the death throes of a chicken (Holland, 1999)." Other critics thought the sound was new and exciting, more capable of expressing extreme vocal drama. In 1840, the production became fashionable and was adapted by Verdi and other opera composers as the sound of a heroic male character. But the productions have now been highly groomed, and the authors do not know what the original sound was.

Featuring the belt quality described here over long and repeated notes may also be treacherous. Prolonged mouth and jaw stretching, along with muscular stretching in the vocal folds to maintain an extremely high pitch, can easily fatigue the voice. Duprez had a short career, retiring at age 49 to become a teacher for the remaining 41 years of his life. The authors do not know if his high C's were belt-like (with a high larynx and a wide-open mouth) or in the operatic style described here (with a slightly lowered larynx, moderate mouth opening, and tracheal resonance). In Duprez's day, the term *belt* had not been invented. In the last century and a half, the male singing voice has been cultivated to the point that any blend between falsetto (the boy voice that does not exhibit a strong second harmonic) and the male belt (which produces the strongest second harmonic) can be obtained with clever vowel modification and registration at the source. By lowering or raising the larynx, and by using either the megaphone or inverted megaphone shape, tenors and high baritones can have more freedom to explore a variety of sound spectra, producing both warmth and brilliance.

VI. CONCLUSIONS

The linear source-filter theory, successfully applied to male speech, is likely to be applicable to male singing when pitches are low enough that significant harmonic-formant interaction does not occur. However, for a male singer with at least a two octave range, pitches in the higher octave, beginning around C_4, may require special vocal tract shapes to enhance self-sustained vocal fold oscillation. Highly gifted singers, with a vocal fold layered structure that easily sustains vocal fold oscillation (Hirano, 1975), may not rely heavily on source-tract interaction. Any vowel shape is possible, but most singers choose caricatures of certain vowels to reinforce a collection of harmonics. This is no need for exact "tuning" of formants to harmonics, as in many man-made musical instruments, but rather a need to find regions *between* the formants where supraglottal inertive reactance and subglottal compliant reactance can be exploited. Some vowel articulation is then still possible, but around a specific vocal tract caricature. In jazz or theater belt, the vowels /æ/ and /a/ provide the highest F_1 so that both F_0 and $2F_0$ can always be kept below F_1. In opera or art song, centralized vowels such as /ʊ/ are often used to lower F_1 so that $2F_0$ (and ultimately F_0 itself) can be lifted over F_1 with a pitch change. In a paper to follow, a similar analysis will be given for female singers across different styles.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant No. 5R01 DC004224-08 from the National Institute on Deafness and Other Communication Disorders. Special thanks are given to Dr. Brad Story who provided the vocal tract shapes of the baritone singer.

Adachi, S., and Sato, M. (**1996**). "Trumpet sound simulation using a two-dimensional lip vibration model," J. Acoust. Soc. Am. **99**, 1200–1209.
Alagna, R. (**2000**). "E lucevan le stele" from Puccini's opera *Tosca* was analyzed. The video is from a filmed performance in 2000, freely available on YouTube retrieved from http://www.youtube.com/watch?v=f6urNGBR95w (Last viewed 2/4/2009).
Alipour, F., Scherer, R. C., and Finnegan, E. (**1997**). "Pressure-flow relationships during phonation as a function of adduction," J. Voice **11**, 187–194.
Appelman, D. R. (**1967**). *The Science of Vocal Pedagogy: Theory and Application* (Indiana University Press, Bloomington, IN).
Ayers, D. (**1998**). "Observation of the brass player's lips in motion," J. Acoust. Soc. Am. **103**, 2873–2874.
Bartholomew, W. (**1934**). "A physical definition of good voice quality in the male voice," J. Acoust. Soc. Am. **6**, 25–33.
Bergan, C., Titze, I. R., and Story, B. (**2004**). "The perception of two vocal qualities in a synthesized vocal utterance: Ring and pressed voice," J. Voice **18**, 305–317.
Bjorkner, E. (**2008**). "Musical theatre and opera singing—Why so different? A study of subglottal pressure, voice source, and formant frequency characteristics," J. Voice **22**, 533–540.
Boersma, P., and Weenick, D. (**2009**). "Doing phonetics by computer," retrieved from www.praat.org (Last viewed 2/4/2009).
Calloway, C. (**1964**). "St. James Infirmary" video, recorded live in 1964 on the Ed Sullivan Show, freely available on YouTube retrieved from http://www.youtube.com/watch?v=DAmxXrjVVsM (Last viewed 2/4/2009).
Chan, R., and Titze, I. R. (**2006**). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," J. Acoust. Soc. Am. **119**, 2351–2362.
Coffin, B. (**1987**). *Coffin's Sounds of Singing: Principles and Applications of Vocal Techniques With Chromatic Vowel Chart*, 2nd ed. (The Scarecrow, Metuchen, NJ).
Dromey, C., Sathopoulos, E. T., and Sapienza, C. M. (**1992**). "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," J. Voice **6**, 44–54.
Estill, J. (**1988**). "Belting and classic voice quality: Some physiological differences," Med. Probl. Perform. Art. **3**, 37.
Fant, G. (**1960**). *The Acoustic Theory of Speech Production* (Moulton, The Hague).
Fant, G. (**1986**). "Glottal flow: Models and interaction," J. Phonetics **14**, 393–399.
Fant, G., and Lin, Q. (**1987**). "Glottal voice source-vocal tract acoustic interaction," J. Acoust. Soc. Am. **81**, S68.
Flanagan, J. L. (**1968**). "Source-system interaction in the vocal tract," Ann. N.Y. Acad. Sci. **155**, 9–17.
Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception* (Springer-Verlag, New York).
Fletcher, N., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments* (Springer, New York).
Fletcher, N. H. (**1993**). "Autonomous vibration of simple pressure-controlled valves in gas flows," J. Acoust. Soc. Am. **93**, 2172–2180.
Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.
Henrich, N. (**2006**). "Mirroring the voice," Logoped. Phoniatr. Vocol. **31**, 3–14.
Hirano, M. (**1975**). "Phonosurgery: Basic and clinical investigations," Otologia (Fukuoka) **21**, 239–440.

Holland, B. (**1999**). Critics Notebook; New York Times Online; http://query.nytimes.com/gst/fullpage.html?res=9903EFDB1430F930A25753C1A96F958260 (Last viewed 2/4/2009).
Hollien, H. (**1974**). "On vocal registers," J. Phonetics **2**, 125–143.
Hollien, H. (**1983**). "A review of vocal registers," in *Transactions of the Twelfth Symposium on Care of the Professional Voice*, edited by V. Lawrence (Voice Foundation, New York, NY), pp. 1–6.
Holmberg, E. B., Hillman, R. E., and Perkell, J. S. (**1988**). "Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice," J. Acoust. Soc. Am. **84**, 511–529.
Jiang, J., and Tao, C. (**2007**). "The minimum glottal airflow to initiate vocal fold oscillation," J. Acoust. Soc. Am. **121**, 2873–2881.
Joliveau, E., Smith, J., and Wolfe, J. (**2004**). "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116**, 2434–2439.
Lilljencrants, J. (**1985**). "Speech synthesis with a reflection-type analog," Doctoral thesis, Royal Institute of Technology, Stockholm, Sweden.
Miller, D. G. (**2008**). *Resonance in Singing: Voice Building Through Acoustic Feedback* (Inside View, Princeton, NJ).
Miller, D. G., and Schutte, H. K. (**2005**). "Mixing the registers: Glottal source or vocal tract?," Folia Phoniatr Logop **57**, 278–291.
Miller, R. (**1986**). *The Structure of Singing: System and Art in Vocal Technique* (Schirmer Books, New York).
Neumann, K., Schunda, P., Hoth, S., and Euler, H. A. (**2005**). "The interplay between glottis and vocal tract during the male passaggio," Folia Phoniatr Logop **57**, 308–327.
Pavarotti, L. (**1994**). Singing the aria "Vesti La Giubba" from Leoncavallo's opera *I Pagliacci*. Video from a 1994 performance at the Metropolitan Opera, freely available on YouTube retrieved from (http://www.youtube.com/watch?v=Ky271W94VHA) (Last viewed 2/4/2009).
Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo), pp. 305–328.
Schutte, H. K., and Miller, D. G. (**1993**). "Belting and pop, nonclassical approaches to the female middle voice: Some preliminary considerations," J. Voice **7**, 142–150.
Schutte, H. K., Miller, D. G., and Duijnstee, M. (**2005**). "Resonance strategies revealed in recorded tenor high notes," Folia Phoniatr Logop **57**, 292–307.
Skilling, H. H. (**1966**). *Electrical Engineering Circuits*, 2nd ed. (Wiley, New York).
Sondhi, M. M., and Schroeter, J. (**1987**). "A hybrid time-frequency domain articulatory speech synthesizer," IEEE Trans. Acoust. Speech Signal Process. **35**, 955–967.
Stathopoulos, E., and Sapienza, C. (**1993**). "Respiratory and laryngeal function of women and men during vocal intensity variation," J. Speech Hear. Res. **36**, 64–75.
Stathopoulos, E., and Sapienza, C. (**1997**). "Developmental changes in laryngeal and respiratory function with variation in sound pressure level," J. Speech Lang. Hear. Res. **40**, 595–614.
Stevens, K. (**1999**). *Acoustic Phonetics (Current Studies in Linguistics)* (MIT, Cambridge, MA).
Stone, R. E., Cleveland, T. F., Sundberg, P. J., and Prokop, J. (**2003**). "Aerodynamic and acoustical measures of speech, operatic, and Broadway vocal styles in a professional female singer," J. Voice **17**, 283–297.
Story, B., Laukkanen, A.-M., and Titze, I. R. (**2000**). "Acoustic impedance of an artificially lengthened and constricted vocal tract," J. Voice **14**, 455–469.
Story, B., Titze, I. R., and Hoffman, E. A. (**2001**). "The relationship of vocal tract shape to three voice qualities," J. Acoust. Soc. Am. **109**, 1651–1667.
Story, B. H. (**1995**). "Speech simulation with an enhanced wave-reflection model of the vocal tract," Ph.D. thesis, University of Iowa, Iowa City, IA.
Story, B. H. (**2005**). "Synergistic modes of vocal tract articulation for American English vowels," J. Acoust. Soc. Am. **118**, 3834–3859.
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.
Story, B. H., Titze, I. R., and Hoffman, E. A. (**1996**). "Vocal tract area functions from magnetic resonance imaging," J. Acoust. Soc. Am. **100**, 537–554.
Sundberg, J. (**1974**). "Articulatory interpretation of the 'singing formant'," J. Acoust. Soc. Am. **55**, 838–844.
Sundberg, J. (**1977**). "The acoustics of the singing voice," Sci. Am. **236**, 82–91.
Sundberg, J. (**1995**). "Vocal fold vibration patterns and modes of phonation," Folia Phoniatr Logop **47**, 218–228.
Sundberg, J., Gramming, P., and Lovetri, J. (**1993**). "Comparisons of pharynx, source, formant, and pressure characteristics in operatic and musical theatre singing," J. Voice **7**, 301–310.
Sundberg, J., Thalén, M., Alku, P., and Vilkman, E. (**2004**). "Estimating perceived phonatory pressedness in singing from flow glottograms," J. Voice **18**, 56–62.
Titze, I. R. (**1984**). "Parameterization of the glottal area, glottal flow, and vocal fold contact area," J. Acoust. Soc. Am. **75**, 570–580.
Titze, I. R. (**1988a**). "The physics of small-amplitude oscillation of the vocal folds," J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**1988b**). "A framework for the study of vocal registers," J. Voice **2**, 183–194.
Titze, I. R. (**2000**). *Principles of Voice Production* (National Center for Voice and Speech, Denver, CO).
Titze, I. R. (**2006**). "Theoretical analysis of maximum flow declination rate versus maximum area declination rate in phonation," J. Speech Lang. Hear. Res. **49**, 439–447.
Titze, I. R. (**2008a**). "Nonlinear source-filter coupling in phonation: Theory," J. Acoust. Soc. Am. **123**, 2733–2749.
Titze, I. R. (**2008b**). "The human instrument," Sci. Am. **298**, 94–101.
Titze, I. R., Riede, T., and Popolo, P. S. (**2008**). "Nonlinear source-filter coupling in phonation: Vocal exercises," J. Acoust. Soc. Am. **123**, 1902–1915.
Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.
Titze, I. R., and Story, B. H. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.
Vennard, W. (**1967**). *Singing: Mechanism and Technique* (Carl Fischer, New York, NY).
Vincent, T. (**2008**). "Love changes everything" from Andrew Lloyd Webber's musical "Aspects of Love." The performance is from the 2008 Summer Olympics and is freely available on YouTube retrieved from http://www.youtube.com/watch?v=flbqN3hospI (Last viewed 2/4/2009).
Walker, E., and Hibbard, S. (**1992**). "Article on Aldophe Nouritt (1802–1839)," in *New Grove Dictionary of Opera*, edited by S. Sadie (Oxford University Press, New York, NY).
Yanagisawa, E., Estill, J., Kmucha, T., and Leder, S. B. (**1991**). "Supraglottic contributions to pitch raising: Videoendoscopic study with spectral analysis," Ann. Otol. Rhinol. Laryngol. **100**, 19–31.

Theoretical limitations of the elastic wave equation inversion for tissue elastography

Ali Baghani,[a] Septimiu Salcudean, and Robert Rohling[b]
Department of Electrical and Computer Engineering, University of British Columbia, Kaiser Building, 2332 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada

(Received 9 December 2008; revised 23 June 2009; accepted 24 June 2009)

This article examines the theoretical limitations of the local inversion techniques for the measurement of the tissue elasticity. Most of these techniques are based on the estimation of the phase speed or the algebraic inversion of a one-dimensional wave equation. To analyze these techniques, the wave equation in an elastic continuum is revisited. It is proven that in an infinite medium, harmonic shear waves can travel at any phase speed greater than the classically known shear wave speed, $\sqrt{\mu/\rho}$, by demonstrating this for a special case with cylindrical symmetry. Hence in addition to the mechanical properties of the tissue, the phase speed depends on the geometry of the wave as well. The elastic waves in an infinite cylindrical rod are studied. It is proven that multiple phase speeds can coexist for a harmonic wave at a single frequency. This shows that the phase speed depends not only on the mechanical properties of the tissue but also on its shape. The final conclusion is that the only way to avoid theoretical artifacts in the elastograms obtained by the local inversion techniques is to use the shear wave equation as expressed in the curl of the displacements, i.e., the rotations, for the inversion. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3180495]

PACS number(s): 43.80.Jz, 43.80.Ev, 43.80.Vj, 43.80.Qf [CCC] Pages: 1541–1551

I. INTRODUCTION

Elastography is a fast developing field of medical imaging which strives to provide images of the mechanical properties of tissue.[1–8] The main mechanical property of interest is the elasticity or Young's modulus E. Other mechanical properties of interest include viscosities, relaxation times, nonlinearity parameters, harmonics, poro-elasticity, and Poisson's ratio.[8–22]

The majority of the elastography techniques have three components in common:

- an *excitation mechanism* which creates quasi-static or dynamic deformations in the tissue,[23–31]
- a medical imaging system augmented by a *motion estimation technique* which is capable of providing displacement (or velocity) images of the tissue while it is being deformed,[25,32–41] and
- an *inversion technique* which transforms the displacement images into elasticity images, the so called elastograms.[25,42–54]

For the inversion, different approaches have been devised. Most of them, however, fall under one of the two following categories.

- *Global inversion techniques*.[44–48,55–57] These techniques typically use finite element modeling of the medium. The goal is to find the best local elasticity values which would create the same displacement pattern as those observed inside the tissue, under similar boundary and excitation conditions. To find the optimum elasticity values, the forward problem is solved iteratively. In each iteration the elasticity values are adjusted until the computed displacements from the model match the measured displacement from the tissue. In addition to the measured displacements, these techniques typically require further information about the shape of the tissue, its boundary conditions, and the excitation.
- *Local inversion techniques*.[11,25,27–30,37,50–52,58–61] Typically in these techniques, a particular *wave equation* is assumed to govern the displacements. This partial differential equation relates the spatial derivatives to the temporal derivatives of the displacement with the local elasticity as the coefficient. If the displacement measurements are available over a range of space and time, the elasticities can be found either by an algebraic inversion of the wave equation or by estimating the phase speed from the gradient of the phase. Section II is devoted to the presentation of these techniques.

It is the objective of this research to study the theoretical limitations of the local inversion techniques for estimating the tissue elasticity, study the effects of the formulation of the wave equation on the inversion results, and study the effects of the choice of the exciter and imaging technique.

To this end, the general form of the elastic wave equation in a linear continuum is studied. The hypothesis to be proven is that the *phase speed*, which is the final estimated quantity in many of the local inversion techniques, does not exactly represent the local mechanical properties of the tissue. The hypothesis is proven by showing that in addition to

[a]Author to whom correspondence should be addressed. Electronic mail: baghani@ece.ubc.ca
[b]Also at Department of Mechanical Engineering, University of British Columbia

the local mechanical properties of the tissue, the phase speed depends on the geometry of the wave as well as the geometry of the tissue itself. This effect is known as the Lamb wave effect specially in the context of thin plates.[62] The practical implication is that artifacts are inevitably present in the elastograms, no matter how well the algorithm is implemented or how accurate the displacement measurements are. Nevertheless, many of the elastography techniques have been proven clinically to provide valuable information about the elasticity of the tissue. The excitation in these techniques is chosen so that the actual displacements created in the tissue satisfy the assumptions made, which in turn justifies the inversion techniques used. These methods include quasi-static constant strain compression, pulsed excitation with external exciters, and acoustic-radiation-force-based techniques.

A previous study of the limitations of the local inversion techniques for estimating the tissue elasticity can be found in Ref. 63. In that work the authors experimentally measured the displacements caused by a circular piston inside rectangular blocks of tissue mimicking material and meat specimens. In their experiments the exciter would start to vibrate at a single frequency (monochromatic excitation). Between the transient and steady state regimes of vibration, a region was identified when transient shear waves are propagating in the medium and phase speed measurements can be carried out. The deleterious effects of the medium boundaries were dealt with by rejecting the steady state regime for phase speed measurements. The effects of the boundary conditions imposed by the exciter were also studied. An approximate model, the Rayleigh–Sommerfeld solution, and the Green's function were used in this analysis. They concluded that the size of the exciter affects the measured phase speed. Another conclusion was that at very low frequencies, the effects of the longitudinal wave on the phase speed cannot be neglected. The final preferred method was to use a pulsed transient excitation with a small piston to get more accurate viscoelastic measurements.

The analysis of the aforementioned article is limited to the particular configuration used, which is very similar to a point source on a semi-infinite medium. In many elastography applications, to get a desired level of displacements inside tissue, it is of interest to use exciters with a larger footprint. The authors have shown previously that if the size of the exciter is comparable to the size of the sample, the phase speed observed is the extensional wave speed.[64] In the present article they look at the problems associated with phase speed measurements from a more general point of view, and discuss some of the phenomena that affect the accuracy of the measurements. They will show that neglecting these phenomena results in unaccounted for artifacts in the computed elastograms.

The authors start by a brief review of the state-of-the-art local inversion techniques in Sec. II. Section III is a critique on the use of the terminology associated with the elastic waves. The main theoretical results of the article are presented in Sec. IV. Simulation results are presented in Sec. V. Section VI is devoted to the conclusion.

II. LOCAL INVERSION OF THE WAVE EQUATION

The mechanical wave equation in an isotropic elastic continuum written in the Cartesian coordinate system is[62]

$$\rho\frac{\partial^2 u_i}{\partial t^2} = (\lambda+\mu)\frac{\partial}{\partial i}\Delta + \mu\nabla^2 u_i, \quad i = x,y,z, \tag{1}$$

where $\mathbf{x}=[x,y,z]^T$ are the Cartesian coordinates which define the spatial location of the point, t is the time variable, $\mathbf{u}(\mathbf{x},t)=[u_x,u_y,u_z]^T$ is the displacement field, and the coefficients $\rho(\mathbf{x})$, $\lambda(\mathbf{x})$, and $\mu(\mathbf{x})$ are the density, the first, and the second Lamé constants, respectively. Here, the authors have adopted the notation of Kolsky.[62] The dilatation Δ is defined by

$$\Delta = \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} + \frac{\partial u_z}{\partial z}, \tag{2}$$

and determines the voluminal changes as the wave propagates. ∇^2 is the Laplacian operator:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{3}$$

The authors denote Poisson's ratio and Young's modulus by ν and E, respectively.

The material of interest in the field of elastography is tissue *in vivo*. This material is found to be nearly incompressible,[51,65] hence the values of Δ are typically very small. Other implications of the incompressibility in terms of the mechanical properties are as follows:

$$\nu \approx 0.5, \quad \lambda \gg \mu \quad E \approx 3\mu. \tag{4}$$

In the rest of this article it is assumed that the relations (4) hold.

To obtain a clinically useful image of the soft tissue, the contrast should be based on a physical quantity which has a high degree of variability among different tissue types. The density of most soft tissue is close to the density of water,[66] 1000 $\mathrm{kg/m^3}$ so ρ is not a candidate. Neither is the first Lamé constant λ which is found to be close to the λ of water,[67] i.e., 2.3 GPa. On the other hand, μ and $E=3\mu$, which determine the stiffness of the tissue, vary over a wide range from a few to a few hundred kilopascals.[68] Moreover, a change in the stiffness of the tissue is often associated with pathology.[69–71] Hence elastograms are preferentially based on the tissue elasticity $E=3\mu$. Given the displacements as a function of time and space, the most that can be recovered from Eq. (1) are the ratios λ/ρ and μ/ρ. In practice, the elastograms are based on E/ρ. Since the variability of ρ is small, the obtained contrast is almost the same as the one obtained by using the absolute value of elasticity E.

A number of issues arise when medical imaging devices are used to estimate the displacement fields. In many cases it is not possible to acquire all three components of the displacement field, i.e., u_x, u_y, and u_z. In other cases the acquired displacement field may not be available over a volume. For instance, in most cases of ultrasound elastography with two-dimensional probes, only $u_x(x,y)$ is available and only on a single plane $z=$constant and not over a volume. Some customized ultrasound systems can estimate the dis-

placements over a volume and in multiple directions.[37,41,72] MRI based techniques on the other hand can acquire the full displacement field over a volume, at the expense of longer acquisition times.[14,25,50,51,73–75]

Different approaches to the inversion of the wave equation have been taken. As mentioned earlier the dilatation $\Delta(\mathbf{x},t)$ levels in soft tissue are well beyond the accuracy of the measurement systems. Therefore the wave equation (1) cannot directly be inverted to obtain μ/ρ. Some state-of-the-art approaches to tackle this problem are as follows:

- *Assuming zero dilatation or zero pressure gradient.*[30,36,37,58,59] In this approach it is assumed that the dilatation is identically zero, $\Delta \equiv 0$, or that the pressure has no spatial variation in the medium. These assumptions reduces the wave equation to

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \mu \nabla^2 u_i, \quad i = x,y,z. \tag{5}$$

  Note that for an anisotropic inversion all three components are necessary. If isotropy is assumed, measuring one component of the displacement field, for instance, u_x, is enough. In this case, it is natural to choose the measurement axis (the x-axis) aligned with the axis on which the wave causes the greatest displacements. If the measurement is only available on a plane, for instance, $u_x(x,y)$, or on a single line, $u_x(x)$, then the following assumptions are usually made

$$\nabla^2 u_x = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) u_x \approx \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) u_x \approx \frac{\partial^2}{\partial x^2} u_x. \tag{6}$$

- *Assuming a shear wave equation.*[52] In this case it is assumed that the wave is a plane shear wave. If the wave causes the greatest displacements in the x-direction, then this component of the displacement field is measured and assumed to satisfy a one-dimensional (1D) shear wave equation,

$$\rho \frac{\partial^2 u_x}{\partial t^2} = \mu \frac{\partial^2 u_x}{\partial y^2}. \tag{7}$$

- *Assuming a thin rod (extensional) wave equation.*[64] When the size of a compressing exciter is comparable to the dimensions of the tissue being imaged the tissue behaves like a thin rod experiencing extensional waves. Under these conditions the component of the displacement field parallel to the direction of the excitation, say, the x-component, satisfies

$$\rho \frac{\partial^2 u_x}{\partial t^2} = E \frac{\partial^2 u_x}{\partial x^2}. \tag{8}$$

- *Writing the shear wave equation for the rotations.*[14,53,72–75] Define the rotation $\mathbf{w}(\mathbf{x},t)$ as half the curl of the displacement field:

$$\mathbf{w}(\mathbf{x},t) = \tfrac{1}{2} \nabla \times \mathbf{u}(\mathbf{x},t). \tag{9}$$

  From Eq. (1) each component of the rotation field satisfies a separate shear wave equation,

$$\rho \frac{\partial^2 w_i}{\partial t^2} = \mu \frac{\partial^2 w_i}{\partial i^2}, \quad i = x,y,z. \tag{10}$$

  Note that this approach does not make any assumptions, but it requires the measurement over a three-dimensional volume of at least two components of the displacement field for an isotropic inversion. For an anisotropic inversion, all the components are necessary.

The local inversion algorithms (with the exception of the last approach) reduce the general form of the wave equation (1) to a 1D wave equation of the form

$$\rho \frac{\partial^2 u_x}{\partial t^2} = f(E) \frac{\partial^2 u_x}{\partial x^2}, \tag{11}$$

where $f(E)$ is a function of the local elasticity. The quantity

$$c_{ph} = \sqrt{f(E)/\rho} \tag{12}$$

in this 1D equation is called the *phase speed*. For a unidirectional harmonic solution,

$$u_x(x,t) = a \exp(j\omega(x/c_{ph} - t)), \tag{13}$$

this quantity gives the propagation speed of the constant phase surfaces.

Given $u_x(x,t)$, it is possible to invert Eq. (11) and find $f(E)/\rho$. This can be done either by directly solving for $f(E)/\rho$:

$$\frac{f(E)}{\rho} = \frac{\partial^2 u_x/\partial t^2}{\partial^2 u_x/\partial x^2}. \tag{14}$$

This method is sometimes called the algebraic inversion of the differential equation.[50] Another method is to use the concept of the phase speed,

$$\sqrt{f(E)/\rho} = c_{ph} = \frac{\omega}{\frac{\partial}{\partial x} \angle u_x}, \tag{15}$$

sometimes called the phase gradient method.[50]

If $f(E)$ were only a function (even an unknown function) of the local mechanical properties of the tissue, the elastograms based on the phase speed $\sqrt{f(E)/\rho}$ would be free from artifacts. However, as the authors show in Sec. IV, $f(E)$ is not only a function of the local mechanical properties of the medium but also a function of the geometry of the wave and the geometry of the medium. An even more important result is proved; the reduction in Eq. (1) to Eq. (11) may not be possible at all since multiple values of the phase speed can coexist.

III. A CRITIQUE ON THE TERMINOLOGY ASSOCIATED WITH THE WAVE EQUATION

Before the authors move on to their results on the phase speed, they present their choice of the terminology. By the fundamental theorem of vector calculus, or Helmholtz decomposition theorem, any sufficiently smooth and decaying vector field can be written as the sum of a divergence-free and a curl-free vector field.[76] The solutions of the wave equation (1) satisfy the conditions of the theorem, and therefore it

is possible to write a vibration pattern inside an elastic medium as the sum of a divergence-free and a curl-free displacement field.

- *Dilatational wave* or better termed *irrotational component* is a component of the wave which is curl-free. This component of the wave phenomena is solely the result of voluminal changes in the medium.
- *Distortion wave* or better termed *equivoluminal component* is a component of the wave which is divergence-free. This component of the wave phenomena preserves the volumes of the infinitesimal elements.

There is some confusion, however, over the definition and application of the terms "shear wave," "transverse wave," "longitudinal wave," and "compressional wave." These terms are sometimes defined as follows.

(a) *Longitudinal wave* or *compressional wave* is a wave for which the particle velocities are parallel to the phase velocity. For a harmonic wave, the particle velocity would be perpendicular to the equiphase surfaces.
(b) *Transverse wave* or *shear wave* is a wave for which the particle velocities are perpendicular to the phase velocity. For a harmonic wave the particle velocity would be tangent to the equiphase surfaces.

These definitions are very intuitive with respect to the literal meanings of the terms. However a major problem exists with these definitions; to the best knowledge of the authors, there is no theorem proving that the quality of a wave being transversal or longitudinal (taken these definitions) is invariant with respect to the wave equation. In other words, a wave propagating in an infinite elastic medium might be purely longitudinal at one instant but not so at a later instant (even without encountering a boundary and undergoing mode conversion). This severely limits the applicability of these terms.

When these definitions are taken, a connection is usually assumed between longitudinal and irrotational waves on one hand and transverse and equivoluminal waves on the other hand. The origin of this connection could be the study of the simple case of a plane wave in an infinite medium. Indeed a longitudinal plane wave propagating in an infinite medium has no equivoluminal components and consists solely of a dilatational component. This wave propagates at a phase speed of $\sqrt{(\lambda+2\mu)/\rho}$, hence the name *longitudinal wave speed*. Similarly a shear plane wave propagating in an infinite medium has no dilatational components and consists solely of an equivoluminal component. This wave propagates at a phase speed of $\sqrt{\mu/\rho}$, hence the name *shear wave speed*.

A better way to define these terms, however, is to take for them the same definitions as presented for the "dilatational waves" and "distortion waves." This is the accepted nomenclature in the physics literature, specially in the field of the electromagnetic waves.[77]

- *Longitudinal wave* or *compressional wave* is a component of the wave which is curl-free.
- *Transverse wave* or *shear wave* is a component of the wave which is divergence-free.

These are the definitions the authors use in this article and recommend.

A number of comments are in order. It is clear from the definitions that a wave can contain both the longitudinal and transversal components, or may lack one. In some cases one is not interested in the decomposition of the displacement field into these components, but rather in the study of the total displacement itself. However, since each of these components separately satisfies a reduced wave equation with a definite wave speed, it is at other times useful to study them separately. The speed of the longitudinal component and the transverse component is equal to $\sqrt{(\lambda+2\mu)/\rho}$ and $\sqrt{\mu/\rho}$, respectively.

Since the wave equation can be reduced to the two aforementioned wave equations, it is evident that in an infinite elastic medium, a purely longitudinal wave will remain purely longitudinal, and a purely transverse wave will remain purely transverse. This invariance property make the terms well-defined. The drawback is that the intuitive relationship between the directions of the particle velocity and the phase velocity is no longer valid. As a matter of fact, the quality of a wave being longitudinal or transversal could be determined solely from the displacement vector field at a single instant. The wave equation preserves this quality as the time evolves.

As will become clear in the Section IV, the relationship between the phase speed and the longitudinal and shear wave speeds, $\sqrt{(\lambda+2\mu)/\rho}$ and $\sqrt{\mu/\rho}$, is a complex relationship which depends on the geometry of the wave, as well as the geometry of the medium.

IV. PHASE SPEED OF MECHANICAL WAVES IN ELASTIC MEDIUMS

A. Dependence of the phase speed on the geometry of the wave: Infinite mediums

Theorem: *Consider an infinite homogeneous linear elastic medium with density* ρ *and Lamé constants* λ *and* μ. *Given the angular frequency* ω *for a harmonic wave*:

- *For any number c such that*

$$\sqrt{\frac{\mu}{\rho}} \leqslant c < \sqrt{\frac{\lambda+2\mu}{\rho}}, \tag{16}$$

there exists a shear beam form for the harmonic wave for which the phase speed is equal to c.
- *For any number c such that*

$$\sqrt{\frac{\lambda+2\mu}{\rho}} \leqslant c, \tag{17}$$

there exists infinitely many beam forms for the harmonic wave for which the phase speed is equal to c. These waves contain both longitudinal and shear components.

The theoretical derivation which follows can be found in classical textbooks on elastic waves[62,78] as part of the solution to the wave equation in cylindrical coordinate systems. However studying the implications in terms of the phase speed, as summarized in the theorem, is novel.

Proof. Consider the elastic wave equation in the cylindrical coordinate system.

$$\rho\frac{\partial^2 u_r}{\partial t^2} = (\lambda + 2\mu)\frac{\partial \Delta}{\partial r} - \frac{2\mu}{r}\frac{\partial \bar{\omega}_z}{\partial \theta} + 2\mu\frac{\partial \bar{\omega}_\theta}{\partial z}, \tag{18}$$

$$\rho\frac{\partial^2 u_\theta}{\partial t^2} = (\lambda + 2\mu)\frac{1}{r}\frac{\partial \Delta}{\partial \theta} - 2\mu\frac{\partial \bar{\omega}_r}{\partial z} + 2\mu\frac{\partial \bar{\omega}_z}{\partial r}, \tag{19}$$

$$\rho\frac{\partial^2 u_z}{\partial t^2} = (\lambda + 2\mu)\frac{\partial \Delta}{\partial z} - \frac{2\mu}{r}\frac{\partial}{\partial r}(r\bar{\omega}_\theta) + \frac{2\mu}{r}\frac{\partial \bar{\omega}_r}{\partial \theta}, \tag{20}$$

where (r,θ,z) are the cylindrical coordinates and $u = (u_r, u_\theta, u_z)$ is the displacement field. The dilatation in the cylindrical coordinates is given by

$$\Delta = \frac{1}{r}\frac{\partial}{\partial r}(ru_r) + \frac{1}{r}\frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z}, \tag{21}$$

and the rotations are given by

$$2\bar{\omega}_r = \frac{1}{r}\frac{\partial u_z}{\partial \theta} - \frac{\partial u_\theta}{\partial z}, \tag{22}$$

$$2\bar{\omega}_\theta = \frac{\partial u_r}{\partial z} - \frac{\partial u_z}{\partial r}, \tag{23}$$

$$2\bar{\omega}_z = \frac{1}{r}\left[\frac{\partial}{\partial r}(ru_\theta) - \frac{\partial u_r}{\partial \theta}\right]. \tag{24}$$

Consider a cylindrical wave beam propagating along the z-axis with the center of the beam on the z-axis. More specifically assume that u_θ is zero, i.e., the particle displacements are confined to the rz-planes. Also assume that the wave is symmetrical around the z-axis, hence $\partial/\partial\theta$ annihilates the variables. From Eqs. (22) and (24) $\bar{\omega}_r = \bar{\omega}_z = 0$. The reduced wave equation in this case would be

$$\rho\frac{\partial^2 u_r}{\partial t^2} = (\lambda + 2\mu)\frac{\partial \Delta}{\partial r} + 2\mu\frac{\partial \bar{\omega}_\theta}{\partial z}, \tag{25}$$

$$\rho\frac{\partial^2 u_z}{\partial t^2} = (\lambda + 2\mu)\frac{\partial \Delta}{\partial z} - \frac{2\mu}{r}\frac{\partial}{\partial r}(r\bar{\omega}_\theta). \tag{26}$$

The authors are interested in harmonic waves propagating along the z-axis, for instance, in the negative z-direction. The general form of such a wave is

$$u_r = U(r)\exp(i(k_z z + \omega t)), \tag{27}$$

$$u_z = W(r)\exp(i(k_z z + \omega t)), \tag{28}$$

where k_z is the wave number and ω is the angular frequency. $U(r)$ and $W(r)$ determine the shape of the beam. Note that the phase speed of this wave is equal to $c_{ph} = \omega/k_z$. Substitution in Eqs. (21) and (23) yields the expressions for the dilatation and rotation:

$$\Delta(r,z,t) = \left[\frac{\partial U(r)}{\partial r} + \frac{U(r)}{r} + ik_z W(r)\right]\exp(i(k_z z + \omega t)), \tag{29}$$

$$2\bar{\omega}_\theta(r,z,t) = \left[ik_z U(r) - \frac{\partial W(r)}{\partial r}\right]\exp(i(k_z z + \omega t)). \tag{30}$$

The forms (27) and (28) simplify the wave equation (25) and (26),

$$-\rho\omega^2 u_r = (\lambda + 2\mu)\frac{\partial \Delta}{\partial r} + 2i\mu k_z \bar{\omega}_\theta, \tag{31}$$

$$-\rho\omega^2 u_z = ik_z(\lambda + 2\mu)\Delta - \frac{2\mu}{r}\frac{\partial}{\partial r}(r\bar{\omega}_\theta). \tag{32}$$

By eliminating $\bar{\omega}_\theta$ and Δ between these equations, the longitudinal and transversal wave equations are derived:

$$\frac{\partial^2 \Delta}{\partial r^2} + \frac{1}{r}\frac{\partial \Delta}{\partial r} + k_\Delta{}^2\Delta = 0, \tag{33}$$

$$\frac{\partial^2 \bar{\omega}_\theta}{\partial r^2} + \frac{1}{r}\frac{\partial \bar{\omega}_\theta}{\partial r} - \frac{\bar{\omega}_\theta}{r^2} + k_{\bar{\omega}_\theta}{}^2\bar{\omega}_\theta = 0, \tag{34}$$

where k_Δ and $k_{\bar{\omega}_\theta}$ are the geometrical beam numbers defined by

$$k_\Delta{}^2 = \frac{\omega^2}{\frac{\lambda + 2\mu}{\rho}} - k_z{}^2, \tag{35}$$

$$k_{\bar{\omega}_\theta}{}^2 = \frac{\omega^2}{\frac{\mu}{\rho}} - k_z{}^2. \tag{36}$$

In the case that the desired phase speed, c, satisfies Eq. (16) the corresponding choice of the wave number,

$$k_z = \frac{\omega}{c}, \tag{37}$$

results in a real k_Δ and an imaginary $k_{\bar{\omega}_\theta}$. In the other case (17) both k_Δ and $k_{\bar{\omega}_\theta}$ are real. The significance of this will become clear shortly.

The solution to Eqs. (33) and (34) can be found by change of variables from r to $k_\Delta r$ and $k_{\bar{\omega}_\theta} r$, respectively. The resulting equations are Bessel equations of the zeroth and first orders, respectively. The physically meaningful solutions, which have bounded values at $r=0$, are

$$\Delta(r,z,t) = G(z,t)J_0(k_\Delta r), \tag{38}$$

$$\bar{\omega}_\theta(r,z,t) = H(z,t)J_1(k_{\bar{\omega}_\theta} r), \tag{39}$$

where $J_0(\cdot)$ and $J_1(\cdot)$ are Bessel functions of the first kind and of zeroth and first orders, respectively. Now the forms for the dilatation and rotation are given in Eqs. (29) and (30). For these forms to match the solutions (38) and (39), $U(r)$ and $W(r)$ should have the following forms;

$$U(r) = C_1\frac{\partial}{\partial r}J_0(k_\Delta r) + C_2 k_z J_1(k_{\bar{\omega}_\theta} r) = -C_1 k_\Delta J_1(k_\Delta r) + C_2 k_z J_1(k_{\bar{\omega}_\theta} r), \tag{40}$$

$$W(r) = C_1 ik_z J_0(k_\Delta r) + C_2 \frac{i}{r}\frac{\partial}{\partial r}[rJ_1(k_{\bar{\omega}_\theta} r)] = C_1 ik_z J_0(k_\Delta r) + C_2 ik_{\bar{\omega}_\theta} J_0(k_{\bar{\omega}_\theta} r), \quad (41)$$

where C_1 and C_2 are two arbitrary constants. From Eqs. (29) and (30), the dilatation and rotation become

$$\Delta(r,z,t) = -2C_1(k_z^2 + k_\Delta^2)J_0(k_\Delta r)\exp(i(k_z z + \omega t)), \quad (42)$$

$$2\bar{\omega}_\theta(r,z,t) = 2iC_2(k_z^2 + k_{\bar{\omega}_\theta}^2)J_1(k_{\bar{\omega}_\theta} r)\exp(i(k_z z + \omega t)). \quad (43)$$

The wave can thus be a mixture of the longitudinal and shear components with C_1 and C_2 defining the respective proportions.

Now if k_Δ is imaginary, the Bessel functions $J_0(k_\Delta r)$ and $J_1(k_\Delta r)$ go to infinity as r goes to infinity. This is not physically meaningful. Thus for a phase speed c which satisfies Eq. (16), C_1 must be zero in Eqs. (40) and (41). In view of Eqs. (42) and (43) this is a purely shear wave beam, which travels at the phase speed c.

On the other hand if the phase speed c satisfies Eq. (17), both C_1 and C_2 can have nonzero values. Therefore, infinitely many beam forms exist for which the wave travels at the phase speed c. Each of these beam forms contains both the longitudinal and shear components. This completes the proof. □

It follows that the phase speed depends not only on the mechanical properties of the medium but also on the geometry of the wave. Even for a purely shear wave ($C_1=0$), the phase speed can have any value which is greater than or equal to $\sqrt{\mu/\rho}$. *The shear wave speed,* $\sqrt{\mu/\rho}$, *is not the phase speed of every shear wave in an infinite medium*. It is, however, the phase speed of the uniform plane shear waves. This issue is in addition to the main drawback of the use of the phase speed[63] for tissue characterization; namely, that the phase speed cannot be defined when multiple waves are traveling in different directions, for instance, when there are reflections of the wave from the boundaries.

B. Dependence of the phase speed on the geometry of the medium: Wave guides

In Sec. IV A, the medium was assumed to be infinite in size. This assumption is also made in many of the elastography techniques.[63,79] However, no part of the human body is infinite in size. In this section we present some classical results on the wave guides and study their implications in the field of elastography. The wave guides are infinite in at least one direction and finite in at least another. Therefore, they cannot model the tissue behavior accurately either. However, they can be considered as an intermediate step in moving from the analysis of an infinite medium to a bounded medium. As such the insight gained from studying them is useful in understanding and designing elastography systems.

The wave guide the authors study is an infinitely long cylindrical rod. Choose the cylindrical coordinate system with the axis of the cylinder on the z-axis. As in Sec. IV A, the authors are interested in harmonic waves propagating along the z-axis, i.e., along the axis of the cylinder, which are symmetrical around the z-axis. The same analysis applies and the wave pattern given by Eqs. (27) and (28) with $U(r)$ and $W(r)$ given by Eqs. (40) and (41) satisfies the wave equation with phase speed c, provided that k_z is chosen to satisfy $\omega/k_z = c$. However, in this case the boundary conditions impose restrictions on the permissible values of c.

The boundary of the cylinder is free from stresses. The expressions for stresses in the cylindrical coordinate system are

$$\sigma_{rr} = \lambda\Delta + 2\mu\frac{\partial u_r}{\partial r}, \quad (44)$$

$$\sigma_{r\theta} = \mu\left[\frac{1}{r}\frac{\partial u_r}{\partial \theta} + r\frac{\partial}{\partial r}\left(\frac{u_\theta}{r}\right)\right], \quad (45)$$

$$\sigma_{rz} = \mu\left[\frac{\partial u_r}{\partial z} + \frac{\partial u_z}{\partial r}\right]. \quad (46)$$

By the assumptions $\sigma_{r\theta}$ vanishes everywhere. The boundary conditions require that σ_{rr} and σ_{rz} vanish on the surface of the cylinder. Denote the radius of the cylinder by a. Substitution of the expressions for u_r and u_z from Eqs. (27), (28), (40), and (41) translates the boundary conditions into the following two equations:

$$C_1\left[2\mu\left.\frac{\partial^2}{\partial r^2}\right|_{r=a} J_0(k_\Delta r) - \frac{\lambda\rho\omega^2}{\lambda+2\mu}J_0(k_\Delta a)\right] + 2C_2\mu k_z\left.\frac{\partial}{\partial r}\right|_{r=a} J_1(k_{\bar{\omega}_\theta} r) = 0, \quad (47)$$

$$2C_1 k_z\left.\frac{\partial}{\partial r}\right|_{r=a} J_0(k_\Delta r) + C_2\left(2k_z^2 - \frac{\rho\omega^2}{\mu}\right)J_1(k_{\bar{\omega}_\theta} a) = 0. \quad (48)$$

Note that these equations depend only on the ratio of C_1/C_2. Eliminating this variable between the two, a single equation is obtained for the unknown k_z. This equation is called the *Pochhammer frequency equation*.[62] For a given material ρ, μ, λ, and a are known. So the equation basically describes the relationship between the wave number k_z and angular frequency ω of the wave, or equivalently the relationship between the phase speed c and ω. As it turns out, unlike the infinite medium, not every phase speed is possible for a given ω. The permissible speeds must satisfy the Pochhammer frequency equation. Substitution into either of the equations (47) and (48) determines the ratio C_1/C_2, i.e., the proportions of the longitudinal and transversal waves that should be mixed together to obtain that particular phase speed.

V. SIMULATION RESULTS

A. Shear waves in a tissue mimicking infinite medium

To gain some insight into the theoretical results which were presented in Sec. IV, the authors simulate the shear wave beam forms in an infinite medium for different angular frequencies and phase speeds. For a chosen pair of angular

TABLE I. Mechanical property values used for the simulation.

ρ (kg/m^3)	λ (GPa)	μ (kPa)
1000	2.3	10

frequency ω and phase speed c, the wave number k_z is found from Eq. (37). The authors choose the mechanical properties of the medium to match those found in the soft living tissue such as the breast. These values are listed in Table I. Given the mechanical properties of the medium, k_Δ and k_{ω_θ} are found from Eqs. (35) and (36), respectively. In the next step $U(r)$ and $W(r)$ are found from Eqs. (40) and (41), by setting $C_1=0$ and choosing an arbitrary value for C_2. Note that C_1 is chosen to be zero to obtain a purely shear wave and the value of C_2 only determines the overall amplitude and phase of the wave, but does not change the waveform. If the time evolution of the wave is desired, $U(r)$ and $W(r)$ can be substituted in Eqs. (27) and (28) to obtain $u_r(r,z,t)$ and $u_z(r,z,t)$.

As shear waves can propagate at any phase speed above $\sqrt{\mu/\rho}=\sqrt{10}$ m/s, the choice of the phase speed, c, becomes arbitrary. The authors simulated the beam forms using two chosen values of $c=5$ m/s and $c=12$ m/s. The results are shown in Figs. 1 and 2, respectively. The first and the second column in these figures show the displacement components $u_r((r,\theta,z),t)$ and $u_z((r,\theta,z),t)$, respectively at $t=0$. The three rows correspond to increasing frequencies of the wave 40, 65, and 100 Hz.

Note that because of the uniqueness of the solution to the wave equation, if a hypothetical planar exciter could be built which would create the same harmonic motion as $u_r((r,\theta,z),t)$ and $u_z((r,\theta,z),t)$ over an infinite cross section of the medium, for instance, at $z=0$, the wave forms due to this cylindrically symmetric infinite exciter, in the steady state, would be the same as those depicted in these figures. The figures show how changing the excitation pattern of such an exciter can result in a completely different phase speed, even if the frequency of the excitation is not changed.

B. Waves in a tissue mimicking cylindrical rod

The Pochhammer equation has been studied for metallic rods such as steel beams. To gain some insight into this equation when dealing with the living tissue, the authors consider a simple example here. They choose the mechanical properties of the medium to match those found in the living tissue. These values are listed in Table II. Using these values, the Pochhammer equation was solved numerically using MATLAB for different frequencies. Figure 3 shows the plots of the values of c obtained for each ω. At higher frequencies, the equation has multiple roots (modes), thus the multiples plots in this figure.

Mode 0 has a constant phase speed $c=\sqrt{\mu/\rho}$, i.e., the shear wave speed for all frequencies. However, substitution of the shear wave speed into Eqs. (40) and (41) results in $U(r)=W(r)=0$. Therefore, this mode is a *trivial solution*. As a matter of fact assuming a phase speed equal to the shear wave speed results in vanishing displacements, independent of the material properties. The implication is that *no (axisymmetric) wave can travel along the cylindrical rod with the shear wave speed.*

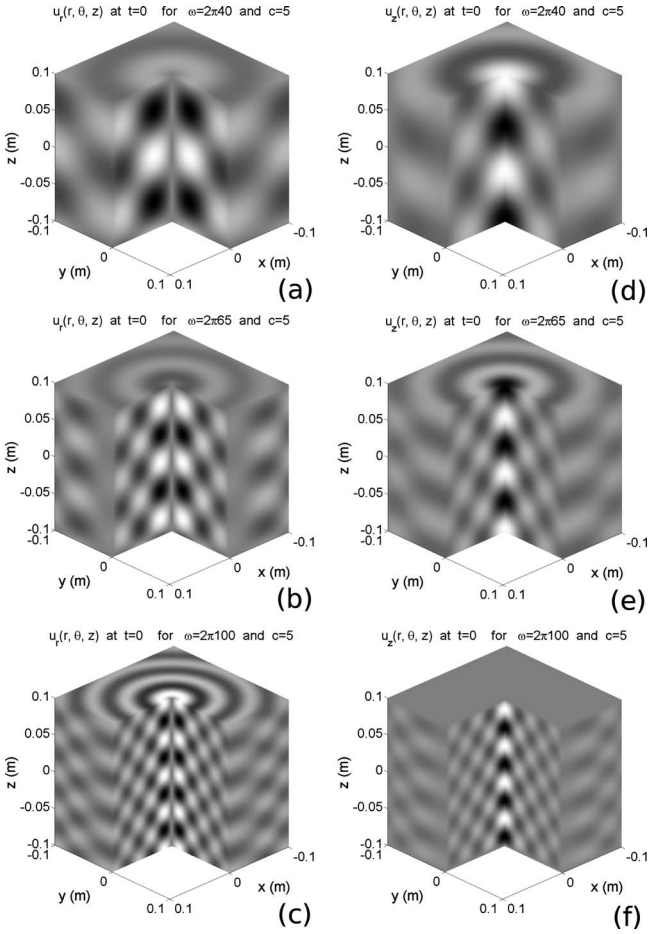


FIG. 1. Infinite medium; shear beam patterns for three different frequencies all sharing the same phase speed of 5 m/s. (a), (b), (c) radial components of the displacement (d), (e), (f) z-component of the displacement at 40 Hz, 65 Hz, and 100 Hz respectively. Note that only a cubic portion of the medium near the axis of symmetry is shown; The medium extends in all directions to infinity.

At low frequencies (below 38 Hz) only mode 1 is present (see Fig. 3). Therefore, the low frequency waves can only travel at a single speed. This is in contrast to the case of the infinite medium studied before, in which the waves, independent of their frequency, could travel at a range of speeds. The low frequency speed is equal to $\sqrt{E/\rho}$ which is the familiar value obtained from the thin rod approximation theory.[64] This also shows the theoretical justification behind the assumptions made in Eq. (8). As the frequency goes higher, other modes start to appear.

At high frequencies (above 38 Hz) multiple waves of the same frequency can propagate simultaneously, each with a different phase speed. For instance, at a frequency of 100 Hz, three phase speeds of 3.0339, 3.719, and 5.630 m/s are possible. The corresponding beam patterns are shown in Fig. 4. In this case it may not be possible to recover a phase speed from studying the displacement patterns inside the material.

To see this more clearly, assume that a harmonic exciter vibrating at a frequency of 100 Hz is placed at infinity on the rod and has caused the following wave pattern:

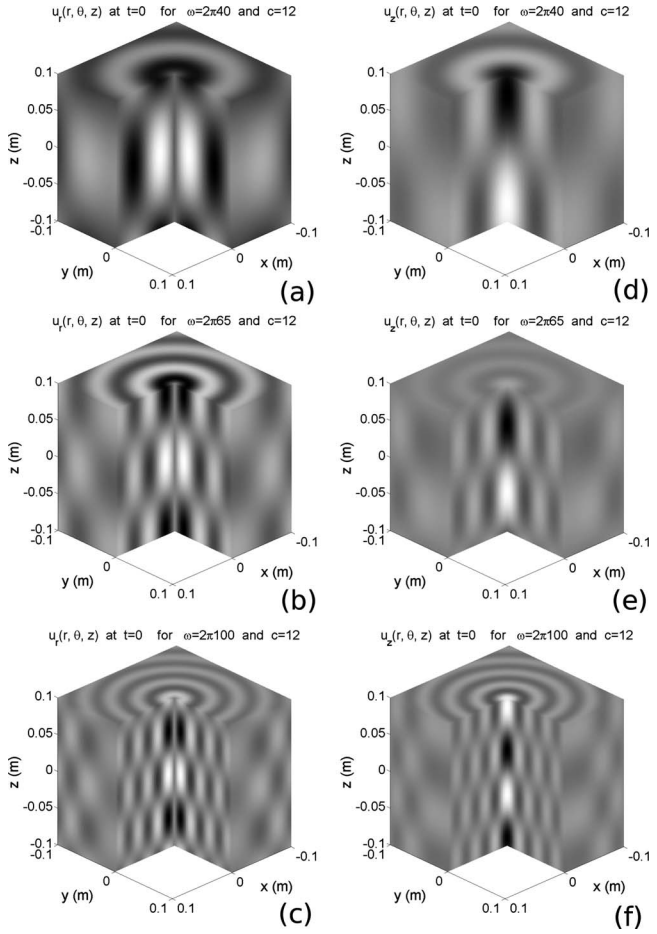


FIG. 2. Infinite medium; shear beam patterns for three different frequencies all sharing the same phase speed of 12 m/s. (a), (b), (c) radial components of the displacement (d), (e), (f) z-component of the displacement at 40 Hz, 65 Hz, and 100 Hz respectively. Note that only a cubic portion of the medium near the axis of symmetry is shown; The medium extends in all directions to infinity.

$$u_r(t) = 3U_{c_1}(r)\exp(i(207.1z + 2\pi 100t)) + 5U_{c_2}(r)\exp(i(168.94z + 2\pi 100t)) + 8U_{c_3}(r)\exp(i(111.60z + 2\pi 100t)), \quad (49)$$

$$u_z(t) = 3W_{c_1}(r)\exp(i(207.1z + 2\pi 100t)) + 5W_{c_2}(r)\exp(i(168.94z + 2\pi 100t)) + 8W_{c_3}(r)\exp(i(111.60z + 2\pi 100t)), \quad (50)$$

where the U_{c_i} and W_{c_i} are the solutions presented in Fig. 4 for the three phase speeds. Because of the linearity of the wave equation, any linear combination of the solutions presented in Fig. 4 is a solution. In particular, the above linear combination with coefficients 3, 5, and 8 caused by the particular exciter used, is a solution. It is not hard to verify by substitution that neither the direct inversion method (14) nor the phase gradient method (15) results in a meaningful mechanical property for the homogeneous cylinder.

TABLE II. Mechanical property values used for the simulation.

ρ (kg/m^3)	λ (GPa)	μ (kPa)	a (m)
1000	2.3	10	0.05

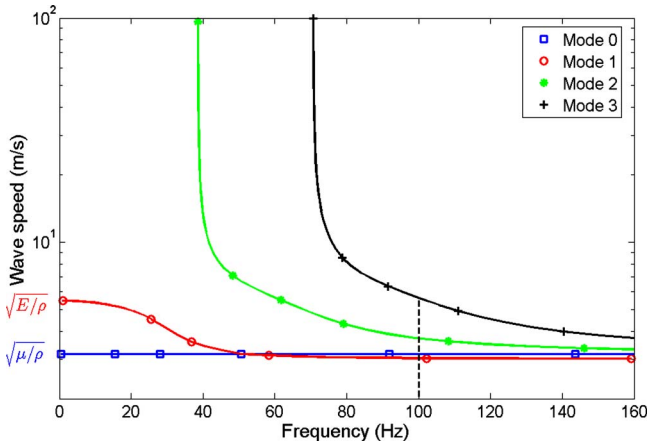


FIG. 3. (Color online) The first four modes of the cylindrical bar

The implication in the context of elastography is that at higher frequencies, multiple modes appear in the measured displacements which makes it impossible to recover a single phase speed and determine the mechanical properties based on that.

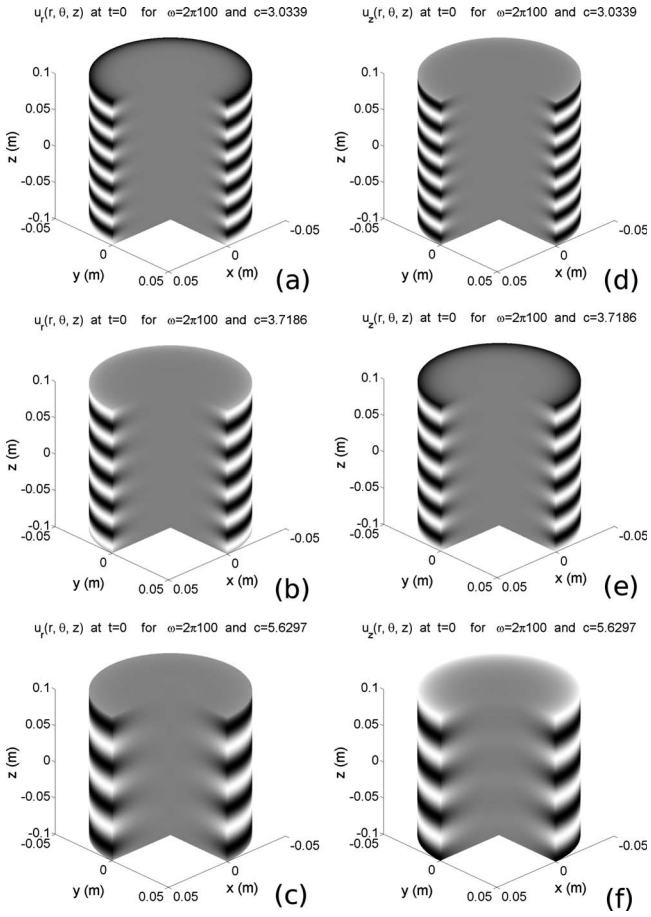


FIG. 4. Permissible beam patterns (modes) in an infinite cylindrical rod for waves having a frequency of 100 Hz. Three (and only three) phase speeds are possible. (a), (b), (c) radial components of the displacement (d), (e), (f) z-component of the displacement for phase speeds of 3.0339 m/s, 3.719 m/s, and 5.630 m/s respectively.

VI. CONCLUSION

The propagation of elastic waves trapped in finite media, such as soft tissue, is a complex phenomenon even without considering the viscous and nonlinear effects. The term wave is generally applied to any vibration pattern inside the tissue. However, the phase velocity cannot be defined for all the cases of the vibration patterns. Usually multiple waves traveling in different directions superimpose to create the vibration pattern. In this case the phase velocity might be well-defined for each wave but not for the resultant of the superposition.

Even in the simple case where a single frequency wave is traveling in a single direction, the geometry of the wave and the medium create multiple permissible phase speeds, and thus make it difficult to recover a meaningful phase speed for the traveling wave. The parameters used in the authors' simulations were chosen to be close to those of the living tissue. The diameter of the cylindrical rod was chosen to be 10 cm which is in the same size range as human organs. Yet multiple phase speeds and modal behavior start to appear at frequencies as low as 38 Hz. This frequency is in the range of frequencies used in many dynamic excitation elastography techniques.

One remedy seems to be the use of the natural decomposition of the vibrations into the longitudinal (dilatational) and transverse (shear) components. Each of these components always satisfies its own wave equation with its own wave speed which is independent of the geometry (note that this is not the phase speed, however). Taking the divergence of the measured displacement field yields the dilatation which satisfies the dilatational wave equation with wave speed $\sqrt{(\lambda+2\mu)/\rho}$,

$$\rho\frac{\partial^2\Delta}{\partial t^2} = (\lambda + 2\mu)\nabla^2\Delta. \tag{51}$$

Since living tissue is incompressible, the dilatations will be very minute and impossible to detect by taking the spatial derivatives of the measured displacements from any of the currently available imaging techniques. Even if accurate measurements were possible, inverting this wave equation would result in the local values of the longitudinal wave speed, $\sqrt{(\lambda+2\mu)/\rho}$, being known. However, the change of this parameter is quite small in the soft tissue (from 1400 m/s for fat to 1540 m/s for muscle). Therefore, the contrast of the formed elastogram would be minimal.

On the other hand, taking the curl of the displacement field results in the rotation fields,

$$\nabla \times (u_x, u_y, u_z)^T = (\bar{\omega}_x, \bar{\omega}_y, \bar{\omega}_z)^T, \tag{52}$$

each of which satisfies a shear wave equation with wave speed $\sqrt{\mu/\rho}$,

$$\rho\frac{\partial^2\bar{\omega}_i}{\partial t^2} = \mu\nabla^2\bar{\omega}_i, \quad i = x,y,z. \tag{53}$$

Since living tissue is incompressible, $E \approx 3\mu$ and inverting these equations results in local information for tissue elasticity E. Moreover since these equations are naturally 1D, it is possible to use the phase speed in their context.

From this discussion it is concluded that *the only theoretically flawless method of imaging tissue elasticity is by taking the curl of the displacement fields*. This requires the measurement of the displacements in all three directions over the volume of interest. In cases where this is not feasible, such as ultrasound elastography, artifacts will always be present. These artifacts are not just due to noisy measurements, poor algorithms, or imperfect setups but due to the theoretical limitations of the systems themselves.

There are some measures which could be taken to minimize the artifacts. The main goal should be to create a single wave propagating in one direction at a single speed.

- The excitation amplitude should be chosen small enough to reduce the reflected wave amplitudes to the level of other measurement noises. The damping present in the tissue helps by reducing the amplitude of the reflected waves from the boundaries of the tissue and bones.
- The frequency of excitation should be kept to a minimum to prevent the appearance of higher modes. However, lower frequencies result in lower damping, and therefore a trade-off exists here.
- The excitation pattern should be chosen so as to excite mainly the lowest mode.

From an engineering point of view, however, many factors beyond the analysis presented here affect the success of an elastography system. Many of the devised elastography techniques are proven to provide clinically valuable images and some of them have even found their way into the market. Quasi-static constant strain, pulsed excitation with external exciters, acoustic-radiation-force-based, and other techniques are each designed to make the displacements created in the clinical setting match the assumptions made about them. This results in the validity of their inversion techniques and their valuable elastograms.

The presence of the artifacts, by itself, does not mean that an elastography system should be dismissed. After all, each of the available medical imaging modalities has its own artifacts, and yet they are very well proven to be useful in diagnosis and treatment.

ACKNOWLEDGMENTS

This research was supported by the Canadian research funding agency NSERC.

[1] J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: A quantitative method for imaging of elasticity of biological tissues," Ultrason. Imaging **13**, 111–134 (1991).
[2] K. Parker, L. Gao, R. Lerner, and S. Levinson, "Techniques for elastic imaging: A review," IEEE Eng. Med. Biol. Mag. **15**, 52–59 (1996).
[3] L. Gao, K. J. Parker, R. M. Lerner, and S. Levinson, "Imaging of the elastic properties of tissue—A review," Ultrasound Med. Biol. **22**, 959–977 (1996).
[4] J. Ophir, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, R. Righetti, and T. Varghese, "Elastographic imaging," Ultrasound Med. Biol. **26**, S23–S29 (2000).
[5] A. Sarvazyan, *Handbook of Elastic Properties of Solids, Liquids and Gases* (Academic, New York, 2001), Vol. **III**.
[6] J. Ophir, S. K. Alam, B. Garra, F. Kallel, E. Konofagou, T. Krouskop, C. Merritt, R. Righetti, R. Souchon, S. Srinivasan, and T. Varghese, "Elastography: Imaging the elastic properties of soft tissues with ultrasound," J. Med. Ultrason. **29**, 155–171 (2002).

[7]J. F. Greenleaf, M. Fatemi, and M. Insana, "Selected methods for imaging elastic properties of biological tissues," Annu. Rev. Biomed. Eng. **5**, 57–78 (2003).
[8]K. J. Parker, L. S. Taylor, S. Gracewski, and D. J. Rubens, "A unified view of imaging the elastic properties of tissue," J. Acoust. Soc. Am. **117**, 2705–2712 (2005).
[9]T. Varghese, J. Ophir, and T. Krouskop, "Nonlinear stress-strain relationships in tissue and their effect on the contrast-to-noise ratio in elastograms," Ultrasound Med. Biol. **26**, 839–851 (2000).
[10]E. E. Konofagou, T. P. Harrigan, J. Ophir, and T. A. Krouskop, "Poroelastography: Imaging the poroelastic properties of tissues," Ultrasound Med. Biol. **27**, 1387–1397 (2001).
[11]S. Chen, M. Fatemi, and J. F. Greenleaf, "Quantifying elasticity and viscosity from measurement of shear wave speed dispersion," J. Acoust. Soc. Am. **115**, 2781–2785 (2004).
[12]J. Bercoff, M. Tanter, M. Muller, and M. FinkJ, "The role of viscosity in the impulse diffraction field of elastic waves induced by the acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 1523–1536 (2004).
[13]R. Righetti, J. Ophir, S. Srinivasan, and T. A. Krouskop, "The feasibility of using elastography for imaging the Poisson's ratio in porous media," Ultrasound Med. Biol. **30**, 215–228 (2004).
[14]R. Sinkus, M. Tanter, T. Xydeas, S. Catheline, J. Bercoff, and M. Fink, "Viscoelastic shear properties of in vivo breast lesions measured by mr elastography," Magn. Reson. Imaging **23**, 159–165 (2005).
[15]R. Righetti, J. Ophir, and T. A. Krouskop, "A method for generating permeability elastograms and Poisson's ratio time-constant elastograms," Ultrasound Med. Biol. **31**, 803–816 (2005).
[16]S. Chen, R. Kinnick, J. F. Greenleaf, and M. Fatemi, "Difference frequency and its harmonic emitted by microbubbles under dual frequency excitation," Ultrasonics **44**, 123–126 (2006).
[17]R. Sinkus, J. Bercoff, M. Tanter, J.-L. Gennisson, C. E. Khoury, V. Servois, A. Tardivon, and M. Fink, "Nonlinear viscoelastic properties of tissue assessed by ultrasound," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **53**, 2009–2018 (2006).
[18]S. Chen, R. R. Kinnick, J. F. Greenleaf, and M. Fatemi, "Harmonic vibroacoustography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **54**, 1346–1351 (2007).
[19]X. Jacob, S. Catheline, J.-L. Gennisson, C. Barriére, D. Royer, and M. Fink, "Nonlinear shear wave interaction in soft solids," J. Acoust. Soc. Am. **122**, 1917–1926 (2007).
[20]A. Thitaikumar, T. A. Krouskop, B. S. Garra, and J. Ophir, "Visualization of bonding at an inclusion boundary using axial-shear strain elastography: A feasibility study," Phys. Med. Biol. **52**, 2615–2633 (2007).
[21]J. Gennisson, M. Rénier, S. Catheline, C. Barriére, J. Bercoff, M. Tanter, and M. Fink, "Acoustoelasticity in soft solids: Assessment of the nonlinear shear modulus with the acoustic radiation force," J. Acoust. Soc. Am. **122**, 3211–3219 (2007).
[22]A. Thitaikumar, L. M. Mobbs, C. M. Kraemer-Chant, B. S. Garra, and J. Ophir, "Breast tumor classification using axial shear strain elastography: A feasibility study," Phys. Med. Biol. **53**, 4809–4823 (2008).
[23]R. Lerner and K. Parker, "Sonoelasticity images, ultrasonic tissue characterization and echographic imaging," in Proceedings of the Seventh European Communities Workshop, Nijmegen, The Netherlands (1987).
[24]R. Lerner, K. Parker, J. Holen, R. Gramiak, and R. Waag, "Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," Acoust. Imaging **16**, 317–327 (1988).
[25]R. Sinkus, J. Lorenzen, D. S. amd, M. Lorenzen, M. Dargatz, and D. Holz, "High-resolution tensor mr elastography for breast tumor detection," Phys. Med. Biol. **45**, 1649–1664 (2000).
[26]R. Souchon, L. Soualmi, M. Bertrand, J.-Y. Chapelon, F. Kallel, and J. Ophir, "Ultrasonic elastography using sector scan imaging and a radial compression," Ultrasonics **40**, 867–871 (2002).
[27]L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging with 2-d transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435 (2002).
[28]L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, and M. Fink, "Shear elasticity probe for soft tissues with 1-d transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 436–446 (2002).
[29]J.-L. Gennisson, S. Catheline, S. Chaffaï, and M. Fink, "Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear wave speeds in muscles," J. Acoust. Soc. Am. **114**, 536–541 (2003).
[30]J. Bercoff, R. Sinkus, M. Tanter, and M. Fink, "Supersonic shear imaging: A new technique for soft tissue elasticity mapping," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 396–409 (2004).
[31]M. W. Urban and J. F. Greenleaf, "Harmonic pulsed excitation and motion detection of a vibrating reflective target," J. Acoust. Soc. Am. **123**, 519–533 (2008).
[32]F. Yeung, S. F. Levinson, D. Fu, and K. J. Parker, "Feature-adaptive motion tracking of ultrasound image sequences using a deformable mesh," IEEE Trans. Med. Imaging **17**, 945–956 (1998).
[33]E. Konofagou, T. Varghese, J. Ophir, and S. Alam, "Power spectral strain estimators in elastography," Ultrasound Med. Biol. **25**, 1115–1129 (1999).
[34]E. E. Konofagou and J. Ophir, "Precision estimation and imaging of normal and shear components of the 3d strain tensor in elastography," Phys. Med. Biol. **45**, 1553–1563 (2000).
[35]T. Varghese, E. Konofagou, J. Ophir, S. Alam, and M. Bilgen, "Direct strain estimation in elastography using spectral cross-correlation," Ultrasound Med. Biol. **26**, 1525–1537 (2000).
[36]M. Fink, L. Sandrin, M. Tanter, S. Catheline, S. Chaffai, J. Bercoff, and J. Gennisson, "Ultra high speed imaging of elasticity," Proc.-IEEE Ultrason. Symp. **2**, 1811–1820 (2002).
[37]J. Bercoff, R. Sinkus, M. Tanter, and M. FinkJ, "3d ultrasound-based dynamic and transient elastography: First in vitro results," Proc.-IEEE Ultrason. Symp. **1**, 28–31 (2004).
[38]K. Hoyt, F. Forsberg, and J. Ophir, "Investigation of parametric spectral estimation techniques for elasticity imaging," Ultrasound Med. Biol. **31**, 1109–1121 (2005).
[39]K. Hoyt, F. Forsberg, and J. Ophir, "Comparison of shift estimation strategies in spectral elastography," Ultrasonics **44**, 99–108 (2006).
[40]K. Hoyt, F. Forsberg, and J. Ophir, "Analysis of a hybrid spectral strain estimation technique in elastography," Phys. Med. Biol. **51**, 197–209 (2006).
[41]J.-L. Gennisson, T. Deffieux, R. Sinkus, P. Annic, M. Pernot, F. Cudeiro, G. Montaldo, M. Tanter, M. Fink, and J. Bercoff, "A 3d elastography system based on the concept of ultrasound-computed tomography for in vivo breast examination," Proc.-IEEE Ultrason. Symp. **1**, 1037–1040 (2006).
[42]A. J. Romano, J. J. Shirron, and J. A. Bucaro, "On the noninvasive determination of material parameters from a knowledge of elastic displacements: Theory and numerical simulation," IEEE Trans. Electromagn. Compat. **45**, 751–759 (1998).
[43]C. Sumi, A. Suzuki, and K. Nakayama, "Estimation of shear modulus distribution in soft tissue from strain distribution," IEEE Trans. Biomed. Eng. **42**, 193–202 (1995).
[44]K. Raghavan and A. E. Yagle, "Forward and inverse problems in elasticity imaging of soft tissues," IEEE Trans. Nucl. Sci. **41**, 1639–1648 (1994).
[45]F. Kallel and M. Bertrand, "Tissue elasticity reconstruction using linear perturbation method," IEEE Trans. Med. Imaging **15**, 299–313 (1996).
[46]M. Doyley, P. Meaney, and J. Bamber, "Evaluation of an iterative reconstruction method for quantitative elastography," Phys. Med. Biol. **45**, 1521–1540 (2000).
[47]P. E. Barbone and J. C. Bamber, "Quantitative elasticity imaging: What can and cannot be inferred from strain images," Phys. Med. Biol. **47**, 2147–2164 (2002).
[48]A. A. Oberai, N. H. Gokhale, and G. R. Feijóo, "Solution of inverse problems in elasticity imaging using the adjoint method," Inverse Probl. **19**, 297–313 (2003).
[49]D. Fu, S. Levinson, S. Gracewski, and K. Parker, "Non-invasive quantitative reconstruction of tissue elasticity using an iterative forward approach," Phys. Med. Biol. **45**, 1495–1509 (2000).
[50]A. Manduca, T. Oliphant, M. Dresner, J. Mahowald, S. Kruse, E. Amromin, J. Felmlee, J. Greenleaf, and R. Ehman, "Magnetic resonance elastography: Non-invasive mapping of tissue elasticity," Med. Image Anal. **5**, 237–2540 (2001).
[51]T. Oliphant, A. Manduca, R. Ehman, and J. Greenleaf, "Complex-valued stiffness reconstruction by for magnetic resonance elastography by algebraic inversion of the differential equation," Magn. Reson. Med. **45**, 299–310 (2001).
[52]S. Catheline, J. Gennisson, G. Delon, M. Fink, R. Sinkus, S. Abouelkaram, and J. Culiolic, "Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach," J. Acoust. Soc. Am. **116**, 3734–3741 (2004).
[53]B. Robert, R. Sinkus, B. Larrat, M. Tanter, and M. Fink, "A new rheological model based on fractional derivatives for biological tissues," Proc.-IEEE Ultrason. Symp. **1**, 1033–1036 (2006).
[54]X. Zhang and J. F. Greenleaf, "Estimation of tissues elasticity with surface

wave speed (l)," J. Acoust. Soc. Am. **122**, 2522–2525 (2007).
[55]A. A. Oberai, N. H. Gokhale, M. M. Doyley, and J. C. Bamber, "Evaluation of the adjoint equation based algorithm for elasticity imaging," Phys. Med. Biol. **49**, 2955–2974 (2004).
[56]M. M. Doyley, S. Srinivasan, S. A. Pendergrass, Z. Wu, and J. Ophir, "Comparative evaluation of strain-based and model-based modulus elastography," Ultrasound Med. Biol. **31**, 787–802 (2005).
[57]M. M. Doyley, S. Srinivasan, E. Dimidenko, N. Soni, and J. Ophir, "Enhancing the performance of model-based elastography by incorporating additional a priori information in the modulus image reconstruction process," Phys. Med. Biol. **51**, 95–112 (2006).
[58]J. Bercoff, S. Chaffai, M. Tanter, L. Sandrin, S. Catheline, M. Fink, J. L. Gennisson, and M. Meunier, "In vivo breast tumor detection using transient elastography," Magn. Reson. Med. **29**, 1387–1396 (2003).
[59]J. Bercoff, M. Muller, M. Tanter, and M. Fink, "Study of viscous and elastic properties of soft tissue using supersonic shear imaging," Proc.-IEEE Ultrason. Symp. **1**, 925–928 (2003).
[60]J. Fromageau, J.-L. Gennisson, C. Schmitt, R. L. Maurice, R. Mongrain, and G. Cloutier, "Estimation of polyvinyl alcohol cryogel mechanical properties with four ultrasound elastography methods and comparison with gold standard testing," IEEE Trans. Electromagn. Compat. **54**, 498–509 (2007).
[61]A. Romano, J. Bucaro, P. Abraham, and S. Dey, "Inversion methods for the detection and localization of inclusions in structures utilizing dynamic surface displacements," Proc. SPIE **5503**, 367–374 (2004).
[62]H. Kolsky, *Stress Waves in Solids* (Dover, New York, 1963).
[63]S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonoelasticity: The acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950 (1999).
[64]A. Baghani, H. Eskandari, T. Salcudean, and R. Rohling, "Measurement of tissue elasticity using longitudinal waves," in Proceedings of the Seventh International Conference on the Ultrasonic Measurement and Imaging of Tissue Elasticity (2008), p. 103.
[65]F. A. Duck, *Physical Properties of Tissue: A Comprehensive Reference Book* (Academic, New York, 1990).
[66]M. M. Burlew, E. L. Madsen, J. A. Zagzebski, R. A. Bahjavic, and S. W. Sum, "A new ultrasound tissue-equivalent material," Radiology **134**, 517–520 (1980).
[67]S. A. Goss, R. L. Johnston, and F. Dunn, "Comprehensive compilation of empirical ultrasonic properties of mammalian tissues," J. Acoust. Soc. Am. **64**, 423–457 (1978).
[68]A. Sarvazyan, O. Rudenko, S. Swanson, J. Fowlkes, and S. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[69]M. Walz, J. Teubner, and M. Georgi, "Elasticity of benign and malignant breast lesions," in *Imaging, Application and Results in Clinical and General Practice*, Eighth International Congress on the Ultrasonic Examination of the Breast (1993), p. 56.
[70]W.-C. Yeh, P.-C. Li, Y.-M. J. amd, Hey-Chi Hsu, P.-L. Kuo, M.-L. Li, P.-M. Yang, and P. H. Lee, "Elastic modulus measurements of human liver and correlation with pathology," Ultrasound Med. Biol. **28**, 467–474 (2002).
[71]T. Krouskop, T. Wheeler, F. Kallel, B. Garra, and T. Hall, "Elastic moduli of breast and prostate tissues under compression," Ultrason. Imaging **20**, 260–274 (1998).
[72]M. Muller, J.-L. Gennisson, T. Deffieux, R. Sinkus, P. Annic, G. Montaldo, M. Tanter, and M. Fink, "Full 3d inversion of the viscoelasticity wave propagation problem for 3d ultrasound elastography in breast cancer diagnosis," Proc.-IEEE Ultrason. Symp. **1**, 672–675 (2007).
[73]L. Huwart, F. Peeters, R. Sinkus, L. Annet, N. Salameh, L. C. ter Beek, Y. Horsmans, and B. E. V. Beers, "Liver fibrosis: Non-invasive assessment with mr elastography," NMR Biomed. **19**, 173–179 (2006).
[74]R. Sinkus, M. Tanter, S. Catheline, J. Lorenzen, C. Kuhl, E. Sondermann, and M. Fink, "Imaging anisotropic and viscous properties of breast tissue by magnetic resonance-elastography," Magn. Reson. Med. **53**, 372–387 (2005).
[75]R. Sinkus, K. Siegmann, T. Xydeas, M. Tanter, C. Claussen, and M. Fink, "MR elastography of breast lesions: Understanding the solid/liquid duality can improve the specificity of contrast-enhanced mr mammography," Magn. Reson. Med. **58**, 1135–1144 (2007).
[76]R. Baierlein, "Representing a vector field: Helmholtz's theorem derived from a Fourier identity," Am. J. Phys. **63**, 180–182 (1995).
[77]F. Rohrlich, "Causality, the Coulomb field, and Newton's law of gravitation," Am. J. Phys. **70**, 411–414 (2002).
[78]K. F. Graff, *Wave Motion in Elastic Solids* (Oxford University Press, New York/Ely House, London, 1975).
[79]L. Sandrin, D. Cassereau, and M. Fink, "The role of the coupling term in transient elastography," J. Acoust. Soc. Am. **115**, 73–83 (2004).

Bottlenose dolphins (*Tursiops truncatus*) moan as low in frequency as baleen whales

Sylvia E. van der Woude
Department of Animal Systematics and Evolution, Institute for Biology, Free University Berlin, 14195 Berlin, Germany and International Laboratory for Dolphin Behaviour Research (ILDBR), c/o Dolphin Reef, Eilat 88100, Israel

(Received 9 January 2009; revised 8 June 2009; accepted 17 June 2009)

Despite a vast number of investigations on the vocal repertoire of bottlenose dolphins, it is still not fully described. This publication reports on a newly discovered tonal low-frequency vocalization in the species at frequencies similar to baleen whale "moans." Dolphin moans are characterized by a slightly modulated fundamental frequency well below 500 Hz that ranges in duration from 0.2 to 8.7 s. Recordings (68 h) were obtained from eight Black Sea bottlenose dolphins residing in an open sea enclosure in Israel. Of 132 unambiguous moans, 49 occurred clearly associated with the release of air from a dolphin's blowhole, which allowed for identifying five moaning individuals. Reasons why this vocalization has not been previously described in any toothed whale are discussed. Moans might not be part of the species' natural repertoire but likewise might have been overlooked due to their inconspicuousness and scarcity, technical limitations, or methodological biases. The function of moaning is unclear; however, the data suggest that moans are signals of anticipating physical satiation provided by humans, i.e., feeding or petting. To further address these questions, verification of moans in other populations and experimental investigation of the properties of moan production and perception are required. [DOI: 10.1121/1.3177272]

PACS number(s): 43.80.Ka [WWA] Pages: 1552–1562

I. INTRODUCTION

Over the past 60 years of acoustic studies on cetaceans, ample evidence demonstrated an increased evolutionary investment of cetaceans in audition compared to other senses (summarized in. e.g., Wartzok and Ketten, 1999). Correspondingly, a great number of cetacean studies has focused on the functions and properties of sound production and perception. However, despite extensive investigation, (1) underwater sound archives still contain recordings of unassigned cetacean vocalizations (Watkins *et al.*, 2004), (2) of some cetacean species no publication on their vocalizations does exist, and (3) even the sound repertoire of the probably most extensively studied cetacean species, the bottlenose dolphin (*Tursiops* spp.), is still not fully described and functionally understood. The objective of this paper is to broaden our knowledge regarding the bottlenose dolphin's vocal repertoire by describing a novel tonal vocalization type, which is remarkably low in frequency for odontocetes.

A. Acoustic repertoire of cetaceans

Cetaceans produce a great variety of vocal and nonvocal sounds (reviewed in, e.g., Herman and Tavolga, 1980; Cranford, 2000). Nonvocal sounds include sounds produced by percussive activity (e.g., "tailslaps"), by respiration (e.g., "coughs"), or digestion. Vocal sounds are generated internally and have been traditionally divided into broad qualitative sound categories, such as "tonal," "pulsive," or "noisy."

Tonal vocalizations are characterized by a continuous sinusoidal waveform and a narrow-band frequency. They are widely attributed to social functions and are classically termed "whistles" and "chirps" in odontocetes and "moans" and "tonal calls" in mysticetes. Although tonal frequencies vary within these two sister groups and some species do not produce tonal sounds (Dawson, 1991), fundamental frequencies are commonly above 1 kHz in odontocetes and below 1 kHz in mysticetes. Pulsive vocalizations are usually broadband and similarly of lower frequencies and longer pulse and interpulse durations in mysticetes. Only for odontocetes are pulsive vocalizations further subdivided functionally and acoustically into "burst pulse sounds" and "echolocation clicks" (Herman and Tavolga, 1980). Functionally, burst pulse sounds appear to serve in communication (Blomqvist and Amundin, 2004), while echolocation clicks are widely accepted to solely serve in detection of the ensonified environment [but for its potential communicational use see, e.g., Dawson (1991) and Herzing (2000)]. Acoustically, burst pulse sounds have, on average, higher pulse repetition rates, lower amplitude, and more energy within the sonic range (Cranford, 2000). While the sonic part of echolocation clicks commonly sounds like a "buzz," burst pulse sequences can take on various harmonic structures that were inconsistently named, primarily based upon how humans perceive them (e.g., as "barks," "brays," "squawks," or "thunks").

Generally, there appears to be an overall difficulty in describing cetacean sounds. This may result from sometimes difficult to define acoustic boundaries (e.g., durations) and hard to quantify acoustic features that increase a signal's complexity and specificity. In tonal sounds, these are features such as continuity (e.g., a tone breaking into units), frequency modulation (spectrographically visible as a "contour"), amplitude modulation (spectrographically visible as "sidebands"), or the shift in energy between harmonics (i.e., the fundamental frequency and its overtones). In addition,

TABLE I. Summary of subjects (f=female; m=male) and number of moans emitted per individual either fully submerged (uw) or with the open blowhole partially or entirely above water (aw). The asterisk (*) denotes subjects observed to produce moans only outside the specified recording times of this study.

Individual (gender)	CIN (m)	SHY (f)	DAN (f)	NAN (f)	JAM (f)	LUN (f)	NIK (f)	SHI (f)
Age (years)	ca. 31	ca. 26	ca. 20	9	5	4	1	⋯
(months)				6–10	6–10	5–9	5–9	7–11
Age class and reproductive stage	Adult	Nursing adult	Pregnant adult	Nursing adult	Subadult	Subadult	Calf of SHY	Calf of NAN
N moans uw (30)	1	⋯	*	14	8	6	1	*
N moans aw (19)	⋯	⋯	⋯	⋯	⋯	19	⋯	⋯

distinct sound types can (1) be sequentially patterned (e.g., in "songs" of mysticetes), (2) be simultaneously produced [indicating different sound production mechanisms (Lilly, 1962; Markov and Ostrovskaya, 1990)], (3) merge into one another [e.g., pulsive into tonal sounds (Murray *et al.*, 1998); burst pulsed sounds into echolocation clicks (Cranford, 2000)], or (4) be fully emitted as intermediates [e.g., pulsed tones (Watkins, 1967b)].

B. Acoustic repertoire of bottlenose dolphins

Most research on dolphin acoustics has focused on the use of whistles and echolocation. Dolphin whistles have received considerable attention in communicational research due to strong indications for their acquisition and employment as referential signals, i.e., as (1) individual-specific "signature whistles" (Caldwell and Caldwell, 1968; most recently summarized in Sayigh *et al.*, 2007; Harley, 2008), (2) alliance-specific whistles (Smolker and Pepper, 1999), or (3) context-specific whistles (Veit, 2002).

Most burst pulse sounds and vocalizations that do not fall into the above mentioned categories have only briefly been described. Examples of the latter are a single pulse often referred to as "jaw clap" (Lilly, 1962) and low-frequency pulses termed "pops" (Connor and Smolker, 1996). In addition, two tonal sound types are distinct from the whistle category due to having an exclusively low fundamental frequency (below 1.3 kHz): the low-frequency narrow-band (LFN) sound (Schultz *et al.*, 1995) and the "gulp" (dos Santos *et al.*, 1995). Both vocalizations are relatively short in duration (<0.4 s), downswept, and produced in series. Gulps are commonly the terminal part of a "bray," a two-part sequence starting with "bray-like" burst pulse sounds (dos Santos *et al.*, 1995). While brays mainly occurred during feeding on salmonids (Janik, 2000), gulps as final elements in sequences of various vocalization types (including brays) mainly occurred during aroused surface behaviors (dos Santos *et al.*, 1995) and agonistic interactions (Veit, 2002). Also LFNs were most often observed during surface behaviors, including mating and chasing (Schultz *et al.*, 1995).

C. Acoustic repertoire of Black Sea bottlenose dolphins

The novel odontocete vocalization type reported on here was produced by captive Black Sea bottlenose dolphins (*T. t. ponticus*) studied in the Red Sea. English-language publications on the acoustics of this geographically and genetically isolated subspecies are scarce. However, the here studied animals confirm that this subspecies produces whistles, echolocation clicks, various burst pulse sounds, jaw claps, and gulps (Veit, 2002), as described for the species. In addition, these animals emit a tonal vocalization type that has hitherto not been described in the genus. Formerly, I termed this vocalization "low-frequency whistle" (LFW) (van der Woude, 2005), but as LFW is already used for odontocete whistles with fundamental frequencies in the low range (below 3 kHz), I renamed the new vocalization type as moan. The term refers to the vocalization's acoustic resemblance to baleen whale tonal calls.

II. METHODS

A. Subjects and study site

Data were obtained from a group of eight Black Sea bottlenose dolphins (Table I) maintained in an open sea enclosure at Dolphin Reef, Eilat, Israel. All subjects were housed in a single, extensive (14,000 m^2), and deep (up to 14 m) "dolphin area" (Fig. 1). For 8 h a day (09:00 a.m. to 17:00 p.m.), visitors could freely enter the adjacent "swimming area" and the three-armed raft inside the dolphin area. Snorkeling and diving inside the dolphin area were restricted to small guided groups limited to certain times and regions, leaving the animals a large retreat area (Brensing *et al.*, 2005). Feedings took place four times a day (usually a few

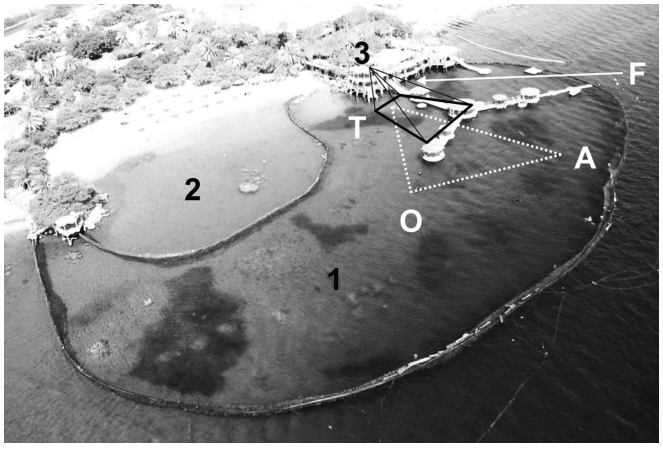


FIG. 1. Dolphin Reef Eilat, Israel, October 2004 (picture taken by Omer Armoza as requested). Location of the dolphin area (1), swimming area for visitors (2), and ILDBR (3). The dashed triangle depicts the approximate positions of three hydrophones (A, O, and T), the white arrow points at the approximate position of an additional hydrophone (F), and the black trapezoid depicts the area shown in the video still frame in Fig. 4(a).

minutes past 10:00 a.m., 12:00 p.m., 14:00 p.m., and 16:00 p.m.) on designated rafts alongside the visitors' raft. Feed amounts increased over the course of the day from about 18% to 38% of the daily ration. Feedings were synchronized and usually short in duration (around 2 min). Due to a non-invasive housing philosophy, the animals were not lured and food rewards were employed only during sporadic medical training and sampling. Interactions with humans were solely voluntary, and any form of harassment and touching specific animals or body parts were strictly prohibited. The area inside the wide-meshed nets (allowing largely unimpeded exchange of fish, sound, etc.) had been naturally preserved, including coral reefs and seagrass meadows. In summary, the dolphins lived as a freely interacting group under conditions reducing restrictions to and interventions in their activities to a minimum.

B. Data collection and recording equipment

Data were collected at the International Laboratory for Dolphin Behaviour Research (ILDBR) situated on a tower approximately 10 m above sea level and overlooking the entire enclosure (Fig. 1). Underwater acoustic recordings were obtained from a triangle of hydrophones (A, O, and T) roughly 31–53 m apart (Fig. 1). A fourth hydrophone (F) was added to the recording system *ad libitum* with an approximately 16–56 m inter-hydrophone distance. Frequency response of hydrophones A, O, and T (High Tech, Inc., HTI-96-min) was 0.002–30 kHz, and that of F (Magrec Ltd., HP/30MT) was 0.2–15 kHz. All hydrophones were fixed in positions allowing for omnidirectional reception and were linked to a PC-soundcard (M-Audio, Delta 1010LT). Received signals were externally amplified (M-Audio, Audio Buddy) and saved as numbered 30-s-wav-files using Avisoft RECORDER (Avisoft Bioacoustics). Although input-device settings (sampling rate: 48 kHz; format: 16 bit; buffer: 0.2 s) and recording levels were unchanged throughout the study, measurements or estimations of source levels were impossible because the hydrophone system was not calibrated.

Most underwater acoustic recordings were supplemented with video and audio recordings taken from the ILDBR balcony. These recordings included verbal comments on the subjects' identities, locations, and behaviors and were saved on VHS (Panasonic, NV-HS900). The VHS stereo audio input was supplied separately, from above water [Hi8 hand camera (SONY, CCD TR805E)] and underwater (channel T forwarded from the PC). Time codes of the video equipment and PC were synchronized daily. Headphones supplied the two channels recorded on VHS and informed the observer about the concurrent underwater and above water sounds simultaneously. Despite equipping the camera with a polarization filter, reflections from intense direct sunlight could not be eliminated, which mostly restricted video recording to the afternoon hours.

C. Data analysis

In total, I analyzed 67.9 h of acoustic recordings and 38.3 h of concurrent visual recordings obtained in 71 days between December 3, 2004 and March 26, 2005. Analysis of acoustic data proceeded in three steps using Avisoft SASLab Pro (Avisoft Bioacoustics). First, the sampling frequency of all recordings was converted to 2 kHz performed with anti-aliasing filtering (accuracy: 128). Second, recordings from all channels were inspected for the presence of moans, both visually (by means of spectrograms) and aurally. At frequencies below 1 kHz, a variety of sounds is present, attributable both to dolphins (e.g., burst pulse sounds and gulps) and to other sources (e.g., boat engines, low-flying airplanes, fish, and scuba divers). Some of these sounds were "moanlike," but most were easily distinguishable from moans. However, a few questionable cases were excluded from further analyses, as were moans that lacked a sufficient signal-to-noise ratio on all channels. Third, for the acoustic characterization of moans, three parameters of the fundamental frequency [duration, minimum frequency (min F_0), and maximum frequency (max F_0)] were measured manually. Note that in the following the terms "fundamental frequency" and "overtone" will be used instead of "harmonic." Correspondingly, the fundamental frequency (F_0) refers to the "first harmonic," the first overtone (H_1) refers to the "second harmonic," etc. (e.g., see Fig. 7 in Lammers *et al.*, 2003).

Moan duration was measured in the amplitude window of the channel at which it was longest. Since moan amplitudes increased and decreased comparatively slowly, the beginning and end of a moan was determined at the appearance and disappearance, respectively, of a sinusoidal wave. Breaks in the contour of some moans were usually shorter than 200 ms and interpreted as parts of the signal with low amplitude. Hence, the duration was measured over the entire sound. Min F_0 and max F_0 were measured from the channel providing the loudest moan and determined by moving the cursor along the moan, both in the spectrogram [parameters: 256-point fast Fourier transform (FFT); Hamming window; 100% frame; 87.5% overlap; 10 Hz frequency resolution; 98.5 ms temporal resolution] and in the power spectrum (with a 7-Hz resolution).

The video recordings were examined for conspicuous behavioral and contextual correlates with moans. In addition, the onset times of moans and feedings were entered into recording log files. A feeding onset was defined as the moment when trainers began calling the animals to their specific feeding positions, i.e., by means of individual underwater acoustic station signals. If only acoustic recordings existed of a feeding, its onset was determined by the first detectable station signal. If the beginning of a feeding could not be detected or had not been recorded, its onset was set on the hour (assuming it to be punctual).

D. Statistical methods

To test whether moans differed between individuals, I ran discriminant function analysis (DFA) (Tabachnick and Fidell, 2001) applied to three subjects (JAM, LUN, and NAN) and three acoustic parameters (duration, min F_0, and max F_0). For this, nine moans per individual were randomly selected and the remaining moans (1–17 per subject) were used for cross-validation. The number of correctly classified as well as the number of correctly cross-classified moans

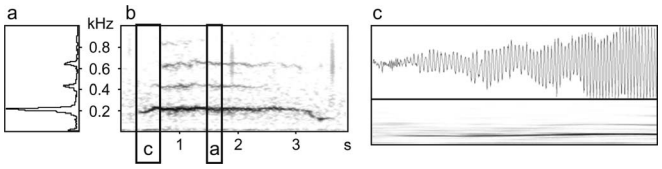


FIG. 2. Spectrogram of a moan (b). The mean power spectrum (a) and the waveform (c) are derived from the respective sections marked in the spectrogram and lack scaling due to unknown amplitudes. Note that the amplitude increases slowly and reaches the limits of the recording equipment while higher order harmonics become spectrographically visible, and vice versa. Spectrogram parameters: 256-point FFT, Hamming window, 100% frame, 87.5% overlap.

were taken as measures of discriminability and it was tested whether these exceeded chance expectations using binomial tests. To avoid undue influence of a specific random selection, the procedure was repeated 10,000 times. This analysis ran with a script written by Mundry for *R* (R Development Core Team 2008) using the function *lda* in the software package MASS (Venables and Ripley, 2002).

To examine temporal correlations between moan production and feedings, only those feedings were considered for analysis, of which recordings from both before (PRE) and after (POST) the feeding onset did exist. The time frame chosen was 30 min PRE and 30 min POST. For the analysis, first the number of moans per hour recorded was determined, separately for PRE and POST and for each individual feeding. Then a Wilcoxon paired-sample signed-ranks test was applied to compare the number of moans PRE and POST. To rule out potential confounding through pooling of the four distinct feedings, the four feedings were also tested separately. Wilcoxon tests were calculated using SPSS 15.0, and indicated *P*-values are exact (Mundry and Fischer, 1998) and two-tailed throughout.

III. RESULTS

A. Sound characteristics

Acoustically, I detected a total of 132 moans with a sufficient signal-to-noise ratio. Moans are narrow-band sounds with peak energy commonly at the fundamental [Fig. 2(a)] and have a sinusoidal waveform without pulsed components [Fig. 2(c)]. Usually, the fundamental frequency was below 400 Hz and ranged between 150 and 240 Hz (Table II). Although moan duration varied greatly between 0.2 and 8.7 s, and their contours were occasionally interrupted, moans were of simple structure. Their fundamental frequency showed little modulation and wavered non-patterned within a narrow frequency band with an average of 85 Hz width (see Figs. 2 and 3 and Table II). In seven moans, peak energy shifted from the fundamental to the first overtone, and in four moans most energy peaked fully at the first overtone. Although amplitude modulation was often evident in the waveform, sidebands were rarely observed. As apparent in Fig. 3, whether or not harmonics or sidebands are spectrographically recognizable presumably depends solely on the received signal strength and signal-to-noise ratio. In Fig. 3, sidebands are noticeable only on channels O and T, whereas harmonics are fully absent on channel A.

TABLE II. Summary of acoustic parameters of moan fundamental frequencies (F_0); bandwidth=difference between min F_0 and max F_0.

N=132	min F_0 (kHz)	max F_0 (kHz)	Bandwidth F_0 (kHz)	Duration F_0 (s)
Range	0.039–0.304	0.085–0.406	0.008–0.235	0.20–8.67
Mean	0.153	0.238	0.085	2.08
±sd	±0.051	±0.067	±0.051	±1.66

B. Bubblestreaming and sender identifications

Of the 132 acoustically recorded moans, 104 occurred during simultaneous video recordings. These revealed that 49 moans clearly coincided with the release of air from a dolphin's blowhole and allowed for an unambiguous identification of five moaning individuals (Table I). Thirty moans were produced by a fully submerged animal expelling a "bubblestream" throughout moaning. These bubblestreams (Fig. 4) clearly differed from any other underwater air emission observed in bottlenose dolphins in both continuity (e.g., compared to eruptive "bubble clouds" occasionally accompanying burst pulse sounds) and thickness, i.e., the release of considerable large amounts of air (compared to delicate "pearl-streams" occasionally accompanying whistles). Twenty-three moans were produced with an open blowhole partially or fully above water, which allowed for sender identification of 19 moans by localizing the corresponding in-air vocalization. Most senders of the remaining 55 moans were unidentified because moans went acoustically undetected during data collection. For cases where no clearly coinciding

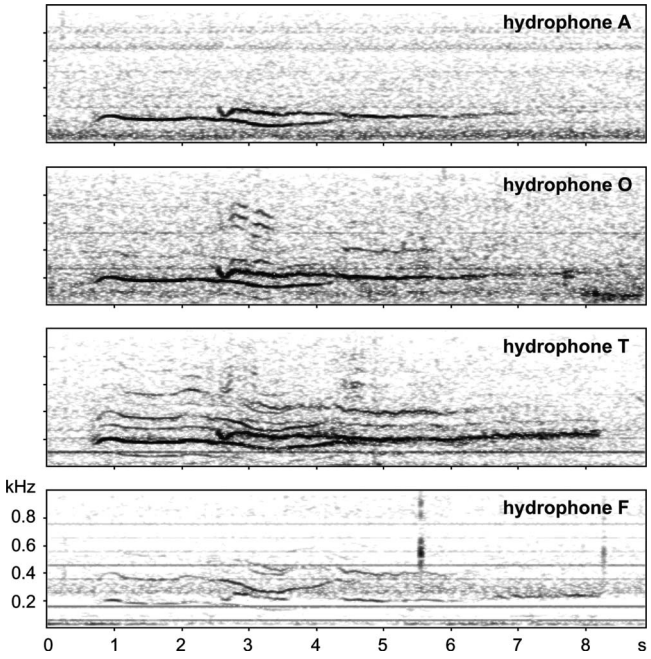


FIG. 3. Spectrograms of two overlapping moans recorded on four hydrophones. See Fig. 4(a) for the corresponding video still frame, revealing the moans to be produced by two subjects (first moan by NAN; second by LUN). The animals increased the distance to hydrophone A, passed by hydrophone O, and approached hydrophones T and F. Correspondingly, on channel A signals became faint and died away, whereas on channel T signals became nearly noisy. Spectrogram parameters: 256-point FFT, Hamming window, 100% frame, 87.5% overlap.

FIG. 4. Video still frames of dolphins bubblestreaming while moaning. Left (a): two dolphins having bubblestreamed simultaneously, but in a slightly staggered manner (emerged NAN followed by submerged LUN). See Fig. 3 for the corresponding spectrograms of concurrent moan production and Fig. 1 for the location of the still frame inside the dolphin enclosure. The respective video is available as supplementary material (video 1, EPAPS). Right (b): NAN blowing bubbles while producing a moan (video filmed on September 4, 2005 by Liron Pinchover as requested).

bubblestream could be detected, moans may have also been produced without or with less conspicuous air emissions. Conversely, all bubblestreams as described above were observed only in association with moan production.

C. Contextual observations

All dolphins observed to moan and bubblestream simultaneously ($N=30$ moans) did so while approaching the area in front of the observation tower. This area was visited by the dolphins regularly before feedings and therefore termed the "vigilance area" at the ILDBR. All approaches occurred in a stereotypical fashion: in a straight line, from a certain direction (ca. only 30° out of potential 180°), at a constant and moderate speed (2–3 m/s), and relatively close to the water surface (Fig. 4). In all cases, moaning was followed by surfacing about a body length later, breathing, and continuing the approach until reaching the entrance to the visitor raft (see supplementary video 1, EPAPS). Hereafter, the animals engaged in other activities (two consecutive moans were not observed) or left the vigilance area, often in order to return in the same manner. Often ($N=23$), a moaner was accompanied by conspecifics swimming parallel or slightly staggered and in the same stereotypical fashion. In two cases a moaning female swam with her calf (NAN with SHI) in infant position, and in two other cases two animals moaned simultaneously, but temporally and spatially staggered [NAN followed by JAM or LUN (as demonstrated in Figs. 3 and 4(a), and supplementary video 1, EPAPS)].

All in-air moans ($N=23$) were also produced in the vigilance area, but by stationary animals floating horizontally or "spyhopping" [positioned vertically with eyes above the water (Shane, 1990)] adjacent to the visitor raft entrance. In-air moans aurally and spectrographically differed from their underwater counterpart by attaining broadband components. They sounded similar to an accelerating car motor overlaid by rapid rattling or buzzing but differed from other in-air vocalizations commonly emitted by captive dolphins and widely known from public shows [e.g., "raspberries" (pulsed sounds; Cranford, 2000) or the "Donald Duck voice" (Lilly, 1962), which is a harmonic vocalization (pulsed sound; Janik and Slater, 1997)].

Throughout, moaning seemed to occur regardless of the presence and activities of nearby conspecifics or humans (divers, snorkelers, or visitors on the raft). Acoustically and visually, no overt response of conspecifics (or humans) to moans could be detected, other than some rare and putative responses of nearby conspecifics to underwater moans: a head turn toward the moaner ($N=1$), two parallel approaching animals turning on their side and swimming bended toward the moaner ($N=1$), and a parallel approaching animal swimming closer to the moaner ($N=5$). In addition, on two days LUN produced an in-air moan during an agonistic interaction while frontally approaching her opponent (JAM or SHY, respectively). In one case, the animals subsequently terminated the agonistic sequence; in the other, they did not.

Prior to this study, the trainers had been unaware of the dolphins producing moans. Afterward, however, they reported (and supplied underwater video footage) of dolphins concurrently moaning and bubblestreaming also in enclosure areas other than the vigilance area and while approaching a trainer in order to subsequently engage in affiliative interactions [see supplementary video 2 (filmed on August 9, 2005 by Omer Armoza, EPAPS)].

D. Individual differences

Acoustic moan parameters tended to differ between individuals. For selected moans, all of the 10,000 random selections revealed a number of correctly classified moans (N_{ccm}) being larger than chance expectation (9), of which 9777 revealed significance ($N_{ccm} \geq 14$, binomial test, $p \leq 0.05$). For cross-classified moans, the result was similar: 9214 random selections revealed N_{ccm} being larger than chance expectation (7.34), of which 3926 revealed significance ($N_{ccm} \geq 12$, binomial test, $p \leq 0.05$). DFA deriving discriminant functions (DFs) for all available moans revealed only the first out of two functions being significant (function 1: $\chi^2=17.1$, df=6, and $p=0.009$; function 2: $\chi^2=0.2$, df=2,

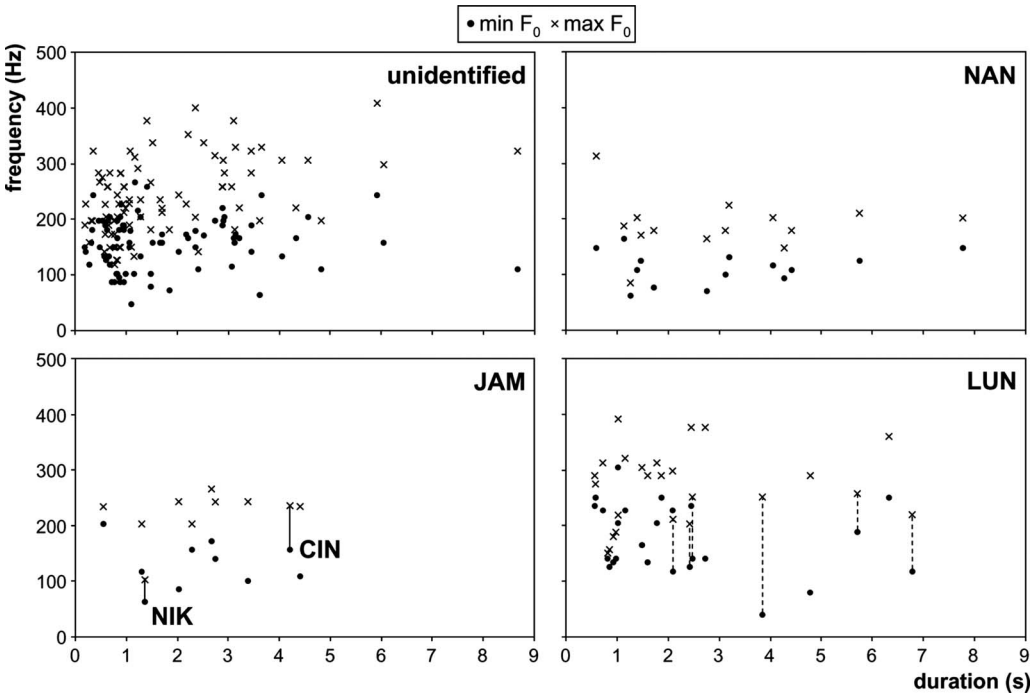


FIG. 5. Acoustic parameters measured for each moan (N=132), separately for each subject (LUN, NAN, and JAM) and for all moans individually unidentified. Measures of the two single moans produced by one animal (NIK and CIN) were added to the chart of subject JAM and are highlighted by a connecting line between the respective minimum and maximum frequencies. The dashed connecting lines in the chart of subject LUN indicate the 6 out of 25 moans that were produced when fully submerged (i.e., the remaining moans were produced with the open blowhole exposed partially or entirely above water).

and p=0.92). The two parameters loading most on the first DF were max F_0 (loading=0.90) and min F_0 (0.76). In fact, NAN seemed to moan lower in frequency than the other subjects (Fig. 5). As indicated in Fig. 5, no significant correlations among sound parameters could be expected (such as the longer the moan, the higher its frequency).

E. Temporal occurrence and feedings

Among 132 moans that were recorded on 30 out of 71 days, three moans were produced on three nights (18:00 p.m. to 05:30 a.m.). Moan production was abundant throughout the day and (a) increased over the course of a day, (b) ceased after the last feeding, and (c) peaked during the half hours before the 14:00 and 16:00 p.m. feedings (Fig. 6). All of the 49 individually identified moans (ID moans) occurred within 48 min preceding a feeding onset (46 within 30 min PRE). Almost all of these moans occurred before the 14:00 p.m. (N=15) and 16:00 p.m. (N=33) feedings. A closer examination revealed that PRE-feeding peaks were generally, and particularly before the two afternoon feedings, significantly higher than expected (see caption of Fig. 6).

IV. DISCUSSION

Moans comprise a readily identifiable vocalization type that is clearly distinct from other bottlenose dolphin tonal vocalization types observed in the studied group and described for the genus so far. Moans have a narrow fundamental frequency range (maximal 0.039–0.406 kHz) far below the numerously reported wide fundamental frequency range of whistles (about 1–20 kHz). Furthermore, moans have a considerably wider duration range (0.2–8.7 s) than gulps (up to 0.18 s; dos Santos *et al.*, 1995; Veit, 2002) and LFN sounds (up to 0.41 s; Schultz *et al.*, 1995). In contrast to gulps and LFN sounds, moans are not downswept and nonrepetitive, and unlike gulps, moans are not combined with

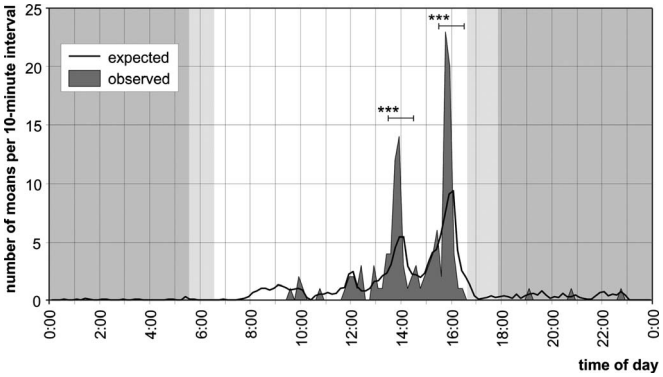


FIG. 6. Expected and observed occurrence of moans (N=132) over the course of a day; for all days combined. The dark gray areas depict the absolute number of moans observed per 10-min interval. The thick solid line depicts the number of moans expected relative to the total amount of analyzed data per 10-min interval. Feedings took place at approximately 10:00 a.m., 12:00 p.m., 14:00 p.m., and 16:00 p.m. [$N_{\text{feedings recorded}}$=5, 10, 26, and 39, respectively; 17 of 80 feedings were not included in analysis due to missing data (i.e., because recordings existed only of either PRE or POST; $N_{\text{feedings analyzed}}$=4, 8, 20, and 31, respectively)]. Moans were significantly more common during the 30 min PRE feeding [Wilcoxon signed-rank test; all feedings pooled: T^+=686, N=38 (25 ties), $P<0.001$; 14:00 p.m. feeding: T^+=104, N=14 (6 ties), $P<0.001$; 16:00 p.m. feeding: T^+=171, N=18 (13 ties), $P<0.001$]. Shaded backgrounds represent night-time on the shortest day (light gray) and longest day (dark gray) during the study (local sunrise/sunset data between December 3, 2004 and March 26, 2005; source: Astronomical Applications Dept., U.S. Naval Observatory, Washington, DC at http://aa.usno.navy.mil/data/docs/RS_OneYear.php).

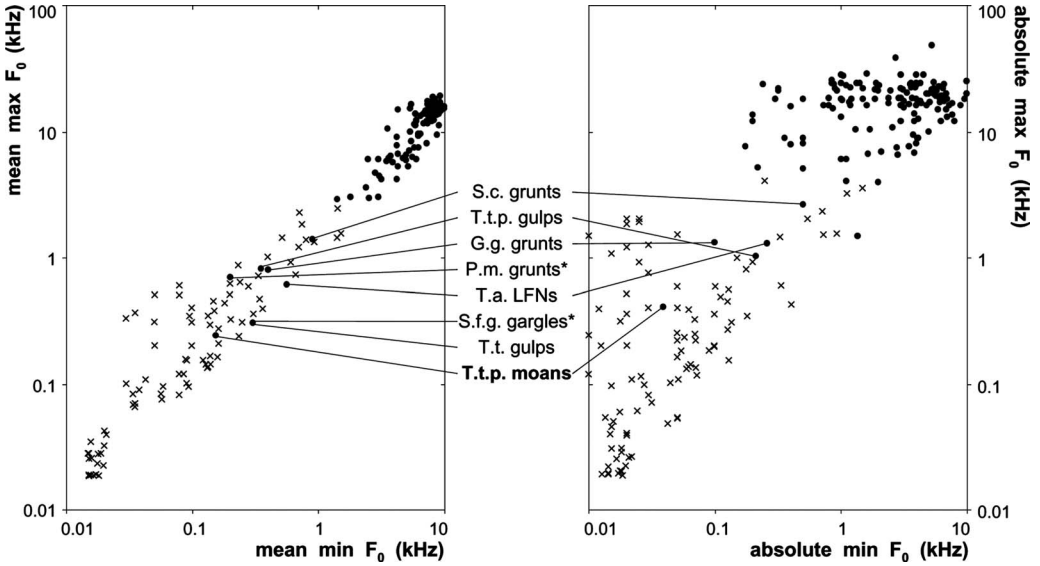


FIG. 7. Mean frequency range (left) and absolute frequency range (right) of the fundamental frequency (F_0) of cetacean tonal vocalization types; data compiled from literature partly reporting on either one pair of parameters (see supplementary table, EPAPS). Solid circles depict data from odontocete species, crosses from mysticete species. Highlighted are odontocete tonal vocalization types with mean max F_0 below 1.4 kHz [$G.g.$ = *Grampus griseus* (grunts described by Corkeron and Van Parijs, 2001), $P.m.$ = *Physeter macrocephalus* (grunts described by Madsen *et al.*, 2003), $S.f.g.$ = *Sotalia fluviatilis guianensis* (gargles described by Monteiro-Filho and Monteiro, 2001), $S.c.$ = *Sousa chinensis* (grunts described by Van Parijs and Corkeron, 2001), $T.a.$ = *Tursiops aduncus* (LFNs=low-frequency narrow-band sounds described by Schultz *et al.*, 1995), $T.t.$ = *Tursiops truncatus* (gulps described by dos Santos *et al.*, 1995), $T.t.p.$ = *Tursiops truncatus ponticus* (gulps described by Veit, 2002; moans described in this study)]. The asterisk (*) denotes vocalization types that were likely but not explicitly stated to be tonal.

other vocalizations. In this study, no other vocalization with similar characteristics or intermediates between moans and other vocalization types could be identified.

To my knowledge, moans have been described neither in bottlenose dolphins nor in any other odontocete before. Both absolute and mean fundamental frequency ranges of moans are by far the lowest reported for odontocete tonal vocalizations, though these measures match those reported for at least half of all mysticete species (Fig. 7). Reasons for odontocete moans going unreported may include: They are (1) unique to the captive study group, (2) specific to Black Sea bottlenose dolphins, or (3) indeed a common odontocete vocalization, but were either not recognized as a distinct type or simply overlooked in past studies. In the following, these possibilities will be discussed.

A. Moans as vocalizations unique to the study group

If moans do not belong to the natural sound repertoire of bottlenose dolphins, the dolphins in Eilat must have newly acquired them either by invention or imitation. Until today, vocal invention has not been clearly demonstrated in bottlenose dolphins since vocal imitation, i.e., the animals' experience with a potential model sound, could not be precluded. Vocal imitation, however, in the form of spontaneous and unreinforced or self-rewarded mimicry of human made sounds has been reported in captive bottlenose dolphins on a few occasions (see a review in Janik and Slater, 1997). Poole *et al.* (2005) claimed to have discovered vocal learning in African elephants (*Loxodonta africana*) by describing a semi-captive elephant presumably imitating remote truck sounds. Likewise, moans may be copies of recreational boat engines. Some specific engine sounds attained remarkable spectrographic and audible similarities to moans owing to a narrow-band concentration of acoustic energy ranging below 300 Hz. Particularly, changes between idling and short drives (i.e., while repositioning boats against the current) caused engine acceleration sounds of variable durations and slowly sloping amplitudes strikingly similar to moans. Also temporally, moans and engine sounds seemed to coincide since the majority of boats frequented the study site before and during the two afternoon feedings (personal observation). I recognized only one other group of potential models for moans: underwater diver vocalizations. They occasionally sounded similar to moans but had a harmonic structure modulated wider in frequency. Since divers were present inside the dolphin area hourly and vocalized rarely, diver vocalizations coincided temporally less with moans than boat engine sounds, but divers could have offered a combined model of sound and concurrent air-release.

However, dolphin imitation of these low-frequency sounds appears fairly speculative for two main reasons. First, in contrast to reports on bottlenose dolphin vocal mimicry (e.g., Reiss and McCowan, 1993), moan production was not noticeably related to or triggered by the assumed models or their sources; i.e., accelerating boats or vocalizing divers. Second, the suggested models appear to be too low in frequency for dolphins to mimic. Dolphin sound imitations were reported to be either of chosen models within (e.g., Miksis *et al.*, 2002) or transposed to (Lilly, 1962; Richards *et al.*, 1984) their preferred vocal frequency range. Qualitative improvement of sound copies over the course of months or years has been described though (e.g., Lilly, 1962; Reiss and McCowan, 1993) and could also apply to the dolphins in Eilat, as they have been producing moans for years (recorded in 2002 through 2009; unpublished data).

B. Moans as specific *T. t. ponticus* vocalizations

To my knowledge, only two publications describing the vocal repertoire of Black Sea bottlenose dolphins are available in English. Aside from whistles and clicks, Burdin *et al.* (1975) proposed a third "main signal class" produced by two captive subjects. This class comprised 11 sounds most frequently emitted by the animals, while their tanks were drained. Due to the lack of a description other than that they lasted up to 2–3 s and "resembled gurgling sounds," it remains unclear whether these were moans. Markov and Ostrovskaya (1990) examined more than 300,000 acoustic signals produced by 20 captive subjects. Despite the fact that these signals were obtained using equipment flat to 50 Hz, no low-frequency vocalizations were documented.

C. Moans as overlooked vocalizations

There are several possible explanations why moans could have been overlooked in past acoustic studies on bottlenose dolphins specifically and on odontocetes in general. First, moans are inconspicuous. In typical spectrograms of dolphin studies (illustrating frequencies up to 15–25 kHz), they are invisible, and even during low ambient sound levels moans became clearly audible only if produced close to a hydrophone. Also, bubblestreams may have been overlooked as concurring sound production and mistaken for occasional release of respiratory gases. The inconspicuousness of moans and bubblestreams is emphasized by two facts: (1) Moans had not been recognized by employees and researchers at Dolphin Reef Eilat before this study. (2) Although I had first detected moans in March 2002, I did not identify their source until December 2004. Second, with an average 0.24 moans per hour and per individual, clearly recognizable moans are relatively rare. Since the number of moans per individual varied greatly (including one subject never having been observed to moan), the number of moans produced in other populations could be even lower. Third, moans can easily go unobserved due to technical limitations, such as interferences or hydrophones with a poor low-frequency response (see hydrophone F in Fig. 3). Fourth, a great number of authors intentionally used equipment with poor low-frequency response, low-frequency cutoffs, or high- and bandpass filters in order to reduce system, flow, ship, or natural ambient noise. Hence, most publications on odontocete sound production that I encountered give little or no attention to frequencies below 500 Hz.

Some authors though have described odontocete whistles with an absolute min F_0 below 500 Hz (see Fig. 7). Potentially, moans were included in these studies but were not recognized as a separate vocalization type. To my knowledge, distinct low-frequency tonal vocalization types have been recognized in only four odontocete species other than bottlenose dolphins (Fig. 7). These findings, along with other newly described vocalization types documented in well-studied cetacean species [e.g., "low-frequency moans" in *Physeter macrocephalus* (Frantzis and Alexiadou, 2008) or "high-frequency whistles" in *Inia geoffrensis* (May-Collado and Wartzok, 2007)], have been mainly published within the past few years. Their topicality may be less attributable to emerging distinct repertoires of newly studied individuals or populations than to recent improvements of recording and analysis methods and of sound type descriptions. Generally, these findings emphasize the need for further studies on cetacean vocal repertoires.

D. Sound source of moans

In this study, the sound production mechanism of moans remains unidentified. Given that all ID moans were emitted during the release of air from the blowhole, it seems likely that moans are produced at or near the blowhole. As described for other cetacean in-air sounds, such as "chuffs" and raspberries of *Stenella frontalis* (Herzing, 2000), "forced blows" and "fart blows" of *Tursiops* sp. (Lusseau, 2006) or "wheezing blows" of *Megaptera novaeangliae* (Watkins, 1967a) moans recorded in air attained a broadband aural character. But in contrast to moans, these in-air sounds were under water either similarly broadband or not audible. As Watkins (1967a) pointed out, the production site of vocalizations recorded simultaneously in air and water is unlikely solely above water, i.e., at the blowhole, since sound produced in one medium barely couples to another. Ridgway *et al.* (1980) reported that the muscles associated with the bottlenose dolphin larynx were active only during the "Bronx cheer or raspberry" involving large amounts of air being forced through the "fluttering lips of the blowhole." From these observations, I conclude that the sound of in-air moans may be a composite of (a) broadband sound generated above water at the constricted blowhole lips and (b) tonal sound generated internally (e.g., laryngeal).

E. Function of moans

Since it is unclear whether in-air moans functionally differ from underwater moans, and whether moans and bubblestreams are a combined display, or whether one of them is a mere by-product, speculations on potential functions of moans must be derived from different perspectives.

Given that nearby conspecifics were sometimes absent during moaning and, if present, appeared neither to be overtly addressed nor to overtly respond, it is probably most parsimonious to assume that moans were emitted uncontrollably or at least without a communicational intention. Moans could simply result from improving bodily conditions, such as air emission and concurrent harrumph or digestive sounds in humans. Ridgway and Carder (1997) reported on a "deaf" bottlenose dolphin that produced no other vocalizations than "low Bronx cheer-like sounds…through a partially open blowhole." This report seems to support the former assumption, although the sounds and their contextual use were not further described, and it was neither reported whether conspecifics responded to them nor whether the "deaf" animal could hear frequencies below 5 kHz.

With regard to acoustic communication, the question whether conspecifics can hear moans arises. Behavioral, electrophysiological, and morphometric studies on bottlenose dolphins demonstrate a low hearing sensitivity to low frequencies. In contrast, several authors reported that bottlenose dolphins responded even at a relatively far range to

low-frequency sounds of prey fish (Gannon *et al.*, 2005) and of conspecifics (e.g., pops; Connor and Smolker, 1996). However, testing bottlenose dolphin sensitivity to tones below 500 Hz has, to my knowledge, only been published thrice. The respective experimental animal responded to frequencies from 500 to 75 Hz (Johnson, 1967), from 300 to 50 Hz (Turl, 1993), and to 300 and 100 Hz (Finneran *et al.*, 2002), but only at high source levels and in close proximity. Although only one study (Finneran *et al.*, 2002) indicated that the relevant cue was acoustic pressure rather than acoustic particle motion, all studies revealed that bottlenose dolphins are capable of perceiving moans.

These results support (1) interpreting occasional turning and swimming to nearby moaners as responses of conspecifics to moans and (2) suggesting that moans serve as signals allowing for detection only by nearby receivers. The latter suggestion seems in line with the lack of clearly noticeable individual differences in moans (i.e., their duration and contour), other than a slight frequency variation (i.e., max F_0 and min F_0) potentially resulting from anatomical differences. In contrast to signature whistles, which are assumed particularly useful in identifying individuals that are out of sight, moans perceived only at close range may not need to be individual-specific. A missing overt response by conspecifics could be explained either by the response being too subtle or by not being requisite for the demonstration of having received a message (such as transmitted in threat signals).

The data suggest that moans are utterances resulting from a state of anticipating certain interactions with humans. On the one hand, most moans appeared to be emitted in a food-anticipatory state since the production of moans increased significantly before feedings and ceased after feeding onsets. In addition, all ID moans were produced in the area where the animals congregate before feedings and were followed by above water exploratory behavior [e.g., spyhopping (Shane, 1990)] adjacent to the specific location where trainers with fish buckets show up when on their way to the feeding rafts. The increased number of moans prior to food delivery may result from generally increased activity during food anticipation, as it has been described in various species (Boulos and Terman, 1980) including bottlenose dolphins (e.g., increased whistle activity; Akiyama and Ohta, 2007). On the other hand, anecdotal evidence of dolphins moaning while approaching a non-feeding human in order to engage in affiliative interactions (i.e., petting) demonstrates that moans were also produced in anticipation of human interactions other than feeding.

With regard to in-air moans, bottlenose dolphin in-air vocalizations were mostly reported in captive subjects, and some of these vocalizations are part of the species' natural underwater vocal repertoire. Increased numbers of vocalizations in air may generally result from various conceivable effects of captivity, such as increased exposure to human in-air sounds and may potentially address above water activities and actors. These assumptions could also hold for in-air moans since most—if not all—of them were produced by one subject being generally far more vocally active in air than other subjects and particularly during feedings (personal observation).

It seems unlikely though that moans were specifically addressing humans since the animals had not always received an immediate human response (previous to this study, humans had not recognized moans). Instead, moaning in bottlenose dolphins may be in line with the suggestion proposed for other captive mammalian species that anticipating a reward may be reinforcing above and beyond the reward itself (de Jonge *et al.*, 2008).

Toward the last feeding of day, the increase in number of pre-feeding moans appears to correlate with the increase in feed amounts. It has been suggested for bottlenose dolphin vocalizations (e.g., whistles or pops) that the intensity of an emotion (i.e., the degree of arousal or threat, respectively) may be encoded in the frequency, duration, amplitude, or number of signals (Caldwell *et al.*, 1990; Connor and Smolker, 1996, respectively). Similarly, I suggest that the number of moans is indicative of the level of anticipation. Due to the overall scarcity of moans, I additionally suggest that simultaneous moan production by two animals swimming in formation is less likely a coincidental event but rather an intended signal amplification displaying shared anticipation, at least by the animal overlapping the other.

F. Function of bubblestreams

Since 63% of moans went visually unobserved, it remains unknown whether all of them were accompanied by bubblestreams. It seems more likely that bubblestreams are characteristic by-products of moans, rather than vice versa, as all observed bubblestreams were associated with moans. Due to bubblestreams being conspicuous displays perceivable by conspecifics (i.e., by means of vision at close range or echolocation at close and far range), an underlying communicative function may be proposed. Previously reported "small" bubblestreams or "whistle trails" accompanying whistles have been suggested to function as indicators of excitement, distress, or location (Fripp, 2005) and to emphasize the acoustic signal or to demarcate its signaler (Pryor, 1990). These suggestions may likewise account for bubblestreams coinciding with moans. In contrast, none of the remaining suggested functions of dolphin underwater air emissions appear to account for bubblestreams since they were clearly not observed to be produced (1) when surprised by an unfamiliar stimulus (Pryor, 1990; McCowan *et al.*, 2000), (2) during agonistic interactions (Shane, 1990; Herzing, 2000), (3) in order to hide from conspecifics or predators (Pryor, 1990), (4) to corral prey fish (Fertl and Wilson, 1997), (5) to support debilitating fish (Simon *et al.*, 2005), or (6) to be manipulated in solitary play behavior (McCowan *et al.*, 2000).

G. Suggestions for future research

To investigate bottlenose dolphin perception of moans, more acoustic properties of moans (such as source levels and harmonic energy content) should be determined and experimentally tested, e.g., by including low-frequency tones in

ongoing audiometric studies. Further, more research is necessary to understand the function of moaning. For instance, the assumption that moans are anticipatory signals could be verified by testing whether moaning also increases in anticipation of an action at a location and time of day where and when moaning is usually not abundant.

But first and foremost, future studies on odontocete acoustics should generally utilize equipment capable of and set to recording frequencies well below 500 Hz. If moans are evidenced in other captive or wild odontocetes, they may become a valuable tool (1) in studies on inter- or intraspecific distinctiveness and (2) in assessing an animal's emotional state. In addition, moans may need to be considered in conservation research (3) as potentially vulnerable to anthropogenic noise and (4) in remote acoustic monitoring projects. Although source levels of dolphin moans are certainly far below those of mysticete moans (see a summary in Wartzok and Ketten, 1999), a dolphin moan emitted close to a hydrophone could be mistaken for a low-level mysticete call.

ACKNOWLEDGMENTS

The study could not have succeeded without the support of numerous people. In particular I thank my supervisors Thomas Bartolomaeus and Roger Mundry, the Dolphin Reef owners, Maya and Roni Zilber, and their staff. Among these, I owe special thanks to Omer Armoza, David Katz, Gil Levy, and Liron Pinchover for hydrophone installation, underwater measurements, and photographic documentation. For assistance in the ILDBR, I particularly thank my colleague Frank Veit. I profited greatly from invaluable aid during planning and installation of my technical system and throughout data analyses by R. Mundry and from immediate software updates by Raimund Specht. Mimi Arandjelovic, R. Mundry, Ulrike Rauthenstrauch, Lori Thomassen, F. Veit, and two anonymous reviewers provided helpful comments on the manuscript. Data collection was financed by the DAAD (German Academic Exchange Service).

Akiyama, J., and Ohta, M. (**2007**). "Increased number of whistles of bottlenose dolphins, *Tursiops truncatus*, arising from interaction with people," J. Vet. Med. Sci. **69**, 165–170.

Blomqvist, C., and Amundin, M. (**2004**). "High-frequency burst-pulse sounds in agonistic/aggressive interactions in bottlenose dolphins, *Tursiops truncatus*," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. F. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 425–431.

Boulos, Z., and Terman, M. (**1980**). "Food availability and daily biological rhythms," Neurosci. Biobehav. Rev. **4**, 119–131.

Brensing, K., Linke, K., Busch, M., Matthes, I., and van der Woude, S. E. (**2005**). "Impact of different groups of swimmers on dolphins in swim-with-the-dolphin programs in two settings," Anthrozoos **18**, 409–429.

Burdin, V. I., Reznik, A. M., Skornyakov, V. M. and Chupakov, A. G. (**1975**). "Communication signals of the Black Sea bottlenose dolphin," Sov. Phys. Acoust. **20**, 314–318.

Caldwell, M. C., and Caldwell, D. K. (**1968**). "Vocalization of naive captive dolphins in small groups," Science **159**, 1121–1123.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego), pp. 199–234.

Connor, R. C., and Smolker, R. A. (**1996**). "'Pop' goes the dolphin: A vocalization male bottlenose dolphins produce during consortship," Behaviour **133**, 643–662.

Corkeron, P. J., and Van Parijs, S. M. (**2001**). "Vocalizations of eastern Australian Risso's dolphins, *Grampus griseus*," Can. J. Zool. **79**, 160–164.

Cranford, T. W. (**2000**). "In search of impulse sound sources in odontocetes," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 109–155.

Dawson, S. M. (**1991**). "Clicks and communication: The behavioural and social contexts of Hector's dolphin vocalisations," Ethology **88**, 265–276.

de Jonge, F. H., Tilly, S.-L., Baars, A. M., and Spruijt, B. M. (**2008**). "On the rewarding nature of appetitive feeding behaviour in pigs (*Sus scrofa*): Do domesticated pigs contrafreeload?" Appl. Anim. Behav. Sci. **114**, 359–372.

dos Santos, M. E., Ferreira, A. J., and Harzen, S. (**1995**). "Rhythmic sound sequences emitted by aroused bottlenose dolphins in the Sado Estuary, Portugal," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein and J. A. Thomas (De Spil, Woerden, The Netherlands), pp. 325–334.

Fertl, D., and Wilson, B. (**1997**). "Bubble use during prey capture by a lone bottlenose dolphin (*Tursiops truncatus*)," Aquat. Mamm. **23**, 113–114.

Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2002**). "Low-frequency acoustic pressure, velocity, and intensity thresholds in a bottlenose dolphin (*Tursiops truncatus*) and white whale (*Delphinapterus leucas*)," J. Acoust. Soc. Am. **111**, 447–456.

Frantzis, A., and Alexiadou, P. (**2008**). "Male sperm whale (*Physeter macrocephalus*) coda production and coda-type usage depend on the presence of conspecifics and the behavioural context," Can. J. Zool. **86**, 62–75.

Fripp, D. (**2005**). "Bubblestream whistles are not representative of a bottlenose dolphin's vocal repertoire," Marine Mammal Sci. **21**, 29–44.

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

Harley, H. E. (**2008**). "Whistle discrimination and categorization by the Atlantic bottlenose dolphin (*Tursiops truncatus)*: A review of the signature whistle framework and a perceptual test," Behav. Processes **77**, 243–268.

Herman, L. M., and Tavolga, W. N. (**1980**). "The communication systems of cetaceans," in *Cetacean Behavior: Mechanisms and Function*, edited by L. M. Herman (Wiley-Interscience, New York), pp. 149–209.

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. (**2000**). "Food-related bray calls in wild bottlenose dolphins (*Tursiops truncatus*)," Proc. R. Soc. London, Ser. B," **267**, 923–927.

Janik, V. M., and Slater, P. B. (**1997**). "Vocal learning in mammals," Adv. Study Behav. **26**, 59–99.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. **2**, pp. 247–260.

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.

Lilly, J. C. (**1962**). "Vocal behavior of the bottlenose dolphin," Proc. Am. Philos. Soc. **106**, 520–529.

Lusseau, D. (**2006**). "Why do dolphins jump? Interpreting the behavioral repertoire of bottlenose dolphins (*Tursiops* sp.) in Doubtful Sound, New Zealand," Behav. Processes **73**, 257–265.

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

Markov, V. I., and Ostrovskaya, V. M. (**1990**). "Organization of communication system in *Tursiops truncatus* Montagu," in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 599–622.

May-Collado, L. J., and Wartzok, D. (**2007**). "The freshwater dolphin *Inia geoffrensis geoffrensis* produces high frequency whistles," J. Acoust. Soc. Am. **121**, 1203–1212.

McCowan, B., Marino, L., Vance, E., Walke, L., and Reiss, D. (**2000**). "Bubble ring play of bottlenose dolphins (*Tursiops truncatus*): Implications for cognition," J. Comp. Psychol. **114**, 98–106.

Miksis, J. L., Tyack, P. L., and Buck, J. R. (**2002**). "Captive dolphins, *Tursiops truncatus*, develop signature whistles that match acoustic features of human-made model sounds," J. Acoust. Soc. Am. **112**, 728–739.

Monteiro-Filho, E. L. A., and Monteiro, K. D. K. A., (**2001**). "Low-frequency sounds emitted by *Sotalia fluviatilis guianensis* (Cetacea: Delphinidae) in an estuarine region in southeastern Brazil," Can. J. Zool. **79**,

59–66.
Mundry, R., and Fischer, J. (**1998**). "Use of statistical programs for nonparametric tests of small samples often leads to incorrect P-values: Examples from animal behaviour," Anim. Behav. **56**, 256–259.
Murray, S. O., Mercado, E., and Roitblat, H. L. (**1998**). "Characterizing the graded structure of false killer whale (*Pseudorca crassidens*) vocalizations," J. Acoust. Soc. Am. **104**, 1679–1688.
Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., and Watwood, S. (**2005**). "Elephants are capable of vocal learning," Nature (London) **434**, 455–456.
Pryor, K. W. (**1990**). "Non-acoustic communication in small cetaceans: Glance, touch, position, gesture, and bubbles," in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 537–544.
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.
Richards, D. G., Wolz, J. P., and Herman, L. M. (**1984**). "Vocal mimicry of computer-generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*," J. Comp. Psychol. **98**, 10–28.
Ridgway, S. H., and Carder, D. A. (**1997**). "Hearing deficits measured in some *Tursiops truncatus*, and discovery of a deaf/mute dolphin," J. Acoust. Soc. Am. **101**, 590–594.
Ridgway, S. H., Carder, D. A., Green, R. F., Gaunt, A. S., Gaunt, S. L. L., and Evans, W. E. (**1980**). "Electromyographic and pressure events in the nasolaryngeal system of dolphins during sound production," in *Animal Sonar Systems*, edited by R.-G. Busnel and J. F. Fish (Plenum, New York), pp. 239–249.
Sayigh, L. S., Esch, H. C., Wells, R. S., and Janik, V. M. (**2007**). "Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*," Anim. Behav. **74**, 1631–1642.
Schultz, K. W., Cato, D. H., Corkeron, P. J., and Bryden, M. M. (**1995**). "Low frequency narrow-band sounds produced by bottlenose dolphins," Marine Mammal Sci. **11**, 503–509.
Shane, S. H. (**1990**). "Behavior and ecology of the bottlenose dolphin at Sanibel Island, Florida," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego), pp. 245–266.
Simon, M., Wahlberg, M., Ugarte, F., and Miller, L. (**2005**). "Acoustic characteristics of underwater tail slaps used by Norwegian and Icelandic killer whales (*Orcinus orca*) to debilitate herring (*Clupea harengus*)," J. Exp. Biol. **208**, 2459–2466.
Smolker, R. A., and Pepper, J. W. (**1999**). "Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops* sp.)," Ethology **105**, 595–617.
Tabachnick, B. G., and Fidell, L. S. (**2001**). *Using Multivariate Statistics*, 4th ed. (Allyn and Bacon, Boston).
Turl, C. W. (**1993**). "Low-frequency sound detection by a bottlenose dolphin," J. Acoust. Soc. Am. **94**, 3006–3008.
van der Woude, S. E. (**2005**). "Tonal low frequency sounds, a novel vocalization in bottlenose dolphins (*Tursiops truncatus*)," in Book of Abstracts of the XX Congress of the International BioAcoustic Council (IBAC), Portorose, Slovenia, September, p. 39(A).
Van Parijs, S. M., and Corkeron, P. J. (**2001**). "Vocalizations and behavior of Pacific humpback dolphins *Sousa chinensis*," Ethology **107**, 701–716.
Veit, F. (**2002**). "Vocal signals of bottlenose dolphins (*Tursiops truncatus*): Structural organization and communicative use," Ph.D. thesis, Free University of Berlin, Berlin, Germany.
Venables, W. N., and Ripley, B. D. (**2002**). *Modern Applied Statistics with S*, 4th ed. (Springer-Verlag, New York).
Wartzok, D., and Ketten, D. R. (**1999**). "Marine mammal sensory systems," in *Biology of Marine Mammals*, edited by J. E. Reynolds, S. A. Rommel (Smithsonian Institution Press, Washington), pp. 117–175.
Watkins, W. A. (**1967a**). "Air-borne sounds of the humpback whale, *Megaptera novaeangliae*," J. Mammal. **48**, 573–578.
Watkins, W. A. (**1967b**). "The harmonic interval: Fact or artifact in spectral analysis of pulse trains," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford), Vol. **2**, pp. 15–43.
Watkins, W. A., Daher, M. A., George, J. E., and Rodriguez, D. (**2004**). "Twelve years of tracking 52-Hz whale calls from a unique source in the North Pacific," Deep-Sea Res., Part I **51**, 1889–1901.
See EPAPS Document No. E-JASMAN-126-013909 for video 1: Concurrent moaning and bubblestreaming by two bottlenose dolphins simultaneously, but temporarily and spatially slightly staggered [NAN followed by LUN (as demonstrated in the corresponding spectrograms in Fig. 3)]; video 2: Concurrent moaning and bubblestreaming by a dolphin (NAN) while approaching a trainer and subsequently engaging in affilitative interactions (filmed on August 9, 2005 by Omer Armoza); table: Cetacean tonal vocalization types and summary of frequency and duration values of fundamental frequencies (F_0). For more information on EPAPS, see http://www.aip.org/pubservs/epaps.html.

Possible occurrence of signature whistles in a population of *Sotalia guianensis* (Cetacea, Delphinidae) living in Sepetiba Bay, Brazil

Luciana Duarte de Figueiredo[a] and Sheila Marino Simão
Departamento de Ciências Ambientais, Laboratório de Bioacústica e Ecologia de Cetáceos, Instituto de Florestas, Universidade Federal Rural do Rio de Janeiro, BR 465 km 7, Seropédica, Rio de Janeiro 23890-000, Brazil

(Received 10 December 2008; revised 13 May 2009; accepted 15 May 2009)

According to the "signature whistle" hypothesis, dolphins emit stereotypic sequential whistles whose function is to transmit the identity and location of the whistling animal. However, it has also been proposed that the information signature may be expressed by distinct acoustical features within a single type of whistle shared by a population of dolphins. In an attempt to detect signature whistles from *Sotalia guianensis* living in Sepetiba Bay, Rio de Janeiro, Brazil, 12 h of vocalizations were recorded. Following analysis of the spectrograms, the whistles were classified according to visual inspection and the contour similarity method. Although the identities of the whistling animals were not established, 202 whistle sequences were selected and classified by visual inspection into 27 different types of potential signature whistles. However, there was a large discrepancy between this classification method and that obtained using the quantitative contour similarity method. The arguments in support of the premise that *S. guianensis* produces signature whistles are discussed and the limitations of the classification systems employed are examined.
 [DOI: 10.1121/1.3158822]

PACS number(s): 43.80.Ka [WWA] Pages: 1563–1569

I. INTRODUCTION

Small cetaceans often communicate through whistles (Hermam and Tavolga, 1980). In fact, a varied repertoire of whistles has been described for the gray dolphin, *Sotalia guianensis* (Monteiro-Filho and Monteiro, 2001; Azevedo and Simão, 2002; Erber and Simão, 2004), lately classified as the marine ecotype of *S. fluviatilis* (Cunha *et al.*, 2005).

The "signature whistle" hypothesis was initially raised by Caldwell and Caldwell (1965) following observations of five captive *Tursiops truncatus* dolphins. According to these authors, each animal tended to emit a unique and distinct type of whistle independent of the circumstances. The individual whistle of each animal was characterized by a specific contour (or stereotype) with a unique pattern of low and high frequency modulations (Tyack, 1986; Caldwell *et al.*, 1990; Sayigh *et al.*, 1990). According to Caldwell *et al.* (1990), the stereotypic contour could be repeated several times within a whistle, the repeated elements being known as "loops" and each of the repeated sequences constituting a signature whistle. The basis of the hypothesis is that the individually distinctive attributes of signature whistles transmits the identity and location of the whistler.

Evidence based on vocal mimicry suggests that the signature whistle contour produced by a dolphin results mainly from vocal learning (Reiss and McCowan, 1993; McCowan and Reiss, 1995; Tyack, 1997; Miksis *et al.*, 2002; Fripp *et al.*, 2005) in which infant dolphins develop their unique signature whistles for use throughout life (Caldwell *et al.*, 1990; Sayigh *et al.*, 1990). Variations of the whistles may serve other functions including the maintenance of group cohesion since, in captivity, the whistles are produced almost exclusively by dolphins isolated from the group (Janik and Slater, 1998).

Most research on signature whistles has been conducted with the species *T. truncatus*, although individual whistles have also been detected in emissions of captive (Caldwell *et al.*, 1973) and free (Herzing, 1996) populations of *Stenella frontalis*, *Lagenorhynchus obliquidens* (Caldwell and Caldwell, 1971), *Delphinus delphis* (Caldwell and Caldwell, 1968), and possibly in *Sousa chinensis* (van Parijs and Corkeron, 2001). The objective of the present paper was to investigate the possible occurrence of signature whistles in a population of *Sotalia guianensis* living freely in their natural habitat, Sepetiba Bay.

II. METHODS

A. Collection of data

Field studies were conducted in Sepetiba Bay, Rio de Janeiro, Brazil (between latitudes 22°54′ and 23°04′S, and longitudes 43°34′ and 44°10′W; Fig. 1) over 22 different days within the period May 1994 to February 1999. Vocalizations of the population of *S. guianensis* in this area were recorded from a stationary boat, located ~20 m from the dolphin group, using a Cetacean Research Technology (Seattle, CA) model C54 hydrophone placed 3 m below the water surface. The output of the hydrophone was connected to a pre-amplifier and to a Sony WM-D3 professional walkman (4 tracks; 2 channels; 1% total harmonic distortion; fre-

[a]Author to whom correspondence should be addressed. Electronic mail: ldsigue@terra.com.br

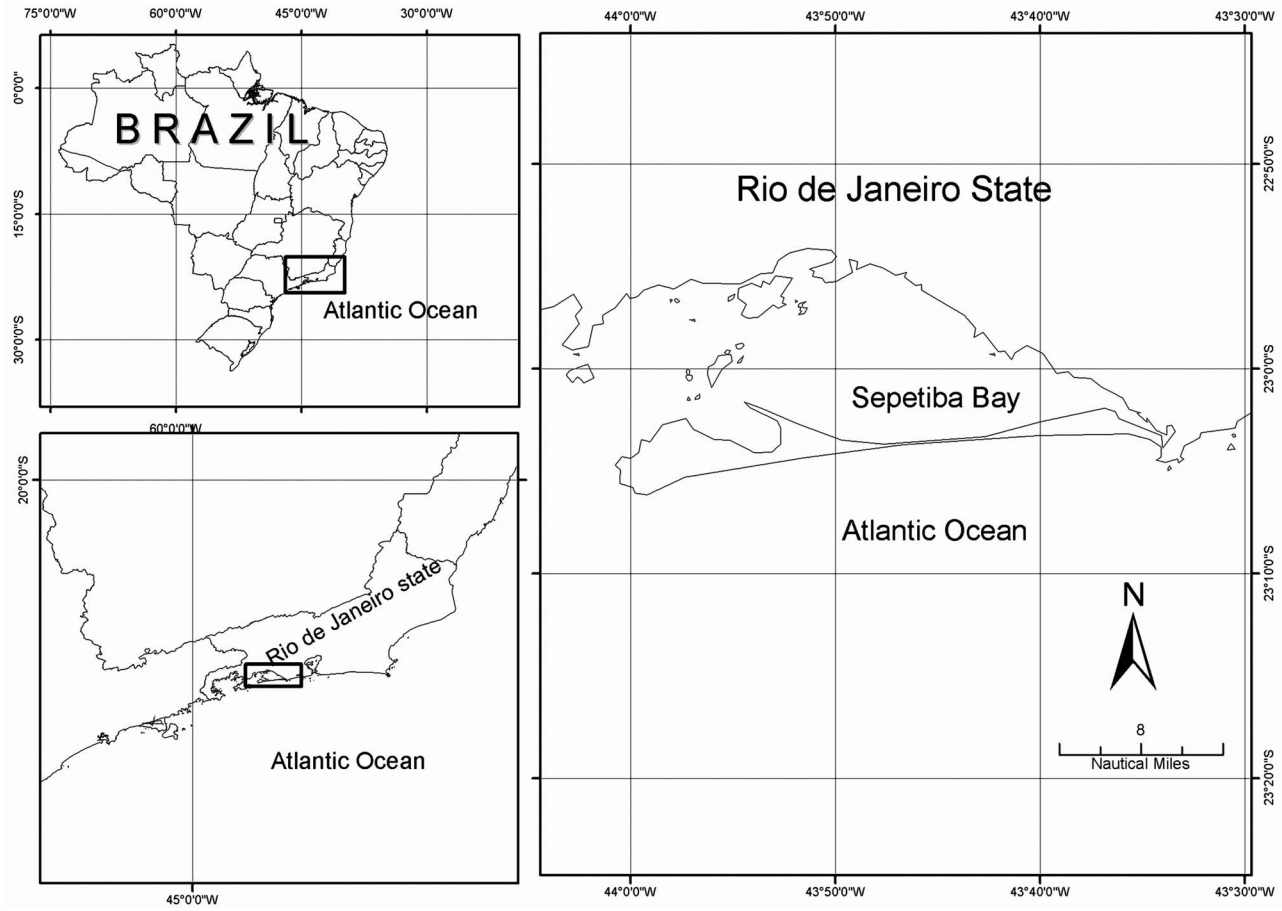


FIG. 1. Geographical location of Sepetiba Bay.

quency response 60–16 000 Hz ± 3 dB) equipped with Sony UCX-S chromium oxide recording tape. A total of 12 h of collective vocalizations, in which the sounds produced by individual animals were not distinguishable, were recorded and analyzed. In parallel with the recordings, information about the climatic conditions, the locations and times of the sightings, and observations concerning the groups of dolphins were noted on field-work forms and on a handheld mini-recorder.

Recorded vocalizations were digitized, oscillograms and spectrograms were generated, and quantitative sound parameters were determined (via Blackman-Harris type fast Fourier transform with a window of 1024 points and 75% overlap) using COOL-EDIT PRO software version 1.2 (Syntrillium Software, Phoenix, AZ). All spectrograms were inspected visually in order to identify stereotypic whistles in sequence. It was impossible to identify the whistler dolphin for each emission. Therefore, a signature whistle was established as a multi-loop one. Once one multi-loop whistle was identified, all one-loop whistles of same contour were considered as part of the sample. Spectrograms corresponding to each of the whistles thus identified were copied, saved in the form of pcx files and labeled with a randomly produced identification number. Each of the selected whistles was analyzed with respect to the number of loops and the presence of a harmonic or lateral band. The following quantitative parameters were extracted from the first loop of each whistle: duration, initial and final frequencies, minimum and maximum frequencies, number of inflexion points, and time intervals between loops.

B. Classification of whistles according to the visual inspection method

The spectrograms of all selected whistles were printed and classified into groups according to the visual perception of the loop contours as determined by two independent researchers. During the classification procedure, the researchers had no access to any other information concerning the whistles. However, since the whistles could not be associated with any particular individual animal, it was not necessary to enlist the participation of a third independent person (unconnected with the project) in order to avoid any tendencies in the classification (McCowan and Reiss, 2001). The two independent classifications obtained were then compared, and whistles with uncertain categorization, together with those that were differently categorized by the two researchers, were disregarded.

C. Classification of whistles according to the contour similarity method

A quantitative method for the classification of whistles emitted by *T. truncatus* has been proposed by McCowan (1995), in which 20 frequency readings were determined for

each whistle contour. The frequencies were obtained by dividing the duration of the loop by 19 and by considering the initial and final frequencies. The frequency measurements were used to produce Pearson's correlation matrix, which was then submitted to principal component analysis to reduce the number of collinear variables. Factors with a value greater than 1.0 were used in the *K*-means cluster analysis to define the different groups of whistles. Finally the cross-validation of whistle types was conducted using stepwise discriminant analysis.

Since the whistle emissions of *S. guianensis* are considerably shorter (mean value=0.102 s; Azevedo and Simão, 2002) than those produced by *T. truncatus* (mean value =0.960 s; Caldwell *et al.*, 1990) the limitations of the hardware and software employed in the present study rendered it impractical to extract 20 frequency readings. For this reason, all the steps were performed taking into consideration 15 frequency readings taken from the first loop of the whistles.

Solutions of the *K*-means cluster analyses were produced in the range $n-2 \leq k \leq n+2$, where n is the number of groups obtained by the visual method and k is the number of groups formed by the cluster analysis. Statistical analyses were performed using STATISTICA for WINDOWS 5.1 software and FITOPAC 1.0 (G. J. Shetherd, UNICAMP, Campinas, São Paulo, Brazil).

In addition the coefficient of frequency modulation (COFM) was calculated for all of the whistles. COFM represents the total magnitude of variation of frequencies in each whistle (McCowan and Reiss, 1995) and was calculated on the basis of the 15 frequency points measured for each whistle according to a modified version of the McCowan and Reiss (1995) equation

$$\text{COFM:} \frac{\sum_{n=1}^{14} |y_{n+1} - y_n|}{10},$$

where Y_n is the frequency (in kilohertz) at the nth frequency point measured.

III. RESULTS

The spectrograms of 346 stereotypic whistles were selected by the visual inspection method from the 12 h of recordings of gray dolphins living freely in Sepetiba Bay, Rio de Janeiro, Brazil, although 144 whistles were later eliminated during the process of visual classification. The 202 remaining whistles had been recorded on 15 of the 22 days of field work (Table I) when the dolphins were engaged in cruising (48%) or random/collective hunting (30%) mainly.

Application of the visual inspection method allowed the division of these 202 whistles into 27 types (Fig. 2); the quantitative parameters are shown in Table II. Seventeen types of whistles (44.5% of the total) were exclusively formed by multi-loops and were recorded on a single day (Table I). Moreover, the time intervals between whistles corresponding to each of these 17 types were very short. These factors all together suggested that each one of these whistle types was emitted by individual dolphins and had the greatest potential to be signature whistles.

TABLE I. Dates when whistles of types 1–27 were recorded.

Type of whistles	Date of recording
1	27/05/1994
2	11/02/1998
3	04/06/1995
4	04/06/1995
5	09/06/1998
6	22/09/1994
7	22/12/1997
8	22/12/1997
9	10/03/98; 21/01/1999
10	21/05/1994; 22/09/1994; 27/08/1995; 02/04/1997; 22/12/1997; 01/03/1998; 10/03/1998; 25/03/1998; 09/08/1998; 21/01/1999
11	04/06/1995; 18/12/1998; 21/01/1999
12	22/12/1997
13	22/12/1997
14	01/03/1998; 09/06/1998; 21/01/1999
15	04/06/1995; 27/08/1995
16	10/03/1998; 26/11/1998
17	01/03/1998
18	22/12/1997
19	27/08/1995; 01/03/1998; 21/10/1999
20	22/09/1994; 27/08/1995; 22/12/1997; 10/03/1998
21	25/03/1998
22	21/01/1999
23	22/12/1997; 10/03/1998; 25/03/1998; 12/05/1998
24	22/12/1997
25	22/12/1997; 10/03/1998; 25/03/1998; 12/05/1998
26	22/09/1994
27	22/12/1997

Analysis of the five solutions of the *K*-means cluster analysis with $25 \leq k \leq 29$ (where k is the number of groups formed by the cluster analysis) showed that classification by the contour similarity method was not in accord with that deriving from the visual inspection method. However, when $k=28$, the two methods exhibited some similarity in relation to 17 possible signature whistles identified by the visual inspection method (Table III), although large discrepancies could still be observed between the two classifications.

IV. DISCUSSION

The fact that 27 distinct types of whistles were detected in the population of *S. guianensis* of Sepetiba Bay underlines the profuse vocal repertoire of this species, as the mean size group was of 30 animals, and confirms the previous findings of the authors' research group (Azevedo and Simão, 2002; Erber and Simão, 2004). In contrast, Monteiro-Filho and Monteiro (2001) reported only four types of whistles in a population of *S. guianensis* living in the estuarine complex of Cananéia (São Paulo, Brazil). This discrepancy may be attributed to the recording equipment used by the earlier authors, as its frequency response had an upper limit of 8000 Hz in comparison with 16 000 Hz for the equipment used in the present study. Since the whistles detected in our studies exhibited average frequencies in the region of 7800 Hz, equipment specification is clearly a factor of considerable importance.

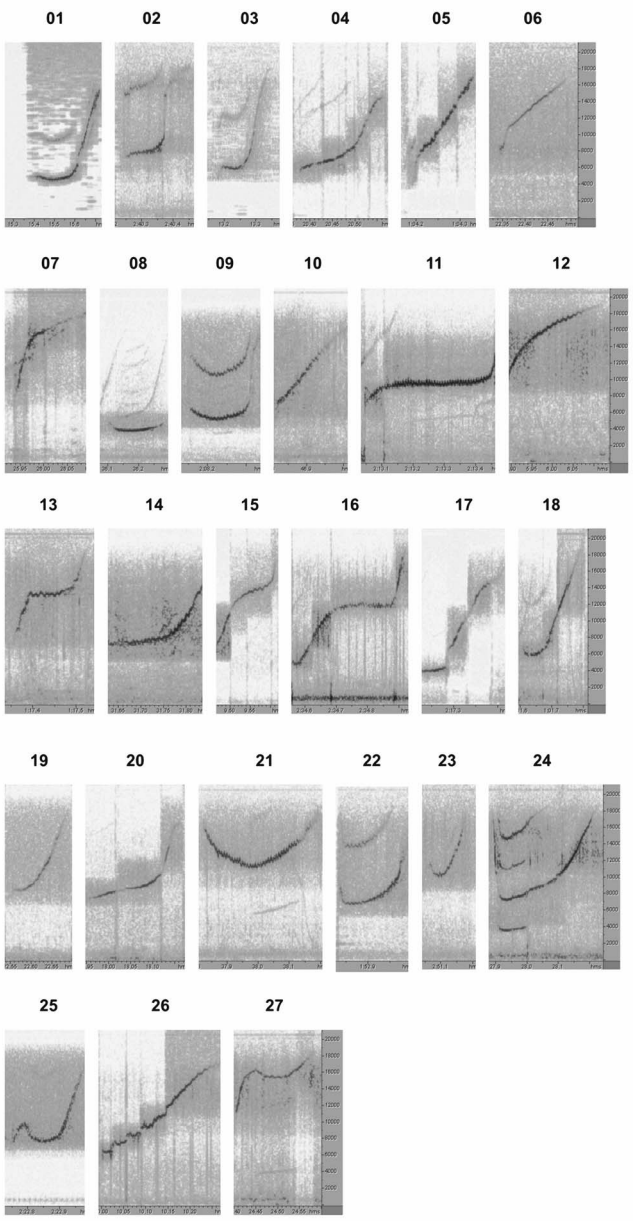


FIG. 2. Spectrograms of the 27 types of potential signature whistles classified by the visual inspection method produced by *Sotalia guianensis* living in Sepetiba Bay, Rio de Janeiro, Brazil.

Most of the previous investigations of signature whistles have involved captive dolphins in which the identity of the vocalizing animal could be readily established (Caldwell *et al.*, 1990; Sayigh *et al.*, 1990). In the few studies relating to free animals, the whistling dolphins were identified either by sub-aquatic filming or after an intense chase of a focal animal (Smolker *et al.*, 1993; Herzing, 1996). In our work it was not possible to establish the identity of the whistling individual since our intention was to interfere as little as possible with the natural behavior of the animals. Furthermore, filming groups of *S. guianensis* would be impracticable since the waters of Sepetiba Bay (Rio de Janeiro, Brazil) are turbid (maximum visibility of ~2.5 m) and gray dolphins are shy animals that tend to avoid proximity to humans.

Hence, among the three characteristics of a signature whistle, namely, stereotypic contour, loop sequences, and individuality (Caldwell *et al.*, 1990), only the first two could be used during the investigation of such sound emissions in the repertoire of this species, resulting in 202 whistle sequences that could constitute signature whistles.

The method of categorizing whistle contours used in the present study was based on that of Janik and Slater (1998) in which visual classification was performed by independent researchers who had no knowledge of the identity and behavior of the whistling animals. By using this strategy, 27 different types of whistles with stereotypic contours, most of which presented multiple loops (74,7%), were identified by the researchers. Although the authors could not confirm that these vocalizations were indeed signature whistles, there are some strong indications supporting this supposition. First, each of the 17 types that presented the highest potential to be signature whistles was detected separately in a single day of recording. Second, these 17 types presented a short time interval between whistles, suggesting that each type of whistle was produced by only one animal. Third, as stated by Tyack (1997), signature whistles allow the individuals to maintain contact with one another, for example, when they are feeding or when one animal is approaching a group. In the present study, 78% of the selected potential signature whistles were recorded in the occasions when the dolphins were cruising or hunting, evidence that is highly consistent with the functions proposed for such type of vocalization. However, Jones and Sayigh (2002) and Cook *et al.* (2004) recorded more signature whistles when *T. truncatus* free-ranging dolphins were socializing. Once only 7.5% of the *Sotalia guianensis* recordings were obtained during socializing events, it is impossible to affirm that the same had not occur for this species in Sepetiba Bay. Finally, our previous studies concerning the gray dolphins of Sepetiba Bay have indicated the existence of an affinity between partners (Simão *et al.*, 2000). According to Beecher (1989), such a relationship between animals is only possible if they are able to recognize distinct individual signs. In the case of dolphins, the hypothesis is that these signs are expressed in the form of signature whistles (Tyack, 1997).

The results of our study are supported by the observations of Ding *et al.* (2001) who, following comparison between whistles produced by different dolphin populations and species, reported the repetition of identical whistles in the repertoire of *S. guianensis*.

The COFM values have demonstrated that the whistles produced by *S. guianensis* are as complex as those of *T. truncatus*, since the COFM value of the former was 0.98 and that of the latter is reported to be 0.88 (McCowan and Reiss, 1995).

The small number of whistles selected (202 sequences in 12 h of recorded vocalization) was expected because the study involved wild animals performing their natural activities. According to Janik *et al.* (1994) and Janik and Slater (1998) signature whistles are produced in all behavioral contexts, but they are more frequent when the animal is isolated from its group, and this is a rare situation in nature. Furthermore, it is believed that when the stress level is reduced and

TABLE II. Quantitative measurements (mean values) of the 27 potential signature whistles classified according to the visual inspection method.

Types	Samples (n)	Duration (ms)	Fi[a] (kHz)	Ff[b] (kHz)	Fmi[c] (kHz)	Fma[d] (kHz)	No. of loops	COFM[e]
1	12	0.286	6.1	15.2	5.6	15.2	1.2	1.09
2	6	0.180	7.2	17.9	7.2	17.9	1.8	1.07
3	7	0.258	4.3	15.4	4.3	15.2	2.1	1.28
4	6	0.768	6.1	14.8	6.1	14.8	1.8	0.87
5	2	0.212	2.3	17.6	2.3	17.6	3.5	1.53
6	2	0.162	6.7	17.4	6.7	17.4	2.0	1.07
7	3	0.127	8.4	17.8	8.4	17.8	2.0	0.93
8	4	0.194	4.5	4.5	4.1	4.5	1.8	0.09
9	6	0.221	8.6	17.5	5.8	17.5	1.3	1.49
10	38	0.161	8.8	16.9	8.8	16.9	2.3	0.83
11	4	0.282	9.7	16.0	9.7	16.0	1.8	0.63
12	6	0.151	9.0	17.5	9.0	17.5	2.0	0.85
13	4	0.122	9.6	17.5	9.6	17.5	1.8	0.80
14	5	0.253	7.8	16.4	7.8	16.4	1.6	0.86
15	6	0.175	5.9	17.2	5.9	17.2	2.0	1.13
16	23	0.219	8.4	17.2	8.4	17.2	2.0	0.85
17	3	0.238	4.3	14.9	4.3	14.9	1.7	1.05
18	2	0.225	6.8	17.6	6.8	17.6	2.0	1.09
19	4	0.132	8.0	16.3	8.0	16.3	2.0	0.83
20	6	0.170	6.6	17.2	6.6	17.2	2.0	1.06
21	4	0.352	16.5	18.0	12.1	18.0	1.8	1.16
22	8	0.203	9.1	15.0	6.9	15.0	1.6	1.01
23	25	0.144	12.2	17.0	10.0	17.0	1.4	0.95
24	6	0.341	14.8	17.3	7.6	17.3	1.5	1.72
25	5	0.243	6.9	16.3	6.9	16.3	1.6	1.50
26	7	0.237	6.8	17.8	6.8	17.8	2.1	1.13
27	7	0.162	11.3	17.6	11.3	17.6	1.4	0.75

[a]Fi=initial frequency.
[b]Ff=final frequency.
[c]Fmi=minimum frequency.
[d]Fma=maximum frequency.
[e]COFM=coefficient of frequency modulation.

the contextual behavior is more diverse, as in animals living freely in the wild, the vocal repertoire is more varied (McCowan, 1995; Smolker and Pepper, 1999). An additional factor that influenced the quantity of whistles selected was the high number of sequences that had to be eliminated either because there was background noise interference or because of their low energy.

The quantitative contour similarity method of McCowan (1995) was employed to compare the similarity between the whistle types selected for two main reasons: First, the technique employs applicable calculus and second the technique allows the categorization of whistles that share similar contours but may present differences with respect to total duration, real frequency, or those that are expanded or compressed with respect to frequency and time. The latter criteria appear to be very important in the classification of whistles, since it has been demonstrated that the signature whistles of *T. truncatus* may vary in duration, frequency, number of loops, etc., while still maintaining the highly distinct pattern of the loop contour (Caldwell *et al.*, 1990).

The results of the authors revealed a large disparity between the visual inspection and the contour similarity methods. Janik (1999), following comparison between different whistle classification methods, reported a similar discrepancy between these two methods. According to this author, the main problem with the contour similarity method appears to be associated with the normalization of the duration and the number of frequency measurements taken for each contour. It is likely that 20 frequency measurements are insufficient to determine the rapid modulations in frequencies that occur in some contours. In order to overcome this problem, modifications were introduced into the method by extracting 60 (McCowan and Reiss, 2001) or 100 (Watwood *et al.*, 2004) frequency points per whistle contour. Following these adjustments, the two classification systems were statistically equivalent and the results produced were more consistent. According to Janik (1999) and Sayigh *et al.* (2007), the contour similarity method is not as reliable as the human observer. Indeed, the authors demonstrated that the whistle contours classified by the visual method were consistently produced by only one isolated animal, thus proving that human observation could recognize meaningful vocalizations, while quantitative techniques could not perceive the differences. Since only 15 frequency measurements could be extracted for each whistle contour in the present study, the

TABLE III. Comparison between the two methods of whistles' classification. Each column represents a whistle type created by McCowan's (1995) method. Each number in columns represents a particular whistle. The visual inspection method's whistle types that are considered more potential signature whistles are bold and, if split, are marked with a similar letter.

Whistle types by McCowan's method																										
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
161	**135**[a]	**204**[b]	**103**[c]	252	**174**	**62**[d]	326	281	**7**	**77**[f]	**17**	63	**218**	283	**60**[d]	149	25	29	**133**[a]	36	3	145	233	21	2	13	48
344	**157**[a]		**110**[c]		**178**	**65**[d]		282	**8**	**78**[f]	**18**	**147**[a]	**219**	286	**195**[b]	235	**49**	171	**137**[a]	**104**[c]	24	**254**	234	**126**	41	37	167
345			**112**[c]			**66**		284	**9**	**79**[f]	21	**148**[a]	**220**		**196**[b]	236	**50**	185	187	**109**[c]	**80**[f]	**255**[f]	237	**139**	43	39	265
346			**113**			**67**[d]		**336**[e]	**11**	**156**[g]	22	**152**[a]	**221**		**198**[b]	244	**52**	250	193	242	**81**[f]	**256**	238	**140**	44	76	266
						68[d]		**338**[e]	**12**	**158**[g]	23		264		**199**[b]		**53**	251	210	258	115		239	**150**	45	93	267
						69[d]			**14**		38				**203**[b]		**54**	269	340	280	**160**[g]		240	180	72	**128**	309
						197[b]			**19**		51				**205**[b]		**55**	**273**	341	**317**[a]	210		241	182	73	**129**	**310**[a]
						201[b]			175		87				**206**[b]		144	**293**			297		243	183	74	303	**311**[a]
									333[e]		88				**208**[b]		168	298					306	184	75	**335**[e]	**312**[a]
									334[e]		89				**209**[b]		304	302						188	114		**313**[a]
											211				**322**[b]		325	305						189	127		**314**[a]
											223						**337**[e]	320						190	131		**315**[a]
											263													192	**136**[f]		**318**[a]
											271														191		331
											272														270		
																									274		
																									287		
																									288		
																									291		
																									292		
																									294		
																									296		

divergence between the two classification methods was already expected.

McCowan and Reiss (2001) investigated whistles produced by captive *T. truncatus* belonging to three different social groups and submitted to two different conditions similar to those predicted in the signature whistle hypothesis (voluntary group separation and temporary forced separation). No stereotypic individual whistles were detected in any of these experiments, with almost all of the individuals producing just one type of whistle characterized by ascending frequency. The authors think that this common whistle could have individual acoustic variations, which might serve as signature information, similar to the situation found with other non-human animals.

In contrast to these authors, a number of recent studies have corroborated the signature whistle hypothesis. For example, Watwood *et al.* (2005) compared signature whistles produced by temporarily restrained *T. truncatus* with whistles produced by free swimming animals. The authors concluded that the types of stereotypic whistles produced by the restrained animals were not a consequence of their confined condition since such whistles were also produced significantly by free interacting animals. Only two of the animals involved in these experiments produced whistle contours similar to those reported by McCowan and Reiss (2001). Sayigh *et al.* (2007) explained the results presented by McCowan and Reiss (2001) as a product of the normalization of whistles' duration that would cause a contour distortion and by the fact that only captive animals were used in the experiment, as Miksis *et al.* (2002) observed that captive-born dolphins can incorporate features of trainers' whistles into their signature whistles.

In the present study, the authors also found a type of whistle (number 10; Fig. 2) similar to that described by McCowan and Reiss (2001) for *T. truncatus*. This type of whistle, which is very simple and presents an ascending contour, has been found in the repertoire of other dolphin species such as *Delphinapterus leucas* (Sjare and Smith, 1986) and *Delphinus delphis* (Moore and Ridgway, 1995) and some other *T. truncatus* populations (Tyack, 1986; Janik *et al.*, 1994; Cook *et al.*, 2004). In this study, whistle number 10 was detected on 11 different days of recorded vocalizations and may be common among the population studied. Whistles of this type that are shared by different species are unlikely to be involved in the transmission of individual information as previously suggested by McCowan and Reiss (2001).

The results presented here agree with signature whistle hypothesis corroborating the suggestion of Sayigh *et al.* (2007) that "individually distinctive signature whistles would appear to be a more promising mechanism for individual recognition than a shared whistle containing subtle signature information, as proposed by McCowan and Reiss (2001)." Further studies involving captive individual dolphins should be conducted in order to confirm the emission of signature whistles by *S. guianensis*. Such a study is feasible since there are two estuarine dolphins in captivity in the Dolphinarium Münster (Munster, Germany; Liebschner *et al.*, 2005).

ACKNOWLEDGMENTS

Part of this research was funded by World Wildlife Foundation (WWF-Brasil). The authors want to thank Romildo Ciniro da Silva for his skill in maneuvering the research boat and for his friendship.

Azevedo, A. F., and Simão, S. M. (**2002**). "Whistles produced by marine tucuxi dolphins (*Sotalia fluviatilis*) in Guanabara Bay, southeastern Brazil," Aquat. Mamm. **28**, 261–266.

Beecher, M. D. (**1989**). "Signalling systems for individual recognition: An information theory approach," Anim. Behav. **38**, 248–261.

Caldwell, M. C., and Caldwell, D. K. (**1965**). "Individualized whistle contours in bottlenosed dolphins (*Tursiops truncatus*)," Nature (London) **207**, 434–435.

Caldwell, M. C., and Caldwell, D. K. (**1968**). "Vocalization of naive captive dolphins in small groups," Science **159**, 1121–1123.

Caldwell, M. C., and Caldwell, D. K. (**1971**). "Statistical evidence for individual signature whistles in Pacific whitesided dolphins, *Lagenorhynchus obliquidens*," Cetology **3**, 1–9.

Caldwell, M. C., Caldwell, D. K., and Miller, J. F. (**1973**). "Statistical evidence for individual signature whistles in the spotted dolphin, *Stenella plagiodon*," Cetology **16**, 1–21.

Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). "Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin," in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego, CA), pp. 199–234.

Cook, M. L. H., Sayigh, L. S., Blum, J. E., and Wells, R. S. (**2004**). "Signature-whistle production in undisturbed free-ranging bottlenose dolphins (*Tursiops truncatus*)," Proc. R. Soc. London, Ser. B **271**, 1043–1049.

Cunha, H. A., da Silva, V. M. F., Lailson-Brito, J., Jr., Santos, M. C. O., Flores, P. A. C., Martin, A. R., Azevedo, A. F., Fragoso, A. B. L., Zanelatto, R. C., and Solé-Cava, A. M. (**2005**). "Riverine and marine ecotypes of *Sotalia dolphins* are different species," Mar. Biol. (Berlin) **148**, 449–457.

Ding, W., Wursing, B., and Leatherwood, S. (**2001**). "Whistles of boto, *Inia geoffresis* and tucuxi, *Sotalia fluviatilis*," J. Acoust. Soc. Am. **109**, 407–411.

Erber, C., and Simão, S. M. (**2004**). "Analyses of whistles produced by the Tucuxi dolphin *Sotalia fluviatilis* from Sepetiba Bay, Brazil," An. Acad. Bras. Cienc. **76**, 381–385.

Fripp, D., Owen, C., Quintana-Rizzo, E., Shapira, A., Buckstaff, K., Jankowski, K., Wells, R. S., and Tyack, P. L. (**2005**). "Bottlenose dolphin (*Tursiops truncatus*) calves appear to model their signature whistles on the signature whistles of community members," Animal Cognition **8**, 17–26.

Hermam, L. M., and Tavolga, W. N. (**1980**). "The communication systems of cetaceans," in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Hermam (Wiley Interscience, New York, NJ), pp. 149–209.

Herzing, D. L. (**1996**). "Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis* and bottlenose dolphins, *Tursiops truncatus*," Aquat. Mamm. **22**, 61–79.

Janik, V. M. (**1999**). "Pitfalls in the categorization of behaviour: A comparison of dolphin whistle classification methods," Anim. Behav. **57**, 133–143.

Janik, V. M., Dehnhardt, G., and Todt, D. (**1994**). "Signature whistle variation in a bottlenosed dolphin, *Tursiops truncatus*," Behav. Ecol. Sociobiol. **35**, 243–248.

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.

Jones, G., and Sayigh, L. (**2002**). "Geographic variation in rates of vocal production of free-ranging bottlenose dolphins," Marine Mammal Sci. **18**, 374–393.

Liebschner, A., Hanke, W., Miersch, L., Dehnhardt, G., and Sauerland, M. (**2005**). "Sensitivity of tucuxi (*Sotalia fluviatilis guianensis*) to airbone sound," J. Acoust. Soc. Am. **117**, 436–441.

McCowan, B. (**1995**). "A new quantitative technique for categorizing whistles using simulated signals whistles from captive bottlenose dolphins *(Delphinidae, Tursiops truncatus*)," Ethology **100**, 177–193.

McCowan, B., and Reiss, D. (**1995**). "Whistle contour development in captive-born infant bottlenose dolphins (*Tursiops truncatus*): Role of learning," J. Comp. Psychol. **109**, 242–260.

McCowan, B., and Reiss, D. (**2001**). "The fallacy of 'signature whistle' in bottlenose dolphins: A comparative perspective of 'signature information' in animal vocalizations," Anim. Behav. **62**, 1151–1162.

Miksis, J. L., Tyack, P. L., and Buck, J. R. (**2002**). "Captive dolphin, *Tursiops truncatus*, develop signature whistles that math acoustic features of human-made model sounds," J. Acoust. Soc. Am. **112**, 728–739.

Monteiro-Filho, E. L. A., and Monteiro, K. D. K. A. (**2001**). "Low-frequency sounds emitted by *Sotalia fluviatilis guianensis* (Cetácea: Delfhinidae) in an estuarine region in southeastern Brazil," Can. J. Zool. **79**, 59–66.

Moore, S. E., and Ridgway, S. H. (**1995**). "Whistles produced by common dolphins from southern California bight," Aquat. Mamm. **21**, 55–63.

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.

Sayigh, L. S., Esch, H. C., Wells, R. S., and Janik, V. M. (**2007**). "Facts about signature whistles of bottlenose dolphins *Tursiops truncatus*," Anim. Behav. **74**, 1631–1642.

Sayigh, L. S., Tyack, P. L., Wells, R. S., and Scott, M. D. (**1990**). "Signature whistles of free-ranging bottlenose dolphins *Tursiops truncatus*: Stability and mother-offspring comparisons," Behav. Ecol. Sociobiol. **26**, 247–260.

Simão, S. M., Pizzorno, J. L. A., Perry, V. N., and Siciliano, S. (**2000**). "Photo-identification method applied to estuarine dolphins, *Sotalia fluviatilis*, (Cetacea, Delphinidae) of Sepetiba Bay," Floresta e Ambiente **7**, 31–39.

Sjare, B. L., and Smith, T. G. (**1986**). "The vocal repertoire of white whales, *Delphinapterus leucas*, summering in Cunningham Inlet, Northwest Territories," Can. J. Zool. **64**, 407–415.

Smolker, R., and Pepper, J. W. (**1999**). "Whistle convergence among allied male bottlenose dolphins (Delphinidae, *Tursiops sp.*)," Ethology **105**, 595–617.

Smolker, R. A., Mann, J., and Smuts, B. B. (**1993**). "Use of signature whistles during separations and reunions by wild bottlenose dolphins mothers and infants," Behav. Ecol. Sociobiol. **33**, 393–402.

Tyack, P. L. (**1986**). "Whistles repertories of two bottlenosed dolphins, *Tursiops truncatus*: Mimicry of signature whistles?," Behav. Ecol. Sociobiol. **18**, 251–257.

Tyack, P. L. (**1997**). "Development and social functions of signature whistles in bottlenose dolphins *Tursiops truncatus*," Bioacoustics **8**, 21–46.

van Parijs, S. M., and Corkeron, P. J. (**2001**). "Evidence for signature whistle production by a Pacific humpback dolphin, *Sousa chinensis*," Marine Mammal Sci. **17**, 944–949.

Watwood, S. L., Owen, E. C. G., Tyack, P. L., and Wells, R. L. (**2005**). "Signature whistle use by temporarily restrained and free-swimming bottlenose dolphins, *Tursiops truncatus*," Anim. Behav. **69**, 1373–1386.

Watwood, S. L., Tyack, P. L., and Wells, R. L. (**2004**). "Whistle sharing in paired male bottlenose dolphins, *Tursiops truncatus*," Behav. Ecol. Sociobiol. **55**, 531–543.

Seasonal changes in the vocal behavior of bowhead whales (*Balaena mysticetus*) in Disko Bay, Western-Greenland

Outi M. Tervo[a)]
Arctic Station, University of Copenhagen, P.O. Box 504, 3953 Qeqertarsuaq, Greenland

Susan E. Parks
Applied Research Laboratory, Pennsylvania State University, P.O. Box 30, State College, Pennsylvania 16804-0030

Lee A. Miller
Institute of Biology, University of Southern Denmark, DK 5230 Odense M, Denmark

(Received 17 December 2008; revised 19 May 2009; accepted 3 June 2009)

Singing behavior has been described from bowhead whales in the Bering Sea during their annual spring migration and from Davis Strait during their spring feeding season. It has been suggested that this spring singing behavior is a remnant of the singing during the winter breeding season, though no winter recordings are available. In this study, the authors describe recordings made during the winter and spring months of bowhead whales in Disko Bay, Western-Greenland. A total of 7091 bowhead whale sounds were analyzed to describe the vocal repertoire, the singing behavior, and the changes in vocal behavior from February to May. The vocal signals could be divided into simple (frequency-modulated) calls (n=483), complex (amplitude-modulated) calls (n=635), and song notes (n=5973). Recordings from the end of February to middle of March were characterized by higher call rates with a greater diversity of call types than recordings made later in the season. This study is the first description of bowhead song from the stock in Western-Greenland during both the winter and spring months, and provides support for the hypothesis that song during the winter months contains more song notes than song from the spring making the winter song more variable. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3158941]

PACS number(s): 43.80.Ka, 43.30.Sf [WWA] Pages: 1570–1580

I. INTRODUCTION

Bowhead whales, *Balaena mysticetus*, are found only in the arctic waters of the Northern Hemisphere (Moore and Reeves, 1993). The global population has traditionally been divided into five stocks: (1) Okhotsk Sea stock, (2) Bering Sea stock, (3) Hudson Bay stock, (4) Davis Strait stock, and (5) Spitsbergen stock. The Hudson Bay stock and Davis Strait stock have been treated separately due to historical whaling records, but in the light of modern satellite tag data and genetic data, it is likely that the two stocks form one population (Heide-Jørgensen *et al.*, 2003, 2006, 2007a). Disko Bay in Western-Greenland is known to be an important aggregation area for the Davis Strait–Hudson Bay bowhead whales (Eschricht and Reinhardt, 1861). Every year from February to May bowhead whales can be seen in a relatively small area in Disko Bay and in increasingly large numbers (Heide-Jørgensen *et al.*, 2006, 2007b). The current estimate of the population size of bowhead whales in Western-Greenland is ~1200 individuals (April–May) (Heide-Jørgensen *et al.*, 2007b). Disko Bay is an aggregation area primarily for adult animals; juvenile animals are rarely seen (Heide-Jørgensen *et al.*, 2007b). Furthermore, in the months of April and May, almost all of the whales (105/130, 81%) occupying the area of Disko Bay are female (Heide-Jørgensen *et al.*, 2007c). Disko Bay is an important feeding area for the bowhead whales in April and May (Laidre *et al.*, 2007), but the behavior, sex ratio, abundance, and movements of the whales earlier in the season are poorly documented. It is possible that more males are present in the area in February and March. Given the late winter presence of bowhead whales in the area, Disko Bay is a potential mating ground for the Davis Strait–Hudson Bay bowhead whale population (Würsig and Clark, 1993; Tyack and Clark, 2000). The few observations of sexual behavior of bowhead whales have been documented in January and February in Disko Bay (Eschricht and Reinhardt, 1861), and the extensive singing behavior of bowhead whales in February and March presented in this study further supports the hypothesis that Disko Bay acts as a mating ground for bowhead whales.

The functions of bowhead whale sounds remain poorly understood despite recording efforts spanning over more than 20 years. Most descriptions of bowhead whale sounds are primarily from recordings of the Bering Sea population near Alaska (Ljungblad *et al.*, 1982, 1984; Clark and Johnson, 1984; Cummings and Holliday, 1987; Blackwell *et al.*, 2007), with a few studies recording sound from the Davis Strait–Hudson Bay population (Richardson and Finley, 1989; Richardson *et al.*, 1995; Tervo, 2006; Stafford *et al.*, 2008). The winter time acoustic behavior for the species is not known (Tyack and Clark, 2000). The migrating bowhead whales off Point Barrow, Alaska have been the subject of several acoustical studies (Ljungblad *et al.*, 1982; Clark and

a)Author to whom correspondence should be addressed. Electronic mail: ote@science.ku.dk

Johnson, 1984; Cummings and Holliday, 1987). The majority of the sounds recorded from the passing bowhead whales were low frequency-modulated (FM) calls (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987) with reported frequency ranges of 25–600 (Ljungblad *et al.*, 1982), 50–300 (Clark and Johnson, 1984), and 25–900 Hz (Cummings and Holliday, 1987). The sounds, or calls, were descending, ascending, constant, or inflecting in frequency (Clark and Johnson, 1984). The duration of all of these calls ranged from short 0.5 s signals to long and melodic 4–5 s tones (Clark and Johnson, 1984). Both song and singing behavior are considered to be advanced forms of vocalization in baleen whales (Clark, 1991). A song is composed of units, phrases, and themes. Units sung in a sequence form phrases, a repetition of a phrase is a theme, and several themes combined create a song (Payne and McVay, 1971). Songs have been recorded from bowhead whales during their spring migration in April–May off Point Barrow when the whales return from their breeding grounds and swim toward their feeding areas (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987) and in Disko Bay, Western-Greenland, during February through May (Tervo, 2006; Stafford *et al.*, 2008). Songs have been documented to change within and between seasons (Clark and Johnson, 1984; Würsig and Clark, 1993; Tervo *et al.*, 2007). It has been suggested that song recorded during the spring migration is a remnant from the winter breeding season and therefore may not represent the entire richness and complexity of the signals that are potentially present earlier in the winter when the bowhead whales are presumed to mate (Würsig and Clark, 1993; Tyack and Clark, 2000).

Song and singing behavior likely have significance in mate choice and sexual selection (Tyack, 1981; Tyack and Clark, 2000; Clark *et al.*, 2003). Four species of baleen whales produce songs: (1) the humpback whale (Payne and McVay, 1971) (*Megaptera novaeangliae*), (2) the blue whale (Cummings and Thompson, 1971) (*Balaenoptera musculus*), (3) the fin whale (Watkins *et al.*, 1987) (*Balaenoptera physalus*), and (4) the bowhead whale (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987). The best studied species is the humpback whale, which produces long and elaborate songs while in their winter grounds (Payne and McVay, 1971; Winn and Winn, 1978) where they breed and calve (Baker and Herman, 1984) and in late spring on their foraging grounds (Clark and Clapham, 2004). The singing behavior has been suggested to have significance in male breeding success (Tyack, 1981; Tyack and Clark, 2000) as only male humpback whales have been documented to sing (Glockner, 1983; Baker and Herman, 1984; Baker *et al.*, 1991). Humpback whale song has been suggested to function as an advertisement display (Tyack and Clark, 2000) and the same could be the case for bowhead whale song (Clark *et al.*, 2003) as well as for fin whale and blue whale song where only males have been observed to sing (McDonald *et al.*, 2001; Croll *et al.*, 2002; Oleson *et al.*, 2007). However, in right whales, a closely related species to bowhead, song has not been described (Clark, 1982; Parks and Tyack, 2005) and females produce the majority of sounds in social surface active groups (SAGs) (Parks and Tyack, 2005). The lack of song in right whales and the prevalence of female vocal behavior in social and sexual contexts make it unclear as to the sex of singing bowheads.

The goal of this study is to describe the changes in the singing behavior of the Davis Strait population of bowhead whales in Disko Bay from winter to spring. Disko Bay is an aggregation area for bowhead whales from January to May providing a unique opportunity to study the variability of vocal behavior of the same population throughout the winter and spring seasons. Understanding the singing behavior during winter is important since bowhead whales are presumed to mate at this time and the changes that occur in the singing behavior during a season can help to pinpoint the timing of sexual behavior of this population. The mating period is one of the most important time periods in the life cycle of a species and more information is needed to ensure that a potential key habitat such as Disko Bay remains accessible for the Davis Strait/Hudson Bay population of bowhead whales during this period.

II. METHODS

A. Collection of acoustic recordings

The study area was located in Disko Bay, Western-Greenland about 69°N and 54°W (Fig. 1). Acoustic recordings of bowhead whale vocalizations were made in Disko Bay offshore from Qeqertarsuaq between 25 February 2005 and 10 May 2005. A total of 890 min of recordings were made over 11 days during the 4 month study period (Table I).

The recordings were made using two hydrophones that were lowered into the water from each side of a dinghy or from R/V Porsild (a 49.5 ft, ice-strengthened, steel vessel) to a depth of ~8 m. One hydrophone (HELWEG, custom built) had a built-in 20 dB amplification and a flat frequency response to 50 kHz (±3 dB). The second hydrophone was a HS 150 (Sonar Research and Development Ltd., Beverley, UK) with a flat frequency response to 150 kHz, which was connected to an Etec amplifier (1 Hz–1 MHz) (Etec, Frederiksværk, Denmark) with high pass filter set at 10 Hz and 26 dB gain. The signals were recorded using a SONY DAT TDC-D8 tape recorder with a sampling frequency of 44.1 kHz. The DAT recorder was the frequency limiting instrument; thus the flat frequency response of the entire recording system was from 20 Hz to 22 kHz.

Simultaneous visual observations of the whales were made whenever possible. The number of individuals and their geographic position (latitude/longitude) were noted. All the recordings were made in the presence of bowhead whales during daylight hours between 8 a.m. and 5 p.m. in February and March and between 8 a.m. and 9 p.m. in April and May.

B. Data analysis

The audio data from the DAT tapes were digitized into standard wave files with BATSOUND software (Petterson Elektronik, Uppsala, Sweden). The acoustic data were analyzed using RAVEN 1.2.1 (Cornell Laboratory of Ornithology, Ithaca, NY) (Hann window, fast Fourier transform (FFT) size 512, with 50% overlap). The recorded signals were divided into the three main categories described by Clark (1991) using a

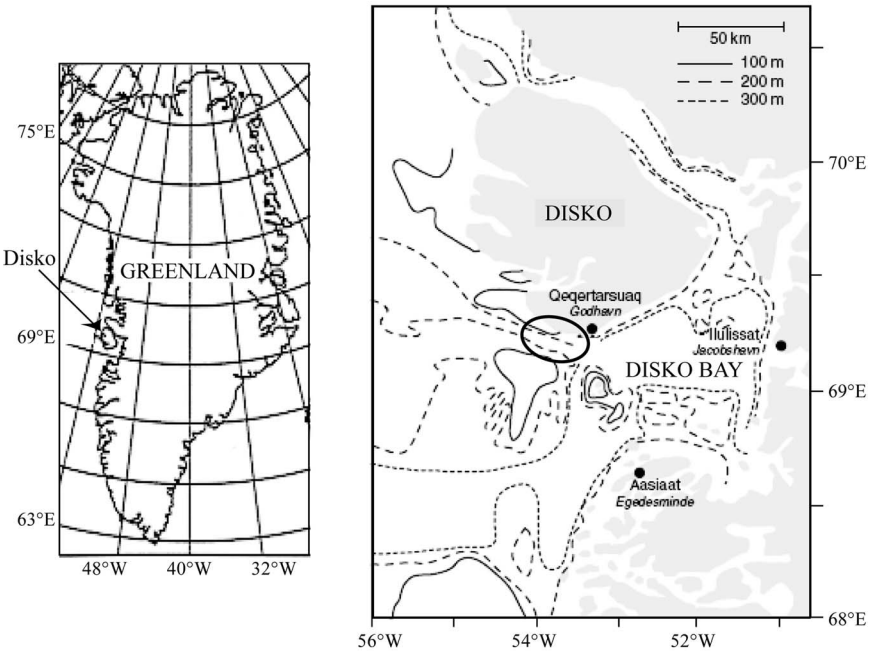


FIG. 1. Map of Greenland showing the location of Disko Bay (courtesy of Torkel Gissel Nielsen). The oval marks the general area where the recordings were made.

combination of spectral and audio qualities: simple FM calls, complex amplitude-modulated (AM) calls, and song notes. Eight variables were measured from simple FM calls and from song notes, including duration, maximum frequency, minimum frequency, frequency range, start frequency, end frequency, number of inflection points, and number of modulation points (Fig. 2). Only four of these parameters were measured for complex AM calls. These were duration, maximum frequency, minimum frequency, and frequency range. Duration of the signal was measured in seconds and was determined from the spectrogram or in some cases from the oscillogram. Maximum and minimum frequencies (Hz) were the highest and lowest frequency points in the signals and were measured from the spectrogram. Frequency range (Hz) refers to the difference between maximum and minimum frequencies of a signal. Start frequency (Hz) refers to the frequency at the start of the signal and end frequency (Hz) to that at the end of the signal. An inflection point refers to a point in the signal where the frequency contour changed from a positive slope to a negative slope or vice versa. A modulation point is a point in the signal showing a smaller degree of frequency modulation that was not strong enough to change the general direction of the signal in frequency and time. A change from a relative constant frequency to a positive or negative frequency slope was considered a modulation point (Fig. 2). Statistical analyses were done using S-PLUS 2000 (MathSoft. Inc., Seattle, WA) and SPSS 12.0.1 (SPSS Inc., Chicago, IL). A Hamming window and a FFT size of 1024 with 75% overlap were used to create the figures of the spectrograms in order to provide good resolution.

III. RESULTS

The bowhead whale was the only baleen whale species present in Disko Bay during the duration of the study and therefore the authors are confident that the acoustic signals

TABLE I. Distribution of sampling days into three time periods, number of different signal types recorded in each period, and the total amount of time in minutes recorded during each of the periods. Note that most singing occurred in the winter (time period 1).

			Simple calls		Complex calls		Song notes		Total	
	Time period	min	*n*	Signals/min	*n*	Signals/min	*n*	Signals/min	*n*	Signals/min
25 February 1 March 3 March 8 March	1	61	11	0.2	51	0.8	1090	17.9	1152	18.9
10 March 11 March 15 March	2	367	464	1.3	569	1.6	2849	7.8	3882	10.6
20 April 3 May 5 May 10 May	3	462	8	0.0	15	0.0	2034	4.4	2057	4.5

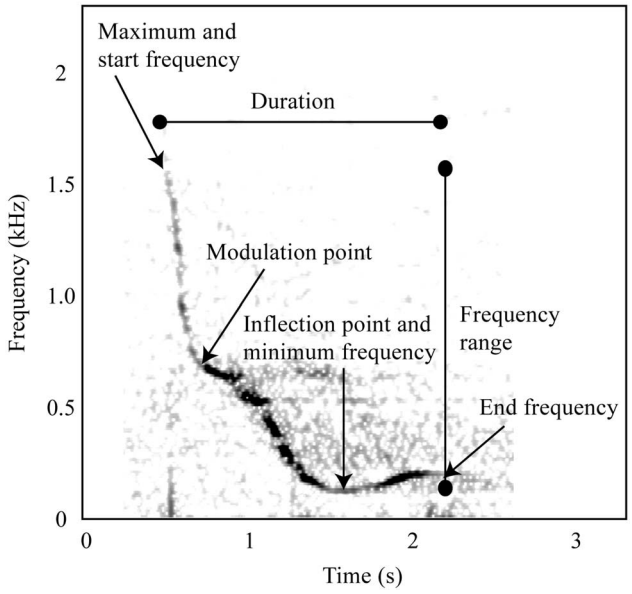


FIG. 2. A spectrogram of a type I song note together with a diagram illustrating the eight different variables measured for simple FM calls and song notes. (Hamming window, FFT size 1024, and 75% overlap).

described in this study were produced by bowhead whales. In addition to the bowhead whale there were two odontocete and five pinniped species inhabiting the study area at the time of the investigation. These were beluga whale *Delphinapterus leucas*, narwhal *Monodon monoceros*, hooded seal *Cystophora cristata*, bearded seal *Erignathus barbatus*, ringed seal *Phoca hispida*, harp seal *Phoca groenlandica*, and walrus *Odobenus rosmarus*. Bearded seal acoustic signals were recorded often from late March through May, but the other six species were not detected in any of the recordings.

Song notes were the most prevalent signal type ($n=5973$) followed by complex AM calls ($n=635$) and simple FM calls ($n=483$). Out of the 5973 song notes, 4115 song notes had a signal/noise ratio of ≥ 6 dB. These selected signals of high quality were analyzed in detail and used in the repertoire description. The remaining song notes ($n=1858$) had the distinctive qualities of song notes, but had signal/noise ratios <6 dB and were therefore used only in the signaling rate measurements.

A. Repertoire of bowhead whale vocalizations

1. Song notes

Songs were composed of song notes, which were narrow band FM signals. Song notes had an average duration of 1.32 ± 0.5 s (Table II). The average minimum and maximum frequencies of song notes were 390 and 982 Hz, respectively. The maximum frequency measured for song notes was 2638 Hz while the lowest frequency measured was 27 Hz. The number of inflection points ranged from 0 to 15, the average value being 1.1. The number of modulation points varied between 0 and 7, and the average value was 0.4. Song notes always appeared in the presence of another song note and in some cases the same song note type was repeated several times. Examples of song notes are shown in Fig. 3.

2. Complex AM calls

Complex calls included pulsative sounds comprised of short broadband pulses [Fig. 4(a)] as well as noisy bursts of sound [Fig. 4(b)] that did not have clear harmonic tonal structure. Complex AM calls had an average duration of 2.7 ± 1.1 s (Table II). The minimum frequency averaged at 91 Hz and maximum frequency at 495 Hz.

3. Simple FM calls and the constant frequency call

Simple calls were divided into three categories: the constant frequency call [Fig. 5(a)], the FM up call, which had an ascending frequency contour [Fig. 5(b)], and the FM down call, which had a descending contour [Fig. 5(c)]. The most common simple call type was by far the constant frequency

TABLE II. Time and frequency parameters of song notes and calls. See Fig. 2 for an illustration of the measured parameters.

		Duration (s)	Min (Hz)	Max (Hz)	Range (Hz)	Start (Hz)	End (Hz)	Inflections	Modulations
Song notes No.=4115	Average	1.3	390.2	981.5	591.2	962.5	466.2	1.1	0.4
	Median	1.3	335.7	952.3	530.8	907.3	404.0	1	0
	St. Dev.	0.5	263.4	282.2	381.0	288.2	281.1	1.5	0.7
	Min	0.2	26.6	125.7	35.9	83.0	26.6	0	0
	Max	7.0	1984.0	2636.4	2275.8	2636.4	2463.0	15	7
Simple FM calls No.=483	Average	1.4	155.9	208.3	52.4	196.6	169.1	0.1	0.0
	Median	171.0	31.2	163.7	131.5	28.6	0.0	0	0
	St. Dev.	0.9	111.0	143.1	68.7	133.2	127.9	0.3	0.3
	Min	0.2	41.0	70.1	7.0	41.0	54.1	0	0
	Max	6.6	1050.8	1263.3	682.5	1263.3	1050.8	2	5
Complex AM calls No.=635	Average	2.7	91.1	494.5	403.4				
	Median	2.8	88.6	432.8	346.1				
	St. Dev.	1.1	35.4	268.2	272.1				
	Min	0.5	20.1	128.7	31.2				
	Max	8.9	287.4	2222.5	2137.0				

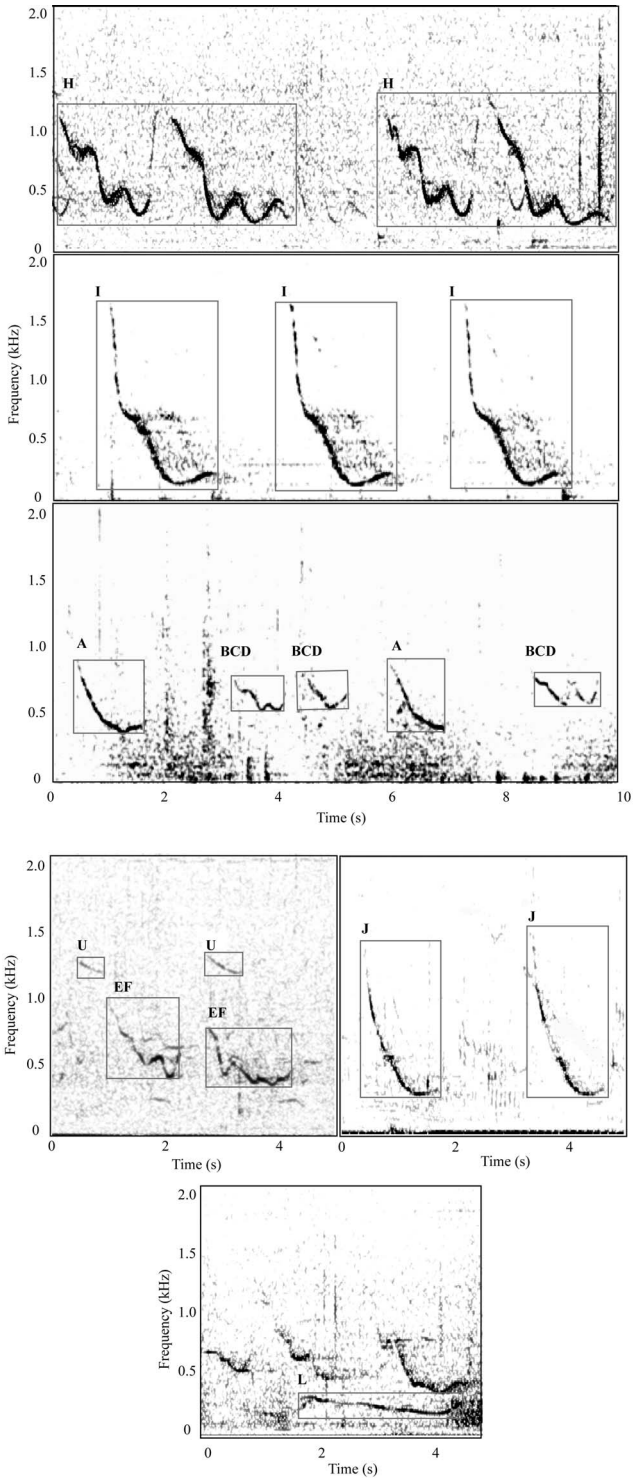


FIG. 3. Spectrograms of the eight song note types H, I, J, L, A, BCD, EF, and U (Hamming window, FFT size 1024, and 75% overlap).

call, which comprised 96.5% (n=466) of all simple calls. The FM up calls comprised 1.9% (n=9) and down calls 1.7% (n=8) of the data. Simple FM calls including the constant frequency call had an average duration of 1.4 ± 0.9 s (Table II). The constant frequency calls had an average minimum frequency of 156 Hz and an average maximum frequency of 208 Hz. Harmonics were present in 24% (n=117) of the simple calls. The number of harmonics ranged from 1 to 13 averaging at 3 ± 2. Harmonics could exceed 2000 Hz. Simple FM calls were not rich in inflection or in modulation points.

B. Classification of song notes

Using visual and audio qualities a classification of song notes was created and named A, B, C, D, E, F, H, I, J, L, and U. All the signals that did not fit into this categorization, but still had the distinctive qualities of a song note were named type x (n=48). The categorization was tested using multinomial log-linear regression analysis in S-PLUS 2000. The variables used in the analysis were duration, minimum frequency, maximum frequency, frequency range, start frequency, end frequency, number of inflection points, and number of modulation points. The results show that eight song note categories, instead of the original 11 categories, could be distinguished from each other using the variables measured (Table III). Using categorizations A, BCD, EF, H, I, J, L, and U the vast majority of the song notes (95.5%) were classified correctly by the multinomial log-linear regression model (Table III).

Each of the eight song note types had a characteristic frequency contour (Fig. 3) that was reflected in the number of inflection and modulation points and in the start and end frequencies of the signals (Table IV). Types BCD and U were short in duration (1.1 and 0.8 s) whereas type H had the longest average duration (3.6 s) (Table IV). Types I and L had both a low average minimum frequency (157 and 158 Hz), but type I had a broad frequency range (956 Hz) whereas type L had a narrow frequency range (237 Hz) (Table IV). Type J had the highest average maximum frequency (1337 Hz) even though the maximum frequency was measured for type I (2517 Hz) (Table IV).

By definition, song notes were always found in the presence of other song notes. Most of the type A song notes (76%, n=285) were encountered in the presence of one or more type BCD song notes. In 39.7% of these occasions (n=149) type A was followed by one type BCD song note. In 30.4% of the occasions (n=114) type A song note was followed by two BCD type song notes and in 5.1% (n=19) by three BCD song notes (Fig. 6). The least common combination consisted of a type A song note followed by four BCD notes, comprising only 0.8% (n=3) of the data. The remaining 24% of the A song notes (n=90) were observed following arbitrary song note types. Type BCD song notes were never observed without song note type A. Types EF, H, I, J, L, and U song notes were never observed singly; they were always observed with multiple copies of the same unit occurring in succession. For example, the following sequences would be observed: H, H or I, I, I (Fig. 3). Duration of song note combinations altered from a few seconds to a few minutes whereas the entire song session could last for hours.

C. Temporal changes in vocalizations

In order to investigate gradual changes in the vocal behavior in time, the study was divided into three time periods. Each period covers approximately 3 weeks of the season depending on the availability of the data. The time periods

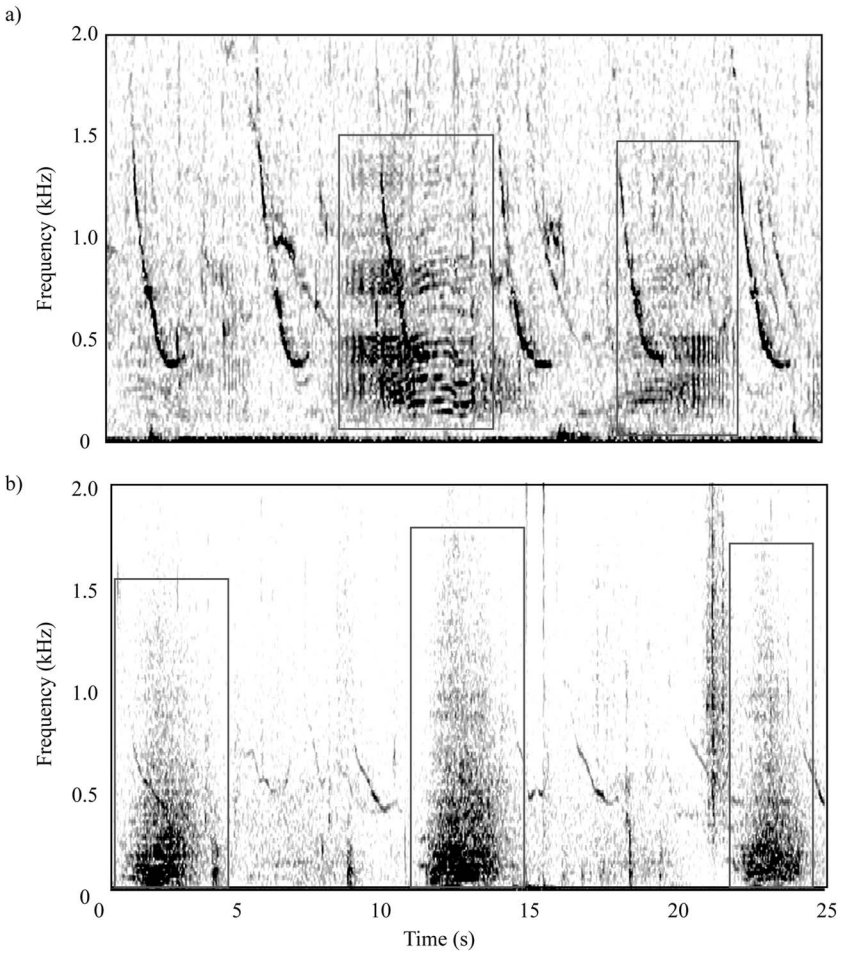


FIG. 4. Spectrograms of two pulsative complex AM calls. Those with harmonics (a) occur simultaneously with FM calls (see Fig. 5). Three complex AM calls with noisy burst-like appearances are shown in (b). All calls are inscribed in boxes (Hamming window, FFT size 1024, and 75% overlap).

were (1) 25 February–1 March 2005, (2) 3 March–15 March 2005, and (3) 20 April–10 May 2005. The overall signaling rate was the highest in the first time period when a total of 1152 signals were recorded in 61 min (18.9 signals min^{-1}) (Table I). The signaling rate decreased to 10.6 signals min^{-1} in the next time period and to 4.5 signals min^{-1} in the third time period. The entire study period was characterized by a high number of song notes per minute compared with the number of complex AM calls and simple FM calls per minute (Table I).

The types of song notes and rates of different song note types in the three time periods were distinctly different (Table V). There were five different types of song notes in the first winter time period and six different song note types in the second winter period. In contrast, only one song note, type I (see Fig. 3), was observed during the third time period, spring. This particular song note was not observed in the two winter time periods. Another difference in the singing between the first two time periods and the third time period was that in winter most of the song notes were emitted by two or

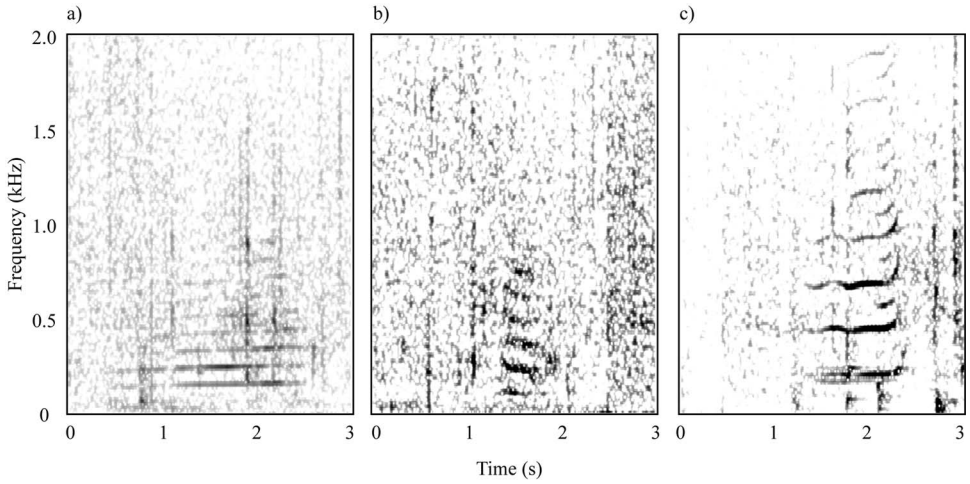


FIG. 5. Spectrogram of simple calls. A constant frequency call is shown in (a), a FM down call in (b), and a FM up call in (c). All have harmonics (Hamming window, FFT size 1024, and 75% overlap).

TABLE III. Multinomial log-linear regression analysis table for eight song note categories from lines to columns. The variables used were duration, minimum frequency, maximum frequency, frequency range, start frequency, end frequency, number of inflection points, and number of modulation points. The values are percentages. The values in bold indicate the percentage of signals that were classified to the type to which they were manually assigned. The numbers in bold print indicate the highest percentage score.

	%	A	BCD	EF	H	I	J	L	U
A	(n=374)	**93.6**	2.1	0.5	0.0	1.1	2.7	0.0	0.0
BCD	(n=309)	2.6	**95.1**	1.9	0.0	0.0	0.0	0.0	0.3
EF	(n=347)	2.9	1.1	**93.4**	0.6	1.7	0.3	0.0	0.0
H	(n=46)	0.0	0.0	10.9	**89.1**	0.0	0.0	0.0	0.0
I	(n=2015)	0.7	0.0	0.1	0.0	**98.3**	0.8	0.0	0.0
J	(n=170)	8.2	0.6	0.6	0.0	35.9	**54.7**	0.0	0.0
L	(n=47)	0.0	0.0	0.0	0.0	0.0	0.0	**100.0**	0.0
U	(n=759)	0.1	0.3	0.0	0.0	0.0	0.0	0.0	**99.6**

more animals at the same time because some of the calls were overlapping in time (Fig. 6). In spring typically only one animal was singing at a time.

There were some indications of the existence of multiple songs (Fig. 6). In the recording made on the 25 February 2005 song notes of type A were associated with song notes of type BCD. During the same sequence, song notes of type EF were also present together with song notes of type H. Signals of types EF and H never overlapped in time, but they were often being produced at the same time as note types A and BCD. This suggests that one individual was singing a song composed of type A and type BCD song notes and another individual sang a song consisting of types EF and H song notes. Alternatively, these combinations may represent two separate phrases from the seasonal song.

IV. DISCUSSION

This study presents the first description of the vocal behavior of bowhead whales in Disko Bay during the winter. The vocal repertoire of bowhead whales in Disko Bay included simple FM calls, complex AM calls, and songs composed of song notes. Song notes were classified into eight distinctive types. The acoustic behavior of bowhead whales during winter was characterized by a broad repertoire of song note types and a high signaling rate. In contrast, the acoustic behavior of the bowhead whales in the spring consisted of only one song note type and a considerably lower signaling rate.

A. Song notes and singing behavior

Song notes were the most common signals recorded in this study. The frequency range, duration, and number of inflection and modulation points were consistent with the values reported in previous investigations (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Richardson and Finley, 1989; Tervo, 2006; Stafford *et al.*, 2008). Stafford *et al.* (2008) reported higher frequencies for two of the song notes in the "Screech" song recorded from bowhead whales in Disko Bay in 2007 than found in this study. The high interannual variability in the song note repertoire of bowhead whales (Clark and Johnson, 1984; Würsig and Clark, 1993; Tervo *et al.*, 2007) makes it difficult to compare results from studies made in different years. A total of eight different song note categories could be identified in this study. The song notes appeared in succession but without locating the vocalizing individuals it is impossible to determine whether the song notes were produced by a single or multiple individuals therefore making it difficult to judge the duration of a "song." Bowhead whales are known to counter call in song like successions of calls (Blackwell *et al.*, 2007). This study indicates that the Davis Strait stock of bowhead whales emits a rich repertoire of diverse song note types, similar to what has been described for the Bering Sea population.

There were clear differences in the singing behavior of bowhead whales between winter and spring illustrated by a decrease in the song note signaling rate and a change in the song note types that were present. As found in previous studies, there were multiple individuals singing at the same time during winter in contrast to spring where typically only one animal was singing at a time (Würsig and Clark 1993; Tervo, 2006; Stafford *et al.*, 2008). The existence of multiple songs during winter is consistent with the previous studies of the Davis Strait population (Tervo, 2006; Stafford *et al.*, 2008) but whether the multiple songs were due to differences between individuals or sexes is not known. Winter singing was characterized by the presence of many different song note types. During the first two time periods from 25 February to 15 March 2005 there were seven different song note types present. The winter song note type U appears similar to the two- and three-syllable bouts documented by Clark and Johnson (1984) (see Figs. 1 and 4 in Clark and Johnson, 1984). Only one song note type, type I, was present in the spring time period between 20 April and 10 May 2005. This is in sharp contrast with the diversity of song note types present earlier in the season described in this and in previous studies (Tervo, 2006; Stafford *et al.*, 2008), and supports the hypothesis that the song recorded in the spring does not represent the entire richness of the repertoire of singing bowhead whales.

Singing is an acoustic display typically performed by males having significance in mate choice and sexual selection (Searcy and Andersson, 1986). Using acoustic cues to advertise fitness is a behavior that has been described from a wide variety of species of insects, amphibians, birds, and whales (Searcy and Andersson, 1986; Payne and McVay, 1971). If the song of bowhead whales is associated with

TABLE IV. Time and frequency parameters of eight song notes. See Fig. 2 for an illustration of the measured parameters and Figs. 4 and 5 for illustrations of the different types of song notes.

		Duration (s)	Min (Hz)	Max (Hz)	Range (Hz)	Start (Hz)	End (Hz)	Inflections	Modulations
Type A, $n=374$	Average	1.5	375.9	877.1	501.2	875.9	417.9	0.4	0.3
	Median	1.5	364.3	878.7	497.5	877.5	410.0	0	0
	St. Dev.	0.3	55.7	141.6	129.3	142.5	75.5	0.5	0.7
	Min	0.6	218.6	482.7	174.7	482.7	218.6	0	0
	Max	2.5	560.5	1293.5	864.4	1293.5	992.7	2	5
Type BCD, $n=309$	Average	1.1	496.3	795.7	299.4	781.1	675.7	1.2	1.1
	Median	1.1	500.9	794.0	288.7	776.4	682.0	1	1
	St. Dev.	0.3	61.9	104.5	95.2	114.2	108.0	0.8	1.2
	Min	0.4	301.1	546.4	105.1	390.4	385.0	0	0
	Max	2.4	633.6	1187.0	681.7	1187.0	994.7	5	6
Type EF, $n=347$	Average	1.6	343.9	1078.0	733.2	1063.9	455.8	3.9	0.3
	Median	1.6	346.0	1079.6	719.5	1065.6	430.5	4	0
	St. Dev.	0.4	61.7	174.9	172.5	186.9	127.6	1.4	0.6
	Min	0.9	216.6	655.5	351.7	83.0	244.6	0	0
	Max	2.6	573.8	1875.6	1263.9	1875.6	1196.9	8	3
Type H, $n=46$	Average	3.6	312.8	1256.9	944.1	1142.2	420.4	8.2	1.0
	Median	3.7	303.3	1236.2	938.1	1148.0	384.0	8	1
	St. Dev.	0.7	55.0	163.7	176.1	195.6	111.6	2.3	0.8
	Min	1.9	232.4	775.1	421.4	707.0	278.4	2	0
	Max	5.6	517.6	1519.8	1221.1	1519.8	738.0	12	3
Type I, $n=2015$	Average	1.4	157.4	1114.9	957.5	1114.9	210.5	0.7	0.6
	Median	1.5	151.7	1053.1	893.5	1053.1	211.0	1	1
	St. Dev.	0.4	62.5	289.6	297.4	289.6	71.9	0.5	0.6
	Min	0.2	26.6	454.2	280.8	454.2	26.6	0	0
	Max	7.0	1552.8	2516.9	2275.8	2516.9	1861.0	3	7
Type J, $n=170$	Average	1.2	352.8	1336.4	983.6	1336.4	376.7	0.4	0.1
	Median	1.2	301.5	1299.3	926.6	1299.3	335.3	0	0
	St. Dev.	0.4	107.7	315.3	322.0	315.3	106.3	0.5	0.3
	Min	0.5	220.3	657.0	369.1	657.0	226.9	0	0
	Max	2.4	666.0	2010.6	1736.8	2010.6	666.0	2	1
Type L, $n=47$	Average	2.5	158.4	395.2	236.8	301.1	379.1	1.0	0.0
	Median	2.5	147.4	319.4	163.7	297.0	311.2	1	0
	St. Dev.	0.3	27.8	203.6	212.4	24.8	212.4	0.1	0.0
	Min	1.8	122.8	278.4	98.3	257.0	163.8	0	0
	Max	3.2	229.3	1002.6	866.2	376.7	1002.6	1	0
Type U, $n=759$	Average	0.8	946.2	1107.7	161.6	1014.9	1038.9	0.0	0.0
	Median	0.8	887.9	1037.8	159.7	893.3	1025.8	0	0
	St. Dev.	0.3	142.6	133.8	37.7	213.1	75.2	0.0	0.0
	Min	0.2	399.9	562.1	45.8	562.1	399.9	0	0
	Max	1.8	1324.6	1510.2	371.2	1457.6	1510.2	1	1

reproductive advertisement, then Disko Bay may be a mating area for the Davis Strait population of bowhead whales in the winter. Previous studies of the conception time of bowhead whales from the Bering Sea indicate that most conceptions take place between early March and mid-April (Reese *et al.*, 2001), which coincides with the timing of complex song found in this and in previous studies (Tervo, 2006; Stafford *et al.*, 2008). Sexual activity early in the season may shift into foraging behavior in the spring when waters off Disko Island are rich in copepods after the spring algae bloom (Madsen *et al.*, 2001; Laidre *et al.*, 2007). This shift in behavior could explain the low signaling rate and the presence of a less complex song in the spring months. The number of bowhead whales and the duration that individual whales reside in Disko Bay in February and March have not been studied. However, large numbers of bowhead whales are present in Disko Bay in April and May (Heide-Jørgensen *et al.*, 2003, 2007c), indicating that the decline in acoustic activity is not a result of a significant decrease in numbers of whales in the area.

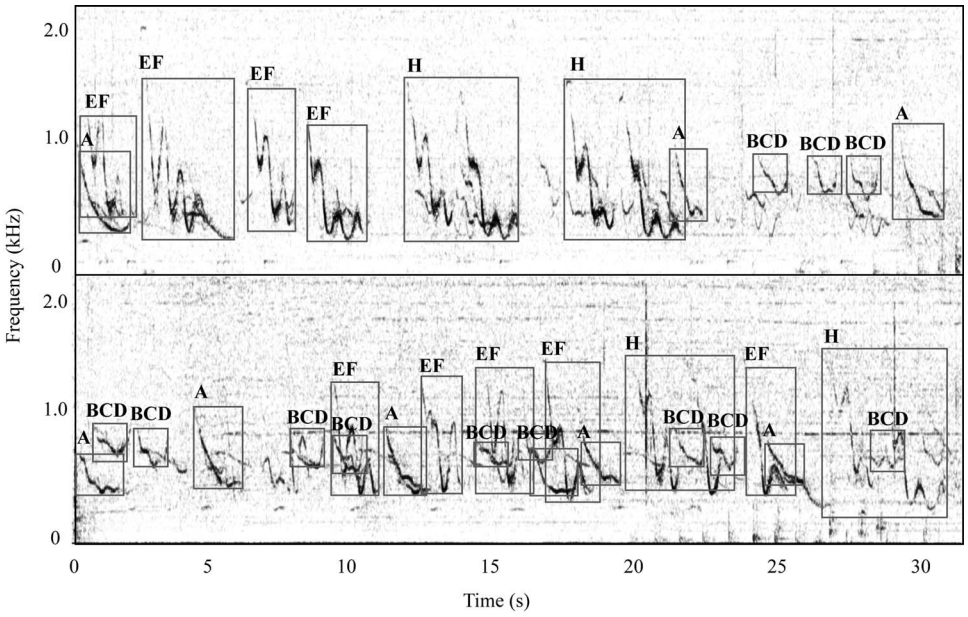


FIG. 6. A spectrogram showing a section of a recording where individual song note types are overlapping in time indicating more than one animal singing at the same time. Types A, BCD, and EF were the dominating song note types. The recording was made on 25 February 2005 (Hamming window, FFT size 1024, and 75% overlap).

B. Simple FM calls and complex AM calls

The simple FM call was the rarest of the three signal types recorded from bowhead whales in present study (Table II). This is in sharp contrast with some of the previous investigations where simple FM calls were the most numerous of the described signal types from the Bering Sea stock (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Ljungblad *et al.*, 1984). Furthermore, simple FM calls have been recorded in May in large number (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Ljungblad *et al.*, 1984; Cummings and Holliday, 1987) whereas the last time period from 20 April to 10 of May 2005 in the present study exhibited lowest signaling rate for simple calls compared with the two previous ones (Table I). However, these differences could be explained by diurnal variation in the signaling rate, which could not be captured in this study.

Simple calls have not been assigned to any particular behavior (Ljungblad *et al.*, 1984), but simple calls with ascending and descending frequencies, referred to as up and down calls (Richardson and Finley, 1989), have been recorded in the presence of socially and sometimes sexually active whales (Ljungblad *et al.*, 1984). The frequency range, duration, and number of inflection and modulation points of simple FM calls were consistent with previously documented values from the Bering Sea population in May (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Ljungblad *et al.*, 1984; Cummings and Holliday, 1987) and September–October (Ljungblad *et al.*, 1982) and from the Davis Strait population in Isabella Bay in August and September (Richardson and Finley, 1989). Simple calls have been also recorded from the Davis Strait population in April (Stafford *et al.*, 2008), but information on call parameters is lacking for comparison.

Complex AM calls were the second most commonly recorded signal type. The average duration measured for the signals in the present study was consistent with measurements from the same kinds of sounds emitted by bowhead whales from the Bering Sea population in May (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Ljungblad *et al.*, 1984; Cummings and Holliday, 1987) and September–October (Ljungblad *et al.*, 1982) and from the Davis Strait population in Isabella Bay in August and September (Richardson and Finley, 1989). A majority of the complex AM calls recorded in Disko Bay had a minimum frequency lower than 100 Hz and the lowest minimum frequency recorded was 20 Hz. Low complex AM calls with minimum frequencies at 25 Hz have been recorded from socially active bowhead whales from the Davis Strait stock in Isabella Bay in August and September (Richardson and Finley, 1989) and from migrating bowhead whales off Point Barrow, Alaska in May and September–October with minimum frequencies starting from 30 Hz (Ljungblad *et al.*, 1982). The average minimum frequency and maximum frequency of complex calls analyzed in this study were consistent with the documented values from the Bering Sea population (Ljungblad *et al.*, 1982; Clark and Johnson, 1984; Cummings and Holliday, 1987) and Davis Strait population (Richardson and Finley, 1989). Complex AM calls have been also recorded

TABLE V. Call rates (signals/min) of song note types A, BCD, EF, H, I, J, L, and U in the three time periods. The last column of the table indicates the number of different song note types recorded in each time period (see Table I). Note the markedly higher rates of calling in the winter time period 1.

Time period	A	BCD	EF	H	I	J	L	U	No. of note types
1	5.90	5.07	4.46	0.70	0	0	0	0.23	5
2	0.04	0	0.20	0.01	0	0.46	0.13	2.03	6
3	0	0	0	0	4.36	0	0	0	1

from the Davis Strait population in April (Stafford *et al.*, 2008), but information on call parameters was not presented for comparison.

Complex calls have not been positively associated with a particular behavior in bowhead whales (Ljungblad *et al.*, 1984; Würsig and Clark, 1993); however, they have been often recorded in the presence of mildly socializing (within a body length) or actively socializing (body contact) bowhead whales (Ljungblad *et al.*, 1984; Würsig *et al.*, 1984) and in the presence of sexually active whale groups (Richardson and Finley, 1989). Complex calls were also characteristic emissions of groups engaging in homosexual activity (Richardson and Finley, 1989). In the present study, complex calls had the highest signaling rate in the second time period from 3 March to 15 March (Table I). Since complex calls in previous studies were often recorded in the presence of socially and sexually active whales, the high signaling rate of complex calls in March observed in this study and the presence of AM calls in early April (Stafford *et al.*, 2008) further support the hypothesis of sexual activity in Disko Bay during winter.

Long-term studies of the acoustic behavior of the Davis Strait population are needed to describe inter-annual variation in the song of the bowhead whales in Disko Bay. Additional comparisons of the repertoires of the Davis Strait population with those of the Bering Sea population can be used to determine the similarities of the acoustic repertoires between the two stocks and furthermore evaluate stock connectivity. Similarities in the vocalizations between two stocks can indicate that the stocks are connected like in the case of humpback whales off Western and Eastern Australia (Noad *et al.*, 2000).

Determining the sex(es) of the singing individuals in this species will be very important to gain insight into the sexual behavior and mate choice strategy of bowhead whales. The bowhead whale is taxonomically closely related to the North Atlantic and southern right whales *Eubalaena glacialis* and *E. australis* (Reeves and Leatherwood, 1985). Males from all the three species possess disproportionately large testes suggesting that sperm competition plays a role in the sexual selection and mating strategy of the species (Brownell and Ralls, 1986). Southern and North Atlantic right whales exhibit signs of a polyandrous mating system where one female mates with multiple males in large mating groups referred to as SAGs (Kraus and Hatch, 2001). Right whales are not known to sing (Clark, 1982), instead, sexual selection pressure on males is thought to take place in the form of sperm competition (Kraus and Hatch, 2001; Mate *et al.*, 2005). Bowhead whales sing complex songs with high inter-annual variation (Clark and Johnson, 1984; Würsig and Clark, 1993; Tervo *et al.*, 2007) much in the manner of humpback whales (Winn and Winn, 1978). Male humpback whales produce long and elaborate songs as an advertisement display fulfilling the criteria of a lekking polygynous species where one male mates with multiple females (Clapham, 1996). Interestingly, bowhead whales seem to exhibit characteristics of both polyandrous and polygynous mating strategies—they have been seen in social groups similar to right whale SAGs (Würsig and Clark, 1993) and they produce songs such as humpback whales. Determining the sex and the size of the singers could greatly enhance the current understanding of the mating system of this species. If male bowhead whales produce songs, it would imply that males may have multiple mating strategies, including acoustic advertisement displays and sperm competition. Which of the two strategies is used could depend on the age and/or the social status of the males as suggested by Würsig and Clark (1993). However, it could be that both sexes produce songs independently or even simultaneously in the form of a duet known from various song bird species (Harcus, 1977; Hall, 2004). If so, this would be the first case of sex-role reversal in cetaceans, where females produce an advertisement display in the form of a complex song.

ACKNOWLEDGMENTS

This study was funded by DANCEA, the Danish Research Council for Natural Sciences (Grant No. 21-03-0171 to L.A.M. and Grant No. 21-04-0391 to Torkel Gissel Nielsen, National Environmental Research Institute, Roskilde, Denmark) and the Centre for Sound Communication, University of Southern Denmark supported by the Danish National Research Council. The Arctic Station of Qeqertarsuaq, University of Copenhagen provided an excellent working platform. The authors want to thank Abel Brandt, Tarfi Mølgaard, Kale Mølgaard, John Jakobsen, and the crew on RV Porsild for their help in data collection. Chris Clark provided valuable assistance in identifying bowhead whale acoustic signals. Rasmus Ejnærs was a great help in the data analysis and Mads Christoffersen gave valuable comments to this manuscript.

Baker, C. S., and Herman, L. M. (**1981**). "Aggressive behavior between humpback whales (*Megaptera novaeangliae*) wintering in Hawaiian waters," Can. J. Zool. **59**, 460–469.

Baker, C. S., Lambertsen, R. H., Weinrich, M. T., Calambodikis, J., Early, G., and O'Brien, S. J. (**1991**). "Molecular genetic identification of the sex of humpback whales (*Megaptera novaeangliae*)," Rep. Int. Whal. Comm. **13**, 105–111.

Blackwell, S. B., Richardson, W. J., Greene, C. R., Jr., and Streever, B. (**2007**). "Bowhead whale (*Balaena mysticetus*) migration and calling behaviour in the Alaskan Beaufort Sea, autumn 2001–04: An acoustic localization study," Arctic **60**, 255–270.

Brownell, R. L., and Ralls, K. (**1986**). "Potential for sperm competition in baleen whales," Rep. Int. Whal. Comm. **8**, 97–112.

Clapham, P. J. (**1996**). "The social and reproductive biology of humpback whales: An ecological perspective," Mammal Rev. **26**, 27–49.

Clark, C. V., Cortopassi, K. A., Ponirakis, D., Fowler, M. C., Fristrup, K. M., and George, J. C. (**2003**). "Seasonal variation in acoustic characteristics of bowhead whale (Balaena mysticetus) sounds during the spring 2001 migration of Pt. Barrow, Alaska," Document No. SC/56/BRG21, International Whaling Commission.

Clark, C. W. (**1982**). "The acoustic repertoire of the southern right whale, a quantitative analysis," Anim. Behav. **30**, 1060–1071.

Clark, C. W. (**1991**). "Acoustic behavior of mysticete whales," in *Sensory Abilities of Cetaceans*, edited by J. Thomas and R. Kastelein (Plenum, New York), pp. 571–583.

Clark, C. W., and Clapham, P. J. (**2004**). "Acoustic monitoring on a humpback whale (Megaptera novaeangliae) feeding ground shows continual singing into late spring," Proc. R. Soc. London, Ser. B, **271**, 1051–1057.

Clark, C. W., and Johnson, J. H. (**1984**). "The sounds of the bowhead whale, *Balaena mysticetus*, during the spring migrations of 1979 and 1980," Can. J. Zool. **62**, 1436–1441.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B. R., Flores, S., Gedamke,

J., and Urban, J. (**2002**). "Only male fin whales sing loud songs," Nature (London) **417**, 809.
Cummings, W. C., and Holliday, D. V. (**1987**). "Sounds and source levels from bowhead whales off Pt. Barrow, Alaska," J. Acoust. Soc. Am. **82**, 814–821.
Cummings, W. C., and Thompson, D. (**1971**). "Underwater sounds from the blue whale, *Balaenoptera musculus*," J. Acoust. Soc. Am. **50**, 1193–1198.
Eschricht, D. F., and Reinhardt, J. (**1861**). *Om Nordhvalen (Balæna Mysticetus L.) Navnlig med Hensyn til Dens Udbredning i Fortiden og Nutiden og til Dens Ydre og Indre Særkjender (The Northern Whale (Balaena Mysticetus L.) with especial reference to its geographical distribution in times past and present, and to its external and internal characteristics)* (Bianco Lunos Bogtrykkeri, Copenhagen).
Glockner, D. A. (**1983**). "Determining the sex of humpback whales (Megaptera novaeangliae) in their natural environment," in *Communication and Behavior of Whales*, edited by R. S. Payne (American Association for the Advancement of Science, Westview, Boulder, CO), pp. 447–464.
Hall, M. L. (**2004**). "A review of hypotheses for the functions of avian duetting," Behav. Ecol. Sociobiol. **55**, 415–430.
Harcus, J. L. (**1977**). "The functions of vocal duetting in some African birds," Z. Tierpsychol **43**, 23–45.
Heide-Jørgensen, M. P., Cosens, S. E., Dueck, L. P., Laidre, K., and Postma, L. (**2007a**). "Baffin Bay–Davis Strait and Hudson Bay–Foxe Basin bowhead whales: A reassessment of the two-stock hypothesis," Document No. SC/60/BRG20, International Whaling Commission.
Heide-Jørgensen, M. P., Laidre, K., Borchers, D., Samarra, F., and Stern, H. (**2007b**). "Increasing abundance of bowhead whales in West Greenland," Biol. Lett. **3**, 577–580.
Heide-Jørgensen, M. P., Laidre, K., Jensen, M. V., Dueck, L., and Postma, L. D. (**2006**). "Dissolving stock discreteness with satellite tracking: Bowhead whales in Baffin Bay," Marine Mammal Sci. **22**, 34–45.
Heide-Jørgensen, M. P., Laidre, K., Wiig, Ø., Bachmann, L., Lindqvist, C., Postma, L., Dueck, L., Lidsay, M., and Tenkula, D. (**2007c**). "Segregation of sexes and plasticity in site selection of bowhead whales," Document No. SC/60/BRG19, International Whaling Commission.
Heide-Jørgensen, M. P., Laidre, K. L., Wiig, Ø., Jensen, M. V., Dueck, L., Schmidt, H. C., and Hobbs, R. C. (**2003**). "First successful satellite tracking of bowhead whales, *Balaena mysticetus*, in Baffin Bay: From Greenland to Canada in ten days," Arctic **56**, 21–31.
Kraus, S. D., and Hatch, J. J. (**2001**). "Mating strategies in the North Atlantic Right whale (*Eubalaena glacialis*)," J. Cetacean Res. Manage. **2**, 237–244.
Laidre, K., Heide-Jørgensen, M. P., and Nielsen, T. G. (**2007**). "Role of the bowhead whale as a predator in West Greenland," Mar. Ecol.: Prog. Ser. **346**, 285–297.
Ljungblad, D. K., Moore, S. E., and Van Schoik, D. R. (**1984**). "Aerial surveys of endangered whales in the Northern Bering, Eastern Chukchi and Alaskan Beaufort seas, 1983: With a five year review, 1979–1983," Naval Ocean Systems Center Technical Report No. 955. Minerals Management Service, Alaska OCS Region, U.S. Department of Interior, San Diego, CA.
Ljungblad, D. K., Thompson, P. O., and Moore, S. E. (**1982**). "Underwater sounds recorded from migrating bowhead whales, *Balaena mysticetus*, in 1979," J. Acoust. Soc. Am. **71**, 477–482.
Madsen, S. D., Nielsen, T. G., and Hansen, B. W. (**2001**). "Annual population development and production by *Calanus finmarchicus, C. glacialis* and *C. hyperboreus* in Disko Bay, Western Greenland," Mar. Biol. (Berlin) **139**, 75–93.
Mate, B., Duley, P., Lagerquist, B., Wenzel, F., Stimpert, A., and Clapham, P. (**2005**). "Observation of a female North Atlantic Right Whale (*Eubalaena glacialis)* in simultaneous copulation with two males: Supporting evidence for sperm competition," Aquat. Mamm. **31**, 157–160.
McDonald, M. A., Calambokidis, J., Teranishi, A. M., and Hildebrand, J. A. (**2001**). "The acoustic calls of blue whales off California with gender data," J. Acoust. Soc. Am. **109**, 1728–1735.
Moore, S. E., and Reeves, R. R. (**1993**). "Distribution and movements," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (The Society of Marine Mammalogy, Lawrence, KS), Special Publication No. 2, pp. 313–386.
Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., and Jenner, K. C. S. (**2000**). "Cultural revolution in whale songs," Nature (London) **408**, 537.
Oleson, E. M., Calambokidis, J., Burgess, W. C., McDonald, M. A., LeDuc, C. A., and Hildebrand, J. A. (**2007**). "Behavioral context of call production by eastern North Pacific blue whales," Mar. Ecol.: Prog. Ser. **330**, 269–284.
Parks, S. E., and Tyack, P. L. (**2005**). "Sound production by North Atlantic right whales (*Eubalaena mysticetus*) in surface active groups," J. Acoust. Soc. Am. **117**, 3297–3306.
Payne, R. S., and McVay, S. (**1971**). "Songs of Humpback Whales," Science **173**, 585–597.
Reese, S. C., Calvin, J. A., George, J. C., and Tarpley, R. J. (**2001**). "Estimation of fetal growth and gestation in bowhead whales," J. Am. Stat. Assoc. **96**, 915–938.
Reeves, R. R., and Leatherwood, S. (**1985**). "Bowhead whale," in *Handbook of Marine Mammals*, edited by S. H. Ridgway and R. Harrison (Academic, London), pp. 305–344.
Richardson, J. W., and Finley, K. J. (**1989**). "Comparison of behavior of bowhead whales of the Davis Strait and Bering/Beaufort Stocks," LGL Limited, Environmental Research Associates, prepared for Minerals Management Service, Alaska OCS Region, U.S. Department of Interior.
Richardson, W. J., Finley, K. J., Miller, G. W., Davis, R. A., and Koski, W. R. (**1995**). "Feeding, social and migration behavior of bowhead whales, *Balaena mysticetus*, in Baffin Bay vs. the Beaufort Sea: Regions with different amounts of human activity," Marine Mammal Sci. **11**, 1–45.
Searcy, W. A., and Andersson, M. (**1986**). "Sexual selection and the evolution of song," Annu. Rev. Ecol. Syst. **17**, 507–533.
Stafford, K. M., Moore, S. E., Laidre, K. L., and Heide-Jørgensen, M. P. (**2008**). "Bowhead whale springtime song off West Greenland," J. Acoust. Soc. Am. **124**, 3315–3323.
Tervo, O. M. (**2006**). "Vocalisation of bowhead whales (Balaena mysticetus) in Disko Bay, Western Greenland, in relation to oceanography and plankton distribution," MSc thesis, University of Southern Denmark, Denmark.
Tervo, O. M., Parks, S. E., and Miller, L. A. (**2007**). "Annual and seasonal changes in the song of the bowhead whale Balaena mysticetus in Disko Bay, Western Greenland," 17th Biennial Conference on the Biology of Marine Mammals, Cape Town, 29 November–3 December.
Tyack, P. L. (**1983**). "Differential response of humpback whales, *Megaptera novaeangliae*, to playback of song or social sounds," Behav. Ecol. Sociobiol. **13**, 49–55.
Tyack, P. L., and Clark, C. W. (**2000**). "Communication and acoustic behavior of dolphins and whales," in *Hearing by Whales and Dolphins*, edited by A. Whitlow, A. Popper, and R. Fay (Springer Handbook for Auditory Research, New York), pp. 156–224.
Watkins, W. A., Tyack, P., Moore, K. E., and Bird, J. E. (**1987**). "The 20-Hz signals of finback whales (*Balaenoptera physalus*)," J. Acoust. Soc. Am. **82**, 1901–1912.
Winn, H. E., and Winn, L. K. (**1978**). "The song of the humpback whale *Megaptera novaeangliae* in the West Indies," Mar. Biol. (Berlin) **47**, 97–114.
Würsig, B., and Clark, C. (**1993**). "Behavior," in *The Bowhead Whale*, edited by J. J. Burns, J. J. Montague, and C. J. Cowles (The Society of Marine Mammalogy, Lawrence, KS), Special Publication No. 2, pp. 157–200.
Würsig, B., Clark, C. W., Dorsey, E. M., Richardson, W. J., and Wells, R. S. (**1983**). "Normal behavior of bowheads 1982," in *Behavior, Disturbance Responses and Distribution of Bowhead Whales Balaena Mysticetus in the Eastern Beaufort Sea, 1982*, edited by W. J. Richardson (LGL Ecological Research Associates for U.S. Minerals Management Service, Reston, VA).

Comparison of directional selectivity of hearing in a beluga whale and a bottlenose dolphin

Vladimir V. Popov and Alexander Ya. Supin[a)]
Institute of Ecology and Evolution, Russian Academy of Sciences, 33 Leninsky Prospect, 119071 Moscow, Russia

(Received 11 February 2009; revised 22 May 2009; accepted 17 June 2009)

Hearing thresholds as a function of sound-source azimuth were measured in a beluga whale *Delphinapterus leucas* and a bottlenose dolphin *Tursiops truncatus* in identical conditions using the auditory evoked-potential method. In both the beluga whale and bottlenose dolphin, the receiving beam width narrowed with frequency increase. At all frequencies, the receiving beam was markedly wider in the beluga whale than in the bottlenose dolphin. In particular, the 3-dB beam width in the beluga whale narrowed from ±33.5° at 8 kHz frequency to ±14.3° at 128 kHz; the 6-dB beam width narrowed from ±56.9° to ±18.9°, respectively. In the bottlenose dolphin, the 3-dB beam width decreased from ±19.9° at 8 kHz to ±6.3° at 128 kHz; the 6-dB beam width decreased from ±33.1° to ±8.4°, respectively. In the bottlenose dolphin, the axis of the low-frequency receiving beam deviated from the midline up to 15°; in the beluga whale, this effect was not detected. The audiograms of both the beluga whale and bottlenose dolphin were azimuth-dependent: from an audiogram featuring the best sensitivity at intermediate frequencies at 0° to that featuring monotonous threshold increase with frequency increase at 90°. In the beluga whale, this dependence was less prominent than in the bottlenose dolphin.
 [DOI: 10.1121/1.3177273]

PACS number(s): 43.80.Lb [WWA] Pages: 1581–1587

I. INTRODUCTION

Odontocetes (toothed whales, dolphins, and porpoises) feature high sensitivity, a wide frequency range, acute frequency tuning, and high temporal resolution of hearing (review Au, 1993; Au *et al.*, 2000; Supin *et al.*, 2001). One more important feature of hearing is its directional sensitivity, which is the dependence of hearing sensitivity on the direction to the sound source. Directional sensitivity forms a receiving beam, which, being directed to a target sound source, allows the minimization of the perception of other interfering sounds.

There have been a few attempts to behaviorally investigate the receiving beam patterns in dolphins (Zaytseva *et al.*, 1975; Au and Moore, 1984). However, in those studies, absolute thresholds were not measured as a function of the sound-source position. Instead of that, masked thresholds were measured as a function of angular distance between the probe and masker sound sources. Based on those results, the directivity indices of the receiving beam were calculated at sound frequencies of 30, 60, and 120 kHz and the indices were estimated as 10.4, 15.3, and 20.6 dB, respectively (Au and Moore, 1984). However, the results obtained in the masking experimental paradigm might depend more on mutual positions of the probe and masker sources instead of the sound-source position relative to the head axis (which gives the true receiving beam pattern). As measurements of a minimum audible angle (Renaud and Popper, 1975; Dubrovskiy, 1990) have demonstrated, dolphins were capable of discriminating the mutual positions of two sound sources with much higher resolution (a few angular degrees) than dictated by the receiving beam width. A recent attempt at the direct behavioral investigation of the receiving beam width in a harbor porpoise (Kastelein *et al.*, 2005) gave somewhat ambiguous results because of rather large steps between the discriminated sound-source positions and because of the experimental design that made the results dependent on the sensory-to-motor coordination ability of the subject.

Direct measurements of the receiving beam pattern based on absolute thresholds as a function of sound-source azimuth have been made using the evoked-potential method in a few dolphin species (Popov and Supin, 1988, 1992; Popov *et al.*, 1992). According to those results, the −3-dB beam width varied from ±6° to ±12°. A more precise investigation of both binaural and monaural receiving beams in a bottlenose dolphin *Tursiops truncatus* at a variety of sound frequencies allowed the characterization of the receiving beam by the directivity index: Within a range from 11.2 to 128 kHz, the directivity index rose from 4.7 to 17.8 dB on the ipsilateral side and from 10.5 to 15.6 dB on the contralateral side (Popov *et al.*, 2006).

It is reasonable to suppose that in creating the directional selectivity in odontocetes, an important role is played by the shadowing of sound by some of the head structures (the skull, air cavities, and others). Therefore, it may be of interest to compare the directional selectivity of hearing in odontocete species featuring different head and skull shapes. For this sort of comparison, the bottlenose dolphin *Tursiops truncates* and beluga whale *Delphinapterus leucas* may be appropriate because the skull of the beluga whale is less elongated than that of the bottlenose dolphin. Both bottlenose

[a)]Author to whom correspondence should be addressed. Electronic mail: alex_supin@mail.ru

dolphins and beluga whales are regularly kept in captivity, so the investigation of these species is feasible. The hearing of the bottlenose dolphin is well investigated in many respects, including spatial selectivity; however, its receiving beam pattern still needs to be characterized in more detail. As to the beluga whale, an investigation of the directional selectivity of its hearing has been attempted (Klishin *et al.*, 2000). However, a confusing result of that study was that directional selectivity did not depend on frequency, whereas consideration of hearing directivity in terms of a receiving aperture predicts directivity increase with frequency, which was really observed in other dolphin species. No explanation was found for that result, which therefore needs to be re-checked.

The goal of the present study was a comparative investigation of the receiving beam patterns in representatives of the two odontocete species, the beluga whale and bottlenose dolphin, in identical conditions. For that, the evoked-potential technique was used, which has demonstrated it effectiveness for the investigation of basic hearing characteristics in odontocetes (Supin *et al.*, 2001).

II. MATERIAL AND METHODS

A. Subjects

The experimental animals were a young beluga whale *Delphinapterus leucas*, male, weighing 155 kg, and a young bottlenose dolphin *Tursiops truncatus*, male, weighing 110 kg. Both subjects were kept in the Utrish Marine Station of the Russian Academy of Sciences, on the Black Sea. Each of the animals was housed in an on-land seawater pool $9 \times 4 \times 1.2$ m^3.

B. Experimental conditions

The experimental conditions were the same for both subjects. During the experiments, the animal was placed into a circular experimental tank filled with seawater, 6 m in diameter, 0.45 m deep (Fig. 1). The animal (1) on a stretcher (2) was positioned in such a way that the main part of its body was in water but its blowhole and a part of its back were above the water. The animal's head was near the center of the tank. A sound-emitting transducer (3) was mounted on a bar (4), which laid on a support (5) and could be rotated around a center above the animal's melon tip. The transducer was located at a distance of 1.2 m from the rotation center. At the mid-distance (0.6 m) between the rotation center and transducer, there were 80-cm long baffles (6) extending 15 cm upward of the bottom and 15 cm downward of the water surface, with a 15-cm slit between them. These baffles were intended for allowing direct sounds to spread from the transducer to the head whereas diminishing spread of sounds reflected from the bottom and surface. Rotation of the bar allowed the placement of the transducer at varying azimuth angles relative to the longitudinal head axis. The rotation center coincided with the tank center, so the distance from the transducer to the nearest tank wall was 1.8 m. Therefore, the path for sounds reflected from the walls to the animal's head was at least 3.6 m longer (the delay at least 2.4 ms longer) than the direct way from the transducer. The position of the animal's head was monitored by a video camera (7) and measurements were done only when deviation of the head midline from the zero directions did not exceed ±5°.

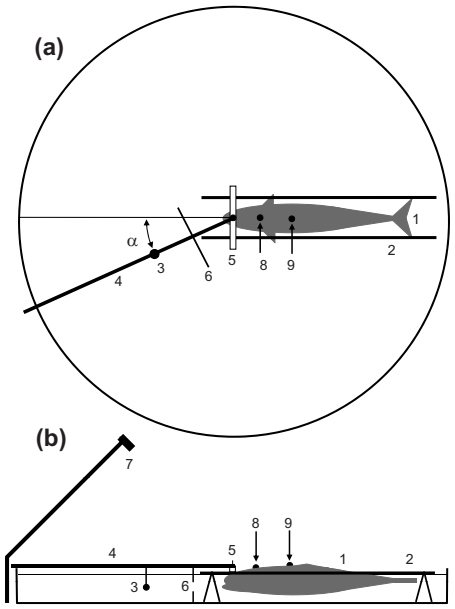


FIG. 1. Experimental design. (a) Dorsal view and (b) lateral view. (1) Animal, (2) stretcher, (3) transducer, (4) rotating bar, (5) bar support, (6) baffles, (7) video camera, (8) active electrode, and (9) reference electrode. α is sound-source azimuth.

C. Evoked-potential recording

For non-invasive evoked-potential recording, suction-cup electrodes were used consisting of a 10-mm stainless-steel disk mounted within a 50-mm silicon suction cup. The active electrode was fixed at the vertex of the animal's head surface, 5 cm behind the blowhole, above the water surface [Fig. 1(8)]. The reference electrode was fixed at the back [Fig. 1(9)]. The electrodes were connected by shielded cables to the input of a custom-made amplifier based on the AD620 (Analog Devices) chip. The amplifier provided 80-dB gain within a frequency range of 200–5000 Hz. The amplified signal was digitized and collected using an E-6040 data acquisition board (National Instruments) and stored in computer memory. To extract signal from noise, the digitized signal was coherently averaged (1000 original records per 1 averaged record) using triggering from the stimulus onset.

D. Acoustic stimulation

The acoustic stimuli were short sound pips designed by modulation of a carrier frequency by one 0.5-ms cycle of a cosine envelope. Carrier frequencies varied from 8 to 128 kHz by $\frac{1}{4}$-octave steps. Stimulus levels are specified below in root-mean-square (rms) levels relative to 1 μPa.

Sound signals were digitally synthesized at an update rate of 500 kHz and digital-to-analog converted by the E-6040 board, amplified, attenuated, and played through a B&K 8104 transducer. The transducer was placed at a distance of 1.2 m from the melon tip of the animal, at a depth of 22 cm (the mid-depth between the tank bottom and water surface). The azimuth position of the transducer (α in Fig. 1) varied within a range of ±90° from the longitudinal head

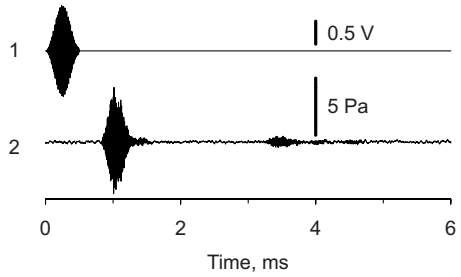


FIG. 2. Stimulus waveforms: digitally synthesized cosine-enveloped pip of 64 kHz carrier frequency, 0.5-ms duration (1), and the acoustic signal received by the hydrophone in the pool center from the transducer 1.2 m away of the center (2).

axis, by steps of 7.5° within a range of ±30° and by steps of 15° within a range 30°–90°. The play-back channel was calibrated both before and after the experiments by positioning a calibrated receiving hydrophone (B&K 8103) at the same location as the animal's head. It showed satisfactory reproduction of the signal waveform in the directly spreading sound with the wall-reflected sounds delayed not less than 2.4–2.5 ms (Fig. 2).

E. Threshold evaluation

For threshold evaluation, the stimulus level was decreased from an obviously supra-threshold to a sub-threshold level. The peak-to-peak auditory evoked-potential (AEP) amplitude was plotted as a function of stimulus level, and a 15- to 20-dB near-threshold part of the plot was approximated by a straight regression line. The intersection of this line with the zero-amplitude level was adopted as a threshold estimate. This threshold-determination procedure was repeated at different sound frequencies and sound-source positions. The resulting threshold-vs-azimuth and threshold-vs-frequency functions were taken as the receiving beam pattern and the audiogram, respectively.

III. RESULTS

A. AEP features

The stimuli provoked AEPs, which displayed all features of the auditory brainstem response, as described in a number of odontocete species. In both of the investigated subjects, the responses looked qualitatively similar (Fig. 3). They had an onset latency of 2.5–3 ms, including the acoustic delay (0.8 ms at the sound-source distance of 1.2 m), and consisted of a few waves each lasting around 1 or less than 1 ms. Quantitatively, the responses differed between the two subjects: In the beluga whale, responses were of longer latency and duration [Figs. 3(a) and 3(b)] than in the bottlenose dolphin [Figs. 3(c) and 3(d)].

A feature of the records obtained at the lateral sound-source positions was the presence of two sets of AEP waves delayed by 2.5–3 ms relative to one another [Figs. 3(b) and 3(d)]. This delay corresponded well to the acoustic delay of the signal reflected from the pool walls relative to that spreading directly from the transducer. Therefore, the first set of AEP waves was considered as a response to the signal directly spreading from the sound source to the head,

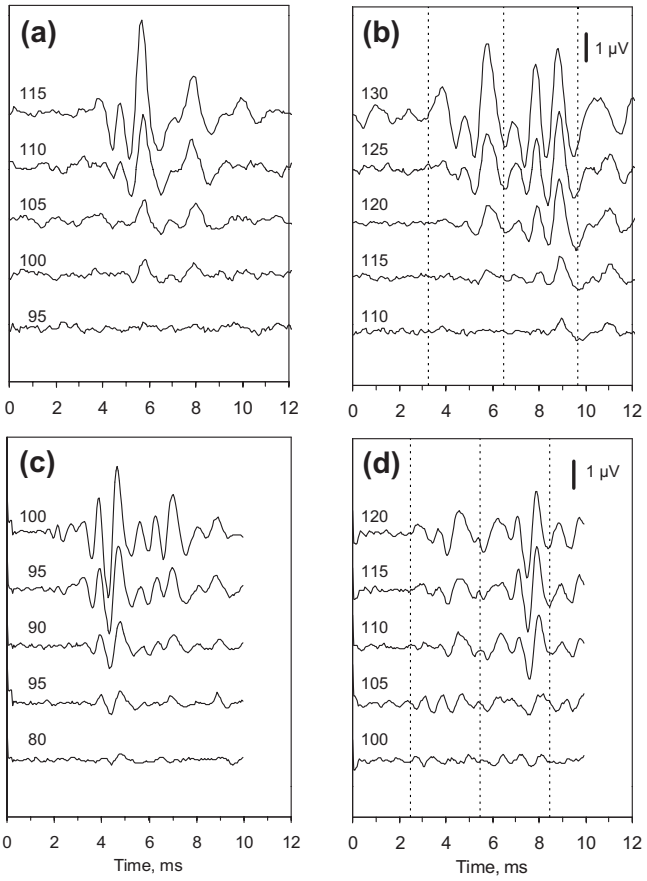


FIG. 3. AEP examples at various stimulus intensities. (a) Beluga whale, 0° azimuth sound-source position, and (b) 90° position. [(c) and (d)] The same for bottlenose dolphin. Stimulus frequency 128 kHz, intensities are specified next to the records. At (b) and (d), vertical dashed lines delimit parts of the records estimated as responses to direct sound propagation (3.3–6.5 ms in the beluga whale and 2.5–5.5 ms in bottlenose dolphin) and to wall-reflected sound (6.5–9.7 ms in beluga whale and 5.5–8.5 ms in bottlenose dolphin).

whereas the second set of waves was considered as a response to the signal reflected from the pool walls. The delay of 2.5–3 ms allowed us to differentiate these two responses and to measure them separately. For the threshold evaluation, only the first response (to the directly spreading signal) was taken into consideration.

The responses were level dependent: As the level decreased, the response amplitude decreased until the response disappeared, as can be seen in Fig. 3. Figure 4 illustrates the procedure of threshold evaluation basing on the response amplitude-vs-level dependence. The amplitude was plotted as a function of sound level (as indicated in Sec. II, only the first of the two evoked-potential complexes was taken into consideration in the records obtained at the 90° sound-source position). The near-threshold 15- to 20-dB part of the plots was approximated by a straight regression line, and this line was extrapolated to the zero-amplitude value. The sound level corresponding to the zero response amplitude was adopted as a threshold estimate.

The AEP sensitivity depended on the position of the sound source. Figure 3 exemplifies that AEPs of almost the same amplitudes were produced by the stimuli of zero sound-source azimuth of 18–30 dB lower intensities [Figs. 3(a) and 3(c)] than at the 90° azimuth [Figs. 3(b) and 3(d)].

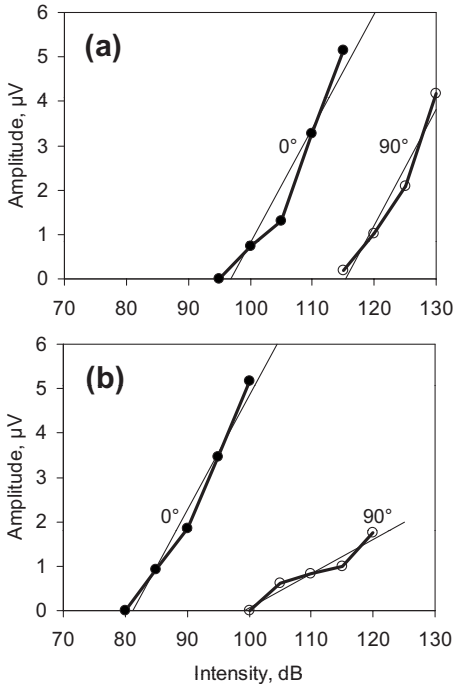


FIG. 4. AEP amplitude dependence on stimulus level. (a) Beluga whale and (b) bottlenose dolphin. Plots obtained at sound-source azimuths of 0° and 90°, as specified. Solid lines with dot symbols—experimental data; straight thin lines—regression lines. The same experiment as exemplified in Fig. 3.

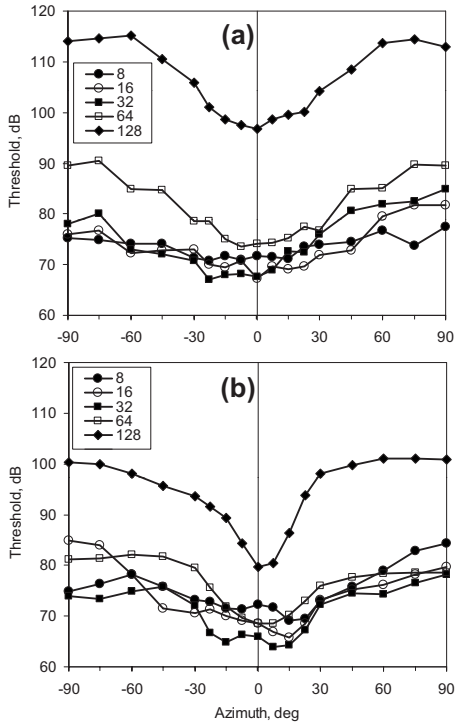


FIG. 5. Threshold-vs-azimuth functions obtained in the beluga whale (a) and bottlenose dolphin (b), keeping stimulus frequency as a parameter (specified in the legends). Negative azimuths—left side; positive azimuths—right side. Thresholds are specified in stimulus rms levels relative to 1 μPa.

Respectively, AEP thresholds were also dependent on the sound-source position. In the beluga whale, the threshold estimate at the zero (midline) azimuth was 96.9 dB, whereas at the azimuth of 90° the threshold was 115.3 dB, i.e., the threshold difference was 18.4 dB [Fig. 4(a)]. In the bottlenose dolphin, thresholds at the same azimuths were 81.2 and 99.3 dB, respectively [an 18.1 dB difference, Fig. 4(b)].

B. Threshold-vs-azimuth functions

Using the technique described above, thresholds were measured as a function of sound-source azimuth. Similar measurements were done at all frequencies from 8 to 128 kHz with octave steps, i.e., 8, 16, 32, 64, and 128 kHz. Because of the limited availability of the subjects for experimentation, only one measurement run was performed at each of the frequencies.

The results of measurements are summarized in Fig. 5 as threshold-vs-azimuth functions, keeping the stimulus frequency as a parameter. All the obtained functions featured threshold dependence on azimuth, with threshold minima near the zero azimuths. Depending on stimulus frequency, the functions were different both in terms of sensitivity (which manifests itself in Fig. 5 in the shift of the plots along the threshold axis) and in terms of the acuteness of the receiving beam (which manifests itself in Fig. 4 in different shapes of the plots).

Since the obtained functions did not feature any systematic asymmetry, the data obtained at symmetrical right and left sound-source positions were averaged, in order to decrease the data scatter. The result is presented in Fig. 6. For better comparison of the plots in terms of spatial selectivity, ignoring the frequency-dependent sensitivity, the thresholds are presented in decibel relative to the zero-azimuth threshold at the particular frequency.

In the both subjects, the general trend was that the steepness of threshold-vs-azimuth functions increased (i.e., the receiving beam became narrower) with increasing stimulus frequency from 8 to 128 kHz. However, the beluga whale showed a less steep threshold dependence on azimuth (a wider receiving beam) [Fig. 6(a)] than the bottlenose dolphin [Fig. 6(b)].

Another difference between the two subjects was that with a lowering of the frequency, the bottlenose dolphin featured a deviation of the best-sensitivity direction from the midline. Whereas the best-sensitivity direction was 0° at 128 kHz frequency, it was 7.5° at 64 kHz, and 15° at 32, 16, and 8 kHz [Fig. 6(b)]. In the beluga whale, this regularity was hardly detectable: Either the zero azimuth featured the lowest threshold, or the threshold difference between the zero and near-zero azimuths was negligible [Fig. 6(a)].

To characterize quantitatively the receiving beam acuteness as a function of stimulus frequency, the authors used the receiving beam width at two arbitrary criterion levels, 3 and 6 dB re peak. These estimates demonstrated the receiving beam width decrease with frequency increase (Fig. 7). In the beluga whale, the regression line within a range 8–128 kHz (4 octaves) passed from ±33.5° to ±14.3° at the 3-dB criterion and from ±56.9° to ±18.9° at the 6-dB criterion, which corresponded to half-width trends of −4.8 and −9.5 deg/octave, respectively. In the bottlenose dolphin, the receiving beam was much narrower: The regression lines for the 3- and 6-dB criteria passed from ±14.9° to ±6.3° and from ±33.1°

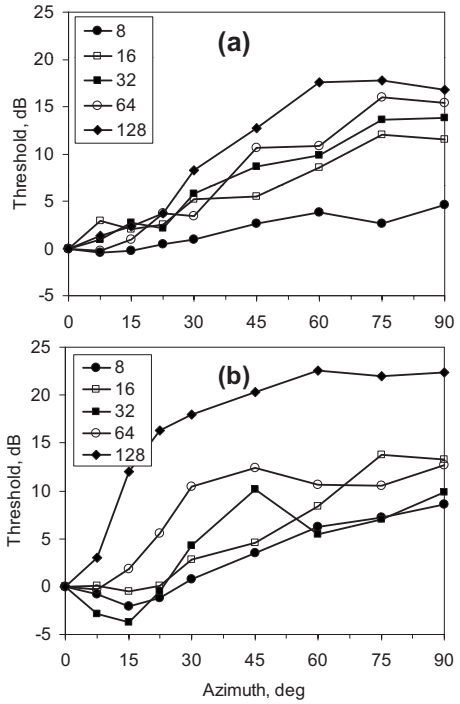


FIG. 6. Threshold-vs-azimuth functions (receiving beam patterns) of the beluga whale (a) and bottlenose dolphin (b), keeping stimulus frequency as a parameter (specified in the legends). Thresholds are specified in decibel relative to the zero-azimuth threshold at each of the frequencies.

to ±8.4°, respectively, which corresponded to half-width trends of −2.2 and −6.2 deg/octave, respectively.

C. Azimuth-dependent audiograms

At several sound-source azimuths, complete audiograms were obtained within a frequency range from 8 to 128 kHz,

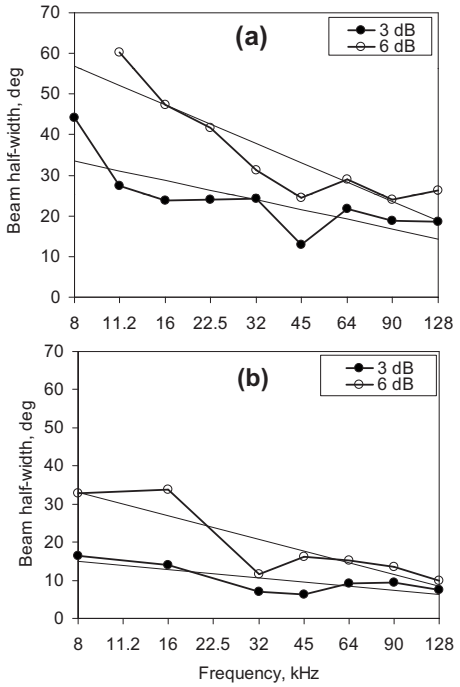


FIG. 7. Receiving beam width of the beluga whale (a) and bottlenose dolphin (b) estimated at 3- and 6-dB levels, as specified in legends. Solid line with dot symbols—estimates obtained from data presented in Fig. 5; thin straight lines—regression lines.

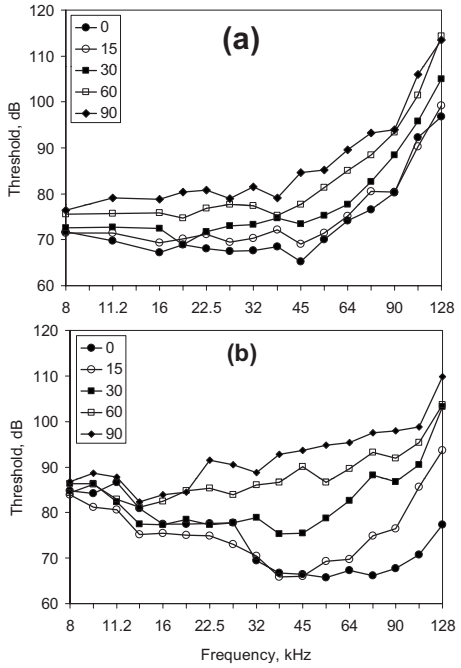


FIG. 8. Audiograms (threshold-vs-frequency functions) of the beluga whale (a) and bottlenose dolphin (b) at different sound-source azimuths, from 0° to 90°, as specified in legends.

with $\frac{1}{4}$-octave steps. Again, because of limited availability of the subjects, only one measurement run was performed at each of the azimuths. At combinations of azimuth and frequency, which were common for both the preceding and the present measurement runs, the differences of threshold estimates were within a range of ±5 dB. This difference was considered as an inevitable consequence of non-detectable differences in experimental conditions or animal states. The audiograms were obtained at both left- and right-side azimuths. Since no systematic difference was found between symmetrical left- and right-sided audiograms, they were averaged in order to decrease the data scatter. Figure 8 presents the final audiograms obtained in the two subjects at azimuths of 0°, 15°, 30°, 60°, and 90°. The 0° audiograms of both the beluga whale and bottlenose dolphin demonstrated features known for a variety of both behavioral and evoked-potential audiograms obtained in odontocetes (review Supin *et al.*, 2001). They featured the lowest thresholds at a certain intermediate frequency range (45 kHz in the beluga whale and 54 kHz in the bottlenose dolphin). Below this best-sensitivity frequency, the threshold raised slowly; above the best-sensitivity frequency, the threshold raised steeper.

With the sound source deviating laterally, the audiograms were deformed in such a way that low-frequency thresholds increased more slowly than high-frequency ones. As a result, at lateral azimuths close to 90°, the audiogram featured no mid-frequency lowest-threshold region. Instead, there was a monotonous threshold increase from lower to higher frequencies.

Qualitatively the deformation of the audiogram depending on the sound-source azimuth was similar in both the beluga whale and bottlenose dolphin. However, quantitatively it was different: It was less prominent in the beluga whale rather than in the bottlenose dolphin, as is obvious

from comparison of Figs. 8(a) and 8(b). The maximal difference between the 0° and 90° thresholds was 19.3 dB (at 45 kHz) in the beluga whale, whereas it was as large as 32.6 dB (at 128 kHz) in the bottlenose dolphin. The most prominent this inter-species difference was in the high-frequency part of the audiograms: The mean difference between 0° and 90° audiograms averaged within a range 32–128 kHz was 15.0 dB in the beluga whale and as large as 28.0 B in the bottlenose dolphin.

IV. DISCUSSION

A. Methodological restrictions

Two methodical restrictions were inevitably admitted in the present study: (i) Only one subject of each species was tested, with only one measurement performed at each of the azimuth-frequency combinations, so no statistical evaluation was available; and (ii) the acoustic field in the small-size experimental bath was far from ideal.

The limited availability of subjects is a common problem for investigations performed with dolphins and whales: Many experimental investigations in these animals have been performed in only one or two individuals. A reasonable approach to this problem is to accept, as a first-step suggestion, that the data really reflect the investigated regularities, until more available data either confirm or deny this suggestion.

The same approach may be used for the non-ideal acoustic field in the experimental tank. The limited size of the tank did not prevent sound reflections from the walls, bottom, and surface, so the sound might reach the subject from other directions than the sound-source position. However, some precautions minimized this effect.

(i) The low water column in the tank (45 cm) made the acoustic conditions approach those of a flat layer, which should minimize the effects of sound reflections from the bottom and surface.
(ii) The effect of reflections from the bottom and surface was additionally reduced by the baffles at the mid-distance between the sound source and the subject. The narrow 15-cm slit between the baffles made possible noticeable diffraction of low-frequency sounds in the vertical plane, but horizontal diffraction should be negligible at the 80-cm long slit.
(iii) Sound reflections from the tank walls were not blocked; however, the delay between the direct and reflected sounds made possible separate measurement of AEP to the directly spreading sound ignoring AEP to reflected sounds.

From these considerations, the authors assume that as a first approximation, the AEP dependence on sound-source azimuth may be assumed as reflecting the hearing directionality of the subjects.

B. Inter-subject differences

The data presented above demonstrated dependence of the receiving beam width on sound frequency both in the beluga whale and bottlenose dolphin. Respectively, the audiogram was azimuth-dependent in the both subjects. For the bottlenose dolphin, this result confirms previous observations (see Introduction), whereas for the beluga whale, this result is in conflict with a study of Klishin *et al.* (2000). The authors cannot explain why the dependence of receiving beam on frequency was not observed in that study. Either individual subject peculiarities or some features of the experimental facilities, or uncontrolled head movements, or some other factors may be suspected. Retrospectively, it is hardly possible to find an answer. In any case, the dependence of the receiving beam on frequency found herein seems more realistic since it agrees with the properties of an equivalent receiving aperture.

Some differences between the two subjects were found.

(i) The beluga whale had less acute spatial selectivity (a wider receiving beam) than the bottlenose dolphin.
(ii) The azimuth-specific audiograms in the beluga whale were less dependent on azimuth than in the bottlenose dolphin.
(iii) The bottlenose dolphin featured dependence of the best-sensitivity direction on frequency (the deviation from the midline with lowering of frequency). Assuming a symmetrical binaural receiving beam, this means the presence of two best-sensitivity axes deviated from the midline. This effect was hardly detectable in the beluga whale.

As to the two best-sensitivity axes, each deviated from the midline, their presence in the bottlenose dolphin may be explained by overlapping of receiving beams of the left and right ears, each featuring an axis deviated from the midline (Popov *et al.*, 2006). In the beluga whale, this effect might not be detected because of the lower spatial selectivity of hearing. Indeed, even in the bottlenose dolphin, this effect is characteristic of lower frequencies. At the lower frequencies, spatial selectivity is not acute, so a deviation of the axis by 7.5°–15° is rather difficult to detect. In the beluga whale, the spatial selectivity was generally less acute than in the bottlenose dolphin, so it was the least acute at low frequencies. In this situation, a small deviation of the best-sensitivity axis from the midline, even if existed, might be undetectable.

A direct consequence of lower spatial selectivity of hearing in the beluga whale than in the bottlenose dolphin is the fact that the audiogram of the beluga whale is less dependent on the sound-source azimuth than that of the bottlenose dolphin. The lowest thresholds of these audiograms (above 60 dB at the zero azimuth) were 25–30 dB higher than typical of behavioral experiments (Johnson, 1967, 1992). This difference should be a result of the short duration of the stimulus used herein: The equivalent rectangular duration of one 0.5-ms cosine envelope is 0.25 ms, whereas behavioral measurements in dolphins revealed temporal summation of hearing as long as tens of millisecond (Johnson, 1968). The same difference between the behavioral and AEP audiograms of odontocetes was observed in a number of studies using short stimuli to provoke AEPs (rev. Supin *et al.*, 2001). Taking into consideration the effect of temporal summation, a correction for about two orders of magnitude of duration, i.e., around 20 dB, must be done for comparison of the thresholds found

in the present study with behavioral data. This correction gives satisfactory agreement between the AEP and behavioral data. So the AEP-based audiograms seem to correctly reflect the hearing sensitivity. The audiogram dependence on azimuth implies that the perceived spectrum of a wide-band sound signal is dependent on the source azimuthal position, which provides additional cues for spatial discrimination of signals.

Because of limited availability of the subjects (only one individual of each of the species and only one measurement series in each subject), the authors cannot posit definitely that the differences reflect real inter-species differences, and not peculiarities of the investigated individuals. Some random data scatter manifesting itself both in difference between left- and right-side data and between measurement runs might also contribute to the inter-subject difference. However, if to accept, as a first-step suggestion, that the data really reflect the species differences, the hearing of the beluga whale species may be estimated as markedly less spatially selective than that of the bottlenose dolphin.

A question arises, whether the difference in spatial selectivity of hearing between the beluga whale and bottlenose dolphin is associated with their difference in the head and skull shape. This association, although not unambiguous, may be considered as a possible explanation. In the beluga whale, the skull is markedly shorter than in the bottlenose dolphin. With a shorter skull, a certain degree of sound shadowing by the head should appear at a larger azimuth angle than with a longer skull.

ACKNOWLEDGMENTS

The study was supported by the Russian Foundation for Basic Research (Grant No. 06-04-48518) and the Russian Ministry of Science and Education (Grant No. NSh-157.2008.4). Valuable assistance of V. Klishin, M. Tarakanov, M. Pletenko, and the staff of the Utrish Marine Station is greatly appreciated.

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

Au, W. W. L., and Moore, P. W. B. (**1984**). "Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncates*," J. Acoust. Soc. Am. **75**, 255–262.

(**2000**). *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. Fay (Springer, New York).

Dubrovskiy, N. A. (**1990**). "On the two auditory systems in dolphins," in *Sensory Abilities of Cetaceans. Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 233–254.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, New York), pp. 247–260.

Johnson, C. S. (**1968**). "Relation between absolute threshold and duration of tone pulse in the bottlenosed porpoise," J. Acoust. Soc. Am. **43**, 757–763.

Johnson, C. S. (**1992**). "Detection of tone glides by the beluga whale," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 241–247.

Kastelein, R. A., Janssen, M., Verboom, W. C., and de Haan, D. (**2005**). "Receiving beam patterns in the horizontal plane of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **118**, 1172–1179.

Klishin, V. O., Popov, V. V., and Supin, A. Ya. (**2000**). "Hearing capabilities of a beluga whale, *Delphinapterus leucas*," Aquat. Mamm. **26**, 212–228.

Popov, V. V., and Supin, A. Ya. (**1988**). "Diagram of auditory directionality in the dolphin *Tursiops truncatus* L," Dokl. Biol. Sci. **300**, 323–326.

Popov, V. V., and Supin, A. Ya. (**1992**). "Electrophysiological study of the interaural intensity difference and interaural time-delay in dolphins," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 257–267.

Popov, V. V., Supin, A. Ya., and Klishin, V. O. (**1992**). "Electrophysiological study of sound conduction in dolphins," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 269–276.

Popov, V. V., Supin, A. Ya., Klishin, V. O., and Bulgakova, T. N. (**2006**). "Monaural and binaural hearing directivity in the bottlenose dolphin: Evoked-potential study," J. Acoust. Soc. Am. **119**, 636–644.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise *Tursiops truncates*," J. Exp. Biol. **63**, 569–585.

Supin, A., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, Boston, MA).

Zaytseva, K. A., Akopian, A. I., and Morozov, V. P. (**1975**). "Noise resistance of the dolphin auditory analyzer as a function of noise detection," Biofizika **20**, 519–521.

Critical ratios in harbor porpoises (*Phocoena phocoena*) for tonal signals between 0.315 and 150 kHz in random Gaussian white noise

Ronald A. Kastelein,[a] Paul J. Wensveen, and Lean Hoek
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands

Whitlow W. L. Au
Hawaii Institute of Marine Biology, University of Hawaii, P.O. Box 1106, Kailua, Hawaii 96734

John M. Terhune
Department of Biology, The University of New Brunswick, P.O. Box 5050, Saint John, New Brunswick E2L 4L5, Canada

Christ A. F. de Jong
MON/Acoustics, TNO Science and Industry, P.O. Box 155, 2600 AD Delft, The Netherlands

(Received 23 February 2009; revised 31 May 2009; accepted 18 June 2009)

A psychoacoustic behavioral technique was used to determine the critical ratios (CRs) of two harbor porpoises for tonal signals with frequencies between 0.315 and 150 kHz, in random Gaussian white noise. The masked 50% detection hearing thresholds were measured using a "go/no-go" response paradigm and an up-down staircase psychometric method. CRs were determined at one masking noise level for each test frequency and were similar in both animals. For signals between 0.315 and 4 kHz, the CRs were relatively constant at around 18 dB. Between 4 and 150 kHz the CR increased gradually from 18 to 39 dB (~3.3 dB/octave). Generally harbor porpoises can detect tonal signals in Gaussian white noise slightly better than most odontocetes tested so far. By combining the mean CRs found in the present study with the spectrum level of the background noise levels at sea, the basic audiogram, and the directivity index, the detection threshold levels of harbor porpoises for tonal signals in various sea states can be calculated.
 [DOI: 10.1121/1.3177274]

PACS number(s): 43.80.Lb, 43.80.Nd [MCH] Pages: 1588–1597

I. INTRODUCTION

Odontocetes often swim in murky water, are active at night, and dive to depths where light hardly penetrates even during the day. Therefore these animals depend for a large degree on sound, rather than on vision, for orientation and communication (Au, 1993, 2000). In the sea, noise is produced naturally by wind, waves, geological events, and biological activities (Urick, 1983). Therefore odontocete hearing must have evolved to function in the presence of interfering noise from natural sources. However, most hearing studies on odontocetes have been conducted in relatively quiet environments in order to obtain audiograms (for a recent overview, see Southall *et al.*, 2007). These studies do not provide information about the effects of masking noise on signal detection.

Underwater noise from human activities (particularly at low frequencies; Ross, 1976) has increased during the past century due to global industrialization. Anthropogenic noises such as those from shipping vessels, oil and gas exploration and exploitation, wind generator parks, and some military sonar systems may affect odontocetes by displacing them from areas used for foraging or reproduction, or by reducing their hearing sensitivity temporarily or permanently. Noise can also compromise hearing by masking a signal (Richardson *et al.*, 1995).

Masking occurs when one sound (the noise) interferes with the detection of another sound (the signal). The degree of interference depends on the amplitudes of the two sounds and on the degree of difference between the frequencies of signal and noise: Masking is greatest when the two sounds have a similar spectrum. The lowest signal-to-noise ratio at which a subject can detect a tonal signal in a broadband masking noise is defined as the critical ratio (CR) (Fletcher, 1940; Hawkins and Stevens, 1950). The lower an animal's CR, the better its ability to detect a signal in noise. CRs are calculated as the ratio of the sound pressure level (SPL) of a just-audible tonal signal ($SPL_{threshold}$ in dB re 1 μPa, rms) and that of the masking noise at the frequency of the signal, which is calculated by using the spectrum level (N_0 in dB re 1 $\mu Pa^2/Hz$) of the masking noise (assuming the noise band is as wide as the critical band or wider). This can be expressed as $CR = SPL_{threshold} - N_0$. The CRs can be used to calculate detection threshold (DT) levels of signals under certain background noise conditions (Scharf, 1970).

As in humans and terrestrial mammals, the hearing thresholds of odontocetes increase when the background noise level increases. This has been demonstrated by using behavioral methods for bottlenose dolphins (*Tursiops trunca-*

[a]Author to whom correspondence should be addressed. Electronic mail: researchteam@zonnet.nl

tus) by Johnson (1968a), Au and Moore (1990), and Finneran *et al.* (2002); for belugas (*Delphinapterus leucas*) by Johnson *et al.* (1989), Finneran *et al.* (2002), and Erbe (2008); and for a false killer whale (*Pseudorca crassidens*) by Thomas *et al.* (1990).

The harbor porpoise (*Phocoena phocoena*) is one of the smallest cetaceans and has very acute hearing over a very wide frequency range. It is found in most coastal waters of the temperate zone of the Northern Hemisphere. In coastal waters, the ambient noise level is often higher than in the open oceans because of the sound of the surf, and because coastal waters support high densities of marine fauna. In addition, humans are relatively active in coastal waters and often produce a lot of underwater noise. Because the harbor porpoise distribution area overlaps with areas used intensively by humans, it is of interest to know how its hearing functions under various noise conditions. Knowledge of the CR of porpoises will allow the prediction of how audible anthropogenic noises are for porpoises under various ambient noise conditions. It is also of interest to discover whether porpoises' social and echolocation signals are masked under various ambient noise conditions.

Some studies on underwater masked hearing in harbor porpoises have already been carried out. Using the auditory evoked potential (AEP) technique, Popov *et al.* (2006) found that the porpoise ear responds as a constant-bandwidth filter rather than as a constant-Q filter in the high frequency region (22.5–140 kHz). Lucke *et al.* (2007) used the AEP method to measure hearing sensitivity for signals with test frequencies of 0.7–16 kHz, using offshore wind turbine noise as a masker. Kastelein and Wensveen (2008) tested a porpoise's masked hearing for 4 kHz tonal signals at two noise levels using the psychoacoustic technique. In all the above-mentioned CR studies with harbor porpoises, only part of the functional hearing range was tested and different masking noises were used. Therefore the aim of the present study was to determine the CRs of harbor porpoises for tonal signals presented in Gaussian white noise over their entire functional hearing range.

II. MATERIALS AND METHODS

A. Subjects

Two male harbor porpoises were used in this study. Both had been stranded on the Dutch coast (on the island of Texel) at the age of about 21 months and had been rehabilitated. Harbor porpoise 01 aged from 2.5 to 3 years old during data collection, his bodyweight varied between 28.1 and 31.8 kg, his body length increased from 124 to 125 cm, and his girth in front of the pectoral fins (at the ear) varied between 71 and 75.5 cm. Harbor porpoise 02 aged from 2.5 to 3 years old during data collection, his bodyweight varied between 30.8 and 34.4 kg, his body length increased from 133 to 135 cm, and his girth in front of the pectoral fins varied between 75.5 and 78.5 cm. Veterinary records showed that neither animal had been exposed to ototoxic medication. Porpoise 01 died from an encapsulated lung abscess when almost 90% of the study was completed. Therefore signals below 2 kHz were only tested with porpoise 02.

The animals each received ~1.8 kg of thawed fish (sprat, *Sprattus sprattus*, herring, *Clupea harengus*, and mackerel, *Scomber scombrus*) per day equally divided over three to four meals (three of which were given during research sessions).

B. Facility

The study was conducted at SEAMARCO's Research Institute, The Netherlands. Its location is remote and quiet, and was specifically selected for acoustic research. The animals were kept in a pool complex built for acoustic research and consisting of an outdoor pool (12 m × 8 m; 2 m deep) connected via a channel (4 m × 3 m; 1.4 m deep) to an indoor pool (8 m × 7 m; 2 m deep; Fig. 1). The study was conducted in the indoor pool. The pool walls were made of plywood covered with polyester. To reduce sound reflections in the pool (mainly above 25 kHz), the walls were covered with 3-cm-thick coconut mats with fibers embedded in 4-mm-thick rubber, and the bottom was covered with a 20-cm-thick layer of sloping sand on which aquatic vegetation grew. The coconut mats extended to 10 cm above the water level, to reduce the splashing noise of waves.

Skimmers kept the water level constant, so that sound conditions were stable. Seawater was pumped directly from the nearby Oosterschelde, a lagoon of the North Sea, into the open system; 80% recirculation through biological and sand filters ensured year-round water clarity.

The water circulation system and the aeration system for the bio-filter were made as quiet as possible. This was done by choosing "whisper" pumps, mounting the pumps on rubber mats, and connecting the pumps to the circulation pipes with very flexible hoses. The average monthly temperature varied during the year between 3 and 20 °C, and the salinity was around 3.4%. There was no current in the pool during the experiments, as all pumps were shut off 10 min before and during sessions. By the time a session started, no water flowed over the skimmers, so there was no flow noise during testing.

During the test sessions, the porpoise not being tested was asked to perform quiet behaviors in the outdoor part of the pool system (10–15 m away from the study animal). The equipment used to produce both types of sound stimuli (tonal test signals and masking noise) was housed out of sight of the study animals, in a research cabin 4 m away from the underwater listening station. The listening station was located at the end of a 3 cm diameter water-filled polyvinylchloride tube. This positioned the porpoise's external auditory meatus 2 m from the sound source, 1 m below the water surface (Fig. 1).

C. Masking noise

The masking noise was random Gaussian white noise produced by a waveform generator (Hewlett Packard, model 33120A; Fig. 2). Depending on the frequency range and transducer used, the white noise was filtered by a variable Butterworth band-pass filter (Krohn-Hite 3500), for two reasons: (1) so that high spectrum levels close to the test frequency (caused by transducer characteristics) could be elimi-

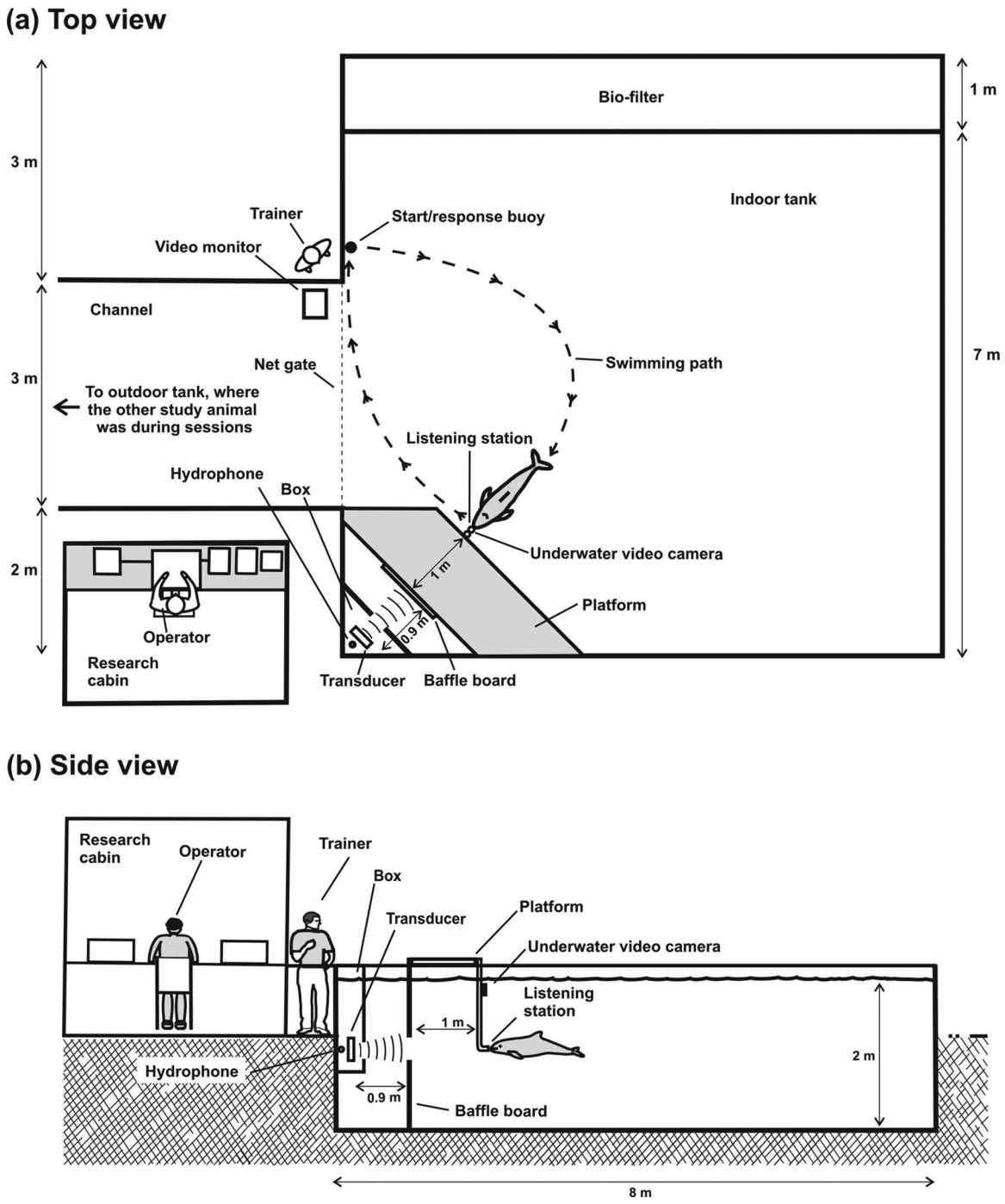


FIG. 1. The study facility, showing a harbor porpoise in position at the listening station: (a) top view, and (b) side view, both to scale. Also shown is the swimming path during a trial.

nated, and (2) to prevent the high noise levels needed to mask the lower frequencies from being uncomfortably high at higher frequencies where porpoise hearing is more sensitive (because of the U-shape of the porpoise audiogram plot; Kastelein *et al.*, 2002). Filter settings used to produce the masking noise bands are shown in Table I. For the RANA transducer (transducers are described below), the output of the filter was equalized with an audio equalizer (Behringer FBQ800) to produce an even spectrum of the noise bands over a 2/3-octave band. It was assumed that the bandwidth of the masking noise (2/3-octave) exceeded the critical bandwidth of the harbor porpoise. This assumption was based on the critical band/auditory filter shape data for humans, bottlenose dolphins, and belugas obtained by means of behavioral techniques (humans: Zwicker, 1961; Glasberg and Moore, 2000; odontocetes: Au and Moore, 1990; Lemonds *et al.*, 2000; Finneran *et al.*, 2002).

Over the frequency range tested, noise spectrum levels of the 2/3-octave bands were approximately 10–30 dB (depending on the frequency) above the 50% detection hearing thresholds of porpoises [Kastelein *et al.*, 2002; Table I, Fig. 3(a)]. In most mammals, masked thresholds for tonal signals increase linearly with the noise level (Fay, 1988). Because this also occurs in harbor porpoises at 4 kHz (Kastelein and

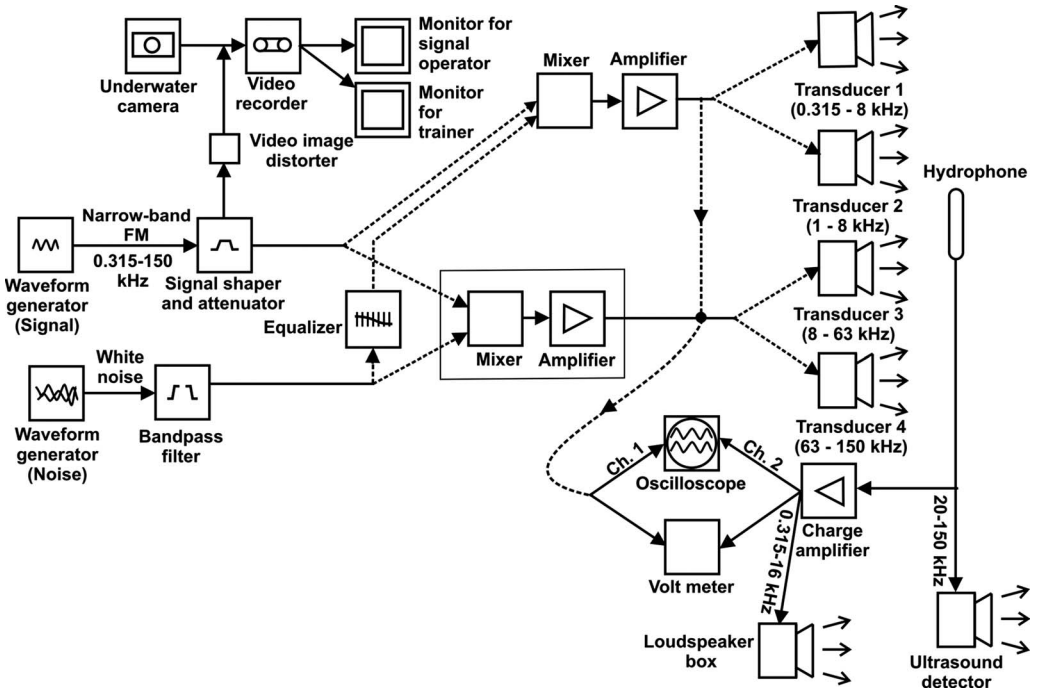


FIG. 2. Block diagram of the signal generation and listening systems.

Wensveen, 2008), masked thresholds were obtained at a single masking noise level per test frequency. The spectrum level of the masking noise used in the present study was 10–60 dB (depending on the frequency) above the background noise spectrum level in the pool [Fig. 3(a)].

D. Test signals

The masked hearing ability of porpoise 01 was quantified for 15 test signals with center frequencies between 2 and 145 kHz inclusive. The masked hearing ability of porpoise

TABLE I. The mean CRs of each of the two male harbor porpoises listening underwater to pure tones and narrowband FM tonal signals (range: 1% of center frequency), in random Gaussian white noise. Each mean is calculated from all the reversals in ten sessions. Also shown are the total number of reversal pairs in ten sessions used in the analyses and the incidences of prestimulus responses (or false alarms) over all trials (signal present and signal absent).

				Harbor porpoise 01			Harbor porpoise 02		
Center frequency (kHz)	FM range (kHz)	Masking noise frequency band (kHz)	Masking noise spectrum level (dB re 1 $\mu Pa^2/Hz$)	Mean CR ± SD (dB)	Reversal pairs	Incidence of prestimulus responses (%)	Mean CR ± SD (dB)	Reversal pairs	Incidence of prestimulus responses (%)
0.315	Pure tone	0.02–2	103	⋯	⋯	⋯	19 ± 5	72	16
0.4	Pure tone	0.02–2	103	⋯	⋯	⋯	18 ± 5	70	18
0.5	0.495–0.505	0.02–2	101	⋯	⋯	⋯	19 ± 4	88	15
1	0.99–1.01	0.02–2	89	⋯	⋯	⋯	17 ± 5	80	12
2	1.98–2.02	0.02–4	95	18 ± 4	95	6	19 ± 5	96	13
4	3.96–4.04	3.2–5	74	20 ± 4	86	12	18 ± 5	83	10
8	7.92–8.08	6.3–10	82	22 ± 4	93	7	20 ± 4	93	7
16	15.84–16.16	12.5–20	69	26 ± 4	94	10	25 ± 4	96	13
25	24.75–25.25	20–32	62	28 ± 4	90	4	27 ± 3	89	11
31.5	31.68–32.32	25–40	57	27 ± 3	91	7	28 ± 4	94	9
40	39.6–40.4	32–50	59	32 ± 4	85	13	28 ± 3	90	8
50	49.5–50.5	40–63	54	32 ± 4	95	8	29 ± 4	80	12
63	63.36–64.64	50–80	65	34 ± 4	81	16	33 ± 3	92	7
80	79.2–80.8	63–100	64	34 ± 4	89	6	32 ± 4	85	6
100	99.0–101.0	80–200	59	35 ± 4	84	2	33 ± 4	90	9
110	108.9–111.1	80–200	62	34 ± 3	103	4	32 ± 3	94	3
125	123.8–126.3	80–200	59	37 ± 4	88	10	34 ± 3	94	8
140	138.6–141.4	80–200	54	39 ± 4	98	4	39 ± 4	80	7
145	143.6–146.4	80–200	71	39 ± 4	87	7	⋯	⋯	⋯
150	148.5–151.5	80–200	69	⋯	⋯	⋯	38 ± 3	92	6

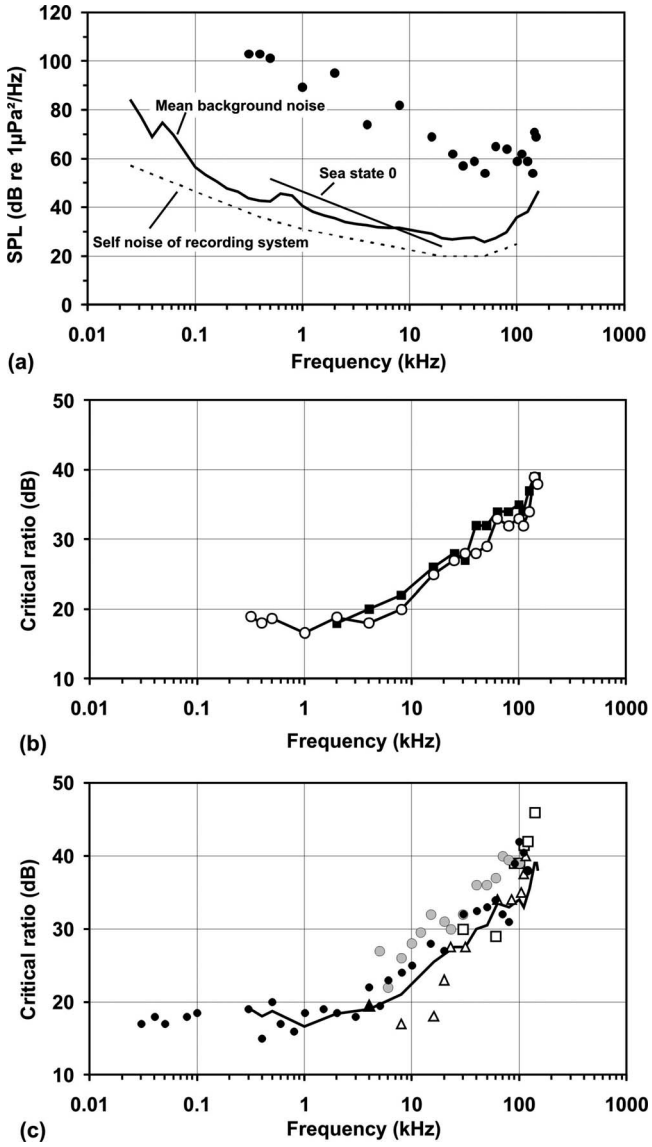


FIG. 3. (a) The mean background noise level in the pool (Leq method), sea state 0, and the masking noise levels (spectrum levels shown as ●) in which the porpoises had to detect the tonal signals. Also shown is the self-noise of the recording system. (b) The mean CRs (based on 50% DTs) of the harbor porpoises 01 (■) and 02 (○) in the present study. Each data point represents the mean of around 90 reversal pairs. (c) The mean CRs of the two harbor porpoises of the present study for the 19 tonal signals (solid line). Also shown are CRs of bottlenose dolphins (gray circles, Johnson, 1968a; □, Au and Moore, 1990), a beluga (●, Johnson *et al.*, 1989), a false killer whale (△, Thomas *et al.*, 1990), and another harbor porpoise (▲, Kastelein and Wensveen, 2008).

02 was quantified for 19 test signals with center frequencies between 0.315 and 150 kHz inclusive. The lowest test frequency for which masking noise could be generated with sufficiently high amplitude was 0.315 kHz. The highest test frequencies (145 kHz for animal 01 and 150 kHz for animal 02) were determined by the rise in the unmasked DTs of the study animals at these frequencies during preliminary tests, indicating the high cut off point of the functional hearing range.

The test signals were produced by a waveform generator (Hewlett Packard, model 33120A; Fig. 2). The 0.315 and 0.4 kHz signals were presented as pure tones. The other test signals were narrowband sinusoidal frequency-modulated (FM) tones (warbles) with center frequencies of 0.5, 1, 2, 4, 8, 16, 25, 31.5, 40, 50, 63, 80, 100, 110, 125, 140, 145, and 150 kHz. As the sonar signals of porpoises consist of narrowband signals peaking at 120–130 kHz (Møhl and Andersen, 1973; Verboom and Kastelein, 1995, 1997; Hansen *et al.*, 2008), extra frequencies in this frequency band were tested. The modulation range of the FM signals was ±1% of the center frequency (the frequency around which the signal fluctuated symmetrically), and the modulation frequency was 100 Hz (for example, if the center frequency was 10 kHz, the frequency fluctuated 100 times per second between 9.9 and 10.1 kHz). Table I shows the frequency ranges of the FM stimuli. Narrowband FM signals were used because such signals are less likely to be dominated by constructive and destructive interference effects (standing waves) in a reverberant pool than pure tones. The received levels are more constant when using narrowband FM signals than when using pure tones (Kastelein *et al.*, 2002, 2005; Finneran and Schlundt, 2007).

A modified audiometer (Madsen Electronics, Midimate 622 with extended frequency range) was used to control the duration and amplitude of the signals. Each acoustic stimulus had a duration of 1000 ms including 50 ms rise and fall times (used to prevent abrupt signal onset and offset transients). The steady state portion of the signal was thus 900 ms. Mammalian ears integrate signals over time. Therefore, signals were used that were probably long enough to yield the lowest possible threshold (based on data from bottlenose dolphins, Johnson, 1968b). The SPL of the test signals at the porpoise's head while it was at the underwater listening station was varied in 5 dB steps.

The masking noise and tonal signals between 0.315 and 8 kHz were mixed by a custom-built digital mixer (summing amplifier; AS 2008-1) and thereafter amplified with one amplifier [Bruel & Kjær (B&K) 2713] for the RANA transducer and another amplifier (Samson HQ VPA2450 MB) for the J-11 transducer. The masking noise and tonal signals between 4 and 200 kHz were mixed and amplified by a sonar system testing apparatus (IJkmonitor type SQM 03/00, MEOB Oegstgeest, stock No. 6625-17-038-3237) of which the attenuators, mixer, and amplifier were used. To test the porpoises' hearing over a wide frequency range, the following four transducers were used, each performing optimally in a specific frequency band (some frequencies were tested with two transducers).

- 0.315–8 kHz sounds were projected with a directional inductive moving coil transducer [Underwater Sound Reference Division (USRD) J-11].
- 1–8 kHz sounds were projected with a directional inductive moving coil transducer (RANA, custom built by TNO, Delft, The Netherlands, resembling a USRD J-9 transducer).
- 8–63 kHz sounds were projected by a custom-built directional underwater piezoelectric transducer [Ocean Engineering Enterprise, North Canton, OH, modified (to produce more output at the lower frequencies) model DRS-12; 30 cm diameter].

- 63–150 kHz sounds were projected by a custom-built directional transducer (WAUq7b) consisting of a disk of 1–3 composite piezoelectric materials (Material Systems Inc., Littleton, MA), with an effective radiating aperture diameter of 4.5 cm. The thickness of the piezoelectric materials was 0.64 cm. The piezoelectric element was a 6.4 cm diameter disk that was encapsulated in degassed polyurethane epoxy.

Because of the output characteristics of the transducers, they could not be used for all the test frequencies in each of the ranges mentioned above. The frequencies that produced relatively flat noise bands were selected. The transducers became available during different phases of the project, which determined to some extent for which frequencies they were used and on which animal. As an indicator of the condition of the transducers, their capacity was checked once a week with a capacity meter (SkyTronic 600.103). During the study period their capacity remained constant. All the transducers were directional sound transmitters, which cause less reflection from the sides of pools than omni-directional transmitters.

To reduce reflections from the walls, water surface, and pool floor, two further measures were taken: (1) The transducers were placed in a corner of the pool in a protective wooden box lined with rubber with an irregular surface, and (2) to prevent reflections from reaching the listening station, a baffle board was placed exactly halfway between the transducer and the animal. The board consisted of 2.4 m high, 1.2 m wide, 4-cm-thick plywood, covered with a 2–cm-thick closed cell rubber mat on the side facing the transducers. A 30 cm diameter hole was made in the board with its center at the same level as the porpoise's head at the listening station and the transducer (1 m below the water surface).

Three of the four transducers were positioned 1.9 m in front of the study animal at the listening station (Fig. 1), and the acoustic axes of the projected sound beams were carefully directed toward the position of the head of the porpoise at the listening station (Fig. 1). The cylinder-shaped RANA, and disk-shaped DRS-12 transducers were kept exactly in position by being suspended from nylon cords. A stainless steel weight was fixed to the lower part of the DRS-12 transducer to compensate for its buoyancy. The exact position of the highly directional small WAUq7b transducer was assured by attaching it to a plank of 2-cm-thick hardwood, which could slide vertically in slits, allowing the transducer to be lowered into the water before each session and removed after each session. The J-11 transducer did not fit in the sound box, and was therefore placed 1.2 m from the listening station between the box and the baffle board. It was kept in position by being suspended from three cords, just in front of the hole in the baffle board.

E. Measurements of background noise, received levels of masking noise, and test signals

The equipment used to measure the background noise in the pool consisted of a hydrophone (B&K 8101), a voltage amplifier system (TNO TPD, 0–300 kHz), and a dual spectrum analyzer system (25 Hz–160 kHz). The system was calibrated with a pistonphone (B&K 4223) and a white noise signal (25 Hz–40 kHz), which was inserted into the hydrophone preamplifier. The measurements were corrected for the frequency sensitivity of the hydrophone and the frequency response of the measurement equipment. The customized analyzer consisted of an analog-to-digital converter (Avisoft UltraSoundGate 116; 0–250 kHz) coupled to a notebook computer. The digitized recordings were analyzed by two parallel analysis systems: a fast Fourier transform narrowband analyzer (25 Hz–160 kHz) and a 1/3-octave band analyzer (25 Hz–160 kHz).

Underwater background noise levels were measured four times during the project under the same conditions as during test sessions. 1/3-octave band background noise SPLs were determined in the frequency range 25 Hz–100 kHz and converted to "spectrum level," via correction for the bandwidth (Kinsler *et al.*, 1982), expressed in dB re 1 μPa^2/Hz. Figure 3(a) shows the background noise in the pool and the self-noise levels of the hydrophone with a built-in preamplifier (taken from the B&K 8101 hydrophone specifications).

The received SPLs (dB re 1 μPa, rms) of each test signal and masking noise were measured approximately once each month at the porpoises' typical head position when the animals were at the listening station during the tests (Fig. 1). During trials, each porpoise's head position at the listening station was carefully monitored by the trainer, and was consistent to within a few centimeters. The SPL was determined as an average of a 5 or 10 s continuous test signal. The signal-to-noise ratio was high enough to allow the signal level to be determined from the measured 1/3-octave band level. Masking band noise levels were measured in 1/3-octave bands and converted to spectrum levels by the same method as the background noise. Figure 3(a) shows the mean masking noise spectrum level at each of the test frequencies.

After the study all equipment setups were reassembled and the received levels of the signals and masking noises used in the study were measured again independently by TNO. The recording and analysis equipment consisted of a hydrophone (B&K 8101) with a power supply (TPD) and a multichannel high frequency analyzer (B&K PULSE 3560 D, sample rate 524288 Hz), and a laptop computer with software (B&K PULSE, LABSHOP version 12.1). The systems were calibrated with a pistonphone (B&K 4223). The SPL values were within 2 dB of those found during the calibration sessions with the sound recording and analysis system mentioned above.

The average received SPL per frequency was calculated from all available calibration sessions per transducer (2–5 depending on the transducer). These averages were used to determine the session thresholds. The maximum received SPL variation between calibration sessions varied per frequency or noise band between 0 and 4 dB, but was generally around 2 dB. No harmonic distortions were present in the test frequencies at the SPLs used in the hearing tests.

The received SPLs were calibrated at a level 15–45 dB (depending on frequency) above the threshold levels found

in the present study. The linearity of the attenuation of the audiometer was checked several times during the study, and was accurate to within 0.1 dB.

F. Experimental procedure

The porpoises' listening environment was made as quiet as possible. Nobody was allowed to move within 15 m of the pool during sessions. A trial began with the animal to be tested at the start/response buoy and the other animal in the outdoor pool (Fig. 1). The masking noise was switched on and was produced during the entire session. The signal operator had set the frequency and the received SPL for the first trial of the session. The amplitude in the first trial of the session was about 15 dB above the DT determined during pre-tests. When the trainer gave a hand signal, the porpoise being tested swam to the listening station (Fig. 1).

To check the animals' positions at the listening station, the animals' behavior was recorded from above by means of an underwater video camera (Mariscope, model Micro), which was attached to the listening station (Fig. 1). The images were visible to the operator in the research cabin. Each porpoise was trained to place the tip of its rostrum at the station and its body axis in line with the beam of the transducer. A maximum deviation of 5° from the beam axis was accepted in all directions. Trials were aborted when the animal was not in the correct position, by knocking on the response buoy. The trial was repeated.

Pre-tests showed that the animals did not need warm-up trials before the actual session began. After the trainer had sent the porpoise to the listening station, he or she moved out of the porpoise's view and watched the porpoise's position at the listening station on a monitor.

In signal-present trials, the porpoise stationed, then had to wait for a period of random duration between 6 and 12 s (established via a random number generator), before the signal operator produced the test signal. If the animal detected the sound, it was trained to leave the station ("go" response) at any time during the transmission of the signal and return to the start/response buoy [Fig. 1(a)]. When the test signal was produced by the audiometer, a generator was activated electronically that produced horizontal white lines on the video image. This helped the operator to determine visually whether or not the porpoise responded during production of the signal. If he responded, the signal operator told the trainer that the response was correct, after which the trainer gave the porpoise a fish reward. If the animal did not respond to the sound ("no-go" response) the signal operator would tell the trainer. The trainer then signaled to the animal (by tapping three times on the side of the pool) that the trial had ended, thus calling him back to the start/response buoy. No reward was given. If the animal responded before a signal was produced (a prestimulus response or false alarm), the signal operator would tell the trainer to ignore the animal for about 10 s. After this short time, a new trial was started by calling the animal to the start buoy and sending it to the listening station. When a prestimulus response was clearly triggered by an external sound, which was also detected by the operator, data from the trial were not used. In such cases the trial was immediately repeated after the external sound had stopped.

In signal-absent (control or catch) trials, the porpoise stationed, then the signal operator told the trainer after a time period of random duration (but between 6 and 12 s) to end the trial by blowing on a whistle. The animal returned to the start/response buoy and received a fish reward. If the porpoise left the station before the whistle was blown during signal-absent trials, this was regarded as a prestimulus response. The same amount of fish was given as a reward for correct responses in signal-present and signal-absent trials. After a correct response trial, the next trial would start immediately after the reward was ingested.

One signal frequency was used in each session. A simple up/down staircase psychometric technique was used (Robinson and Watkins, 1973). If the animal heard a signal and responded to it (a hit), the next signal presented was 5 dB lower. If the animal did not hear a signal and remained at the station (a miss), the next signal offered was 5 dB higher. Prestimulus responses did not result in a change in signal amplitude. A session usually consisted of ~20 trials and lasted for about 15 min per animal. Each session consisted of 2/3 signal-present and 1/3 signal-absent trials offered in random order. There were never more than three consecutive signal-present or signal-absent trials. In order to end with a positive event, the last trial was always one in which the animal responded correctly and received a reward. This methodology kept the study animals motivated throughout each session (they received a reward after approximately 75% of the trials). Each session, one of four data collection sheets with different random number series, was used (to determine the random time between the animal stationing and signal presentation and the random order of signal-present and signal-absent trials). Each porpoise had its own set of four data collection sheets. The trainer never knew beforehand whether a trial was a signal-present trial or a signal-absent trial. When the porpoise left the station, the operator observed the animal's behavior on a monitor in the cabin [Fig. 1(a)], told the trainer whether or not to reward the porpoise, and recorded the animal's responses. Other behaviors were solicited from the porpoises between trials to occupy them during occasional short periods of visible or audible disturbances outside the building. Behavioral maintenance sessions, in which no data were collected, were conducted occasionally in order to correct the porpoises' position at the listening station.

Thresholds were determined for 15 or 19 test frequencies (depending on the animal). Each day, the porpoises' hearing was tested for one frequency. To prevent the learning process from affecting the results, the test frequency was varied from day to day, and adjacent frequencies were usually tested on successive days (going from the lowest to the highest frequency of the test frequency range of each transducer and *vice versa*). This way the difference in frequency tested on successive days was limited, reducing the potential need for the study animals to adapt to a new frequency.

Before each session, the voltage output of the emitting system to the transducer and the voltage output of the sound

receiving system were checked. If they were same as during the SPL calibration sessions, a test session could begin (Fig. 2). Also, the background noise level was checked to make sure it was not too high for testing.

Two to three experimental sessions per day were conducted on 5 days/week (at 0830, 1130, and 1615 h). Data were collected between November 2007 and December 2008.

G. Analysis

A switch from a test signal amplitude that the porpoise responded to (a hit), to an amplitude that it did not respond to (a miss), and *vice versa* is called a reversal. The mean 50% DT for a test signal was determined by taking the mean of all reversal pairs in the 10 sessions for that test frequency (70–103 reversal pairs per test signal depending on the frequency, average ~90 reversal pairs). Occasional sessions with more than 20% prestimulus responses were eliminated. These sessions were usually simultaneous with external noise caused by distant activities (aircraft flying, construction, and doors of the nearby lock opening and closing) or general restless behavior of an animal during that particular part of the day. In total around 6800 trials were used in the analysis.

III. RESULTS

The mean CRs of each of the two porpoises are shown in Table I and Fig. 3(b). For the 14 signal frequencies that were tested with both animals, the CRs were similar (mean difference = 1.5 ± 1.45 dB, range 0–4 dB).

Between 0.315 and 4 kHz the CRs were relatively constant [mean 18.3 dB (re 1 Hz), range 17–20 dB]. Between 4 and 150 kHz the CR increased gradually from 18 to 39 dB (~3.3 dB/octave). From 4 to 150 kHz, the mean CRs (of the two porpoises) can best be described by the equation

$$\mathrm{CR} = 10.7 + 12.1 \log_{10} f, \tag{1}$$

where f is the signal frequency in kHz ($r^2 = 0.963$, $t = 18.12$, $N = 28$, and $P < 0.000\,01$). The high r^2 value suggests that the CR varies with the common log of frequency.

The incidence of prestimulus responses varied between 2% and 16% in animal 01 and between 3% and 18% in animal 02, depending on the test frequency (Table I).

IV. DISCUSSION

A. Evaluation of the data

Because both porpoises were tested during the same sessions, any differences between the thresholds obtained for the two animals must have been due to differences in their hearing sensitivity and/or individual differences in their response criteria, motivational state, or behavior. Differences could not have been caused by differences in equipment, equipment settings, methodology, personnel, or background noise. For the frequencies that were tested on both animals, the CRs were within 4 dB of one another. Therefore the CRs found in the present study are probably representative for young harbor porpoises.

The masking noise was produced during the entire session (not only during trials). This meant that the test situation resembled what harbor porpoises would experience in the wild under conditions of background noise.

In nearly all mammals, CR increases as frequency increases, except at very low frequencies (Fay, 1988). The present study shows that harbor porpoise hearing follows the general mammalian pattern.

B. Comparison with other odontocete masking studies

Kastelein and Wensveen (2008) tested a harbor porpoise's masked hearing for 4 kHz tonal signals at two noise levels using the psychoacoustic technique. The subject's CR at 4 kHz was estimated to be between 18 and 21 dB. This CR range agrees well with the findings of the present study, which yielded CRs of 18 and 20 dB at 4 kHz.

Overall, the pattern of variation of CRs with frequency is consistent for all odontocetes studied to date (Johnson, 1968a; Johnson *et al.*, 1989; Au and Moore, 1990; Erbe, 2008). The average CRs of the two porpoises in the present study were generally slightly lower than those found in most other odontocetes tested so far [Fig. 3(c)].

Popov *et al.* (2006) obtained high frequency (22.5–140 kHz) tuning curves for three porpoises by using the AEP technique, and provided evidence for the existence of constant-bandwidth filters, rather than constant-Q filters, in the porpoise auditory system. In addition, Ketten (2000) found an "acoustic fovea" in the porpoise cochlea around 110 kHz: a sign of increased hearing sensitivity and acute frequency resolution. These findings can be interpreted as conflicting with the results of the present study (because increasing CRs are associated with constant-Q auditory filters). However, although CRs are often used to estimate the width of critical bands, CRs are probably better described as predictors of processing efficiency, than of filter width, because other signal-detection processes, which follow the cochlear filter, are involved (Patterson *et al.*, 1982). An increase in frequency selectivity (e.g., as a result of acoustic foveae) does not necessarily lead to a dramatic decrease in CR. The CR metric is very sensitive to variation in masking noise and DT levels, and the average threshold increase per increase in frequency is small (~3.3 dB/octave between 4 and 150 kHz). As such, several "plateaus" can be seen in Fig. 3(b), which could be interpreted as constant-bandwidth frequency regions (e.g., CR = 32–35 dB between 63 and 110 kHz in both animals), or as results of data variability.

C. Ecological significance

Assuming that the masking noise is continuous, the DT of a harbor porpoise for a tonal signal in ambient noise can be calculated as follows with the commonly used sonar equation:

$$\mathrm{DT} = N_{ss} + \mathrm{CR} - \mathrm{DI}, \tag{2}$$

where N_{ss} is the noise spectral density of the ambient noise in dB re 1 $\mu\mathrm{Pa}^2/\mathrm{Hz}$, CR is the critical ratio in dB re 1 Hz, and DI is the received directivity index in dB. As an example, the

authors take an estimated wind-generated ambient noise spectrum based on the Knudsen curves: $N_{ss}=56+19\log_{10}(\mathrm{ss})-17\log_{10} f$, where ss stands for sea state and f is the center frequency of the noise in kHz (Urban, 2002). The DI is known for harbor porpoises only for 16, 64, and 100 kHz from receiving beam patterns in the horizontal plane (Kastelein *et al.*, 2005). In noise levels characteristic of sea state 0, the DTs are close to the harbor porpoise basic hearing thresholds for the three frequencies. For noise levels at sea states above sea state 0, the DTs for these three frequencies are above the basic hearing thresholds.

Three caveats should be borne in mind when calculating DTs with Eq. (2). First, the Knudsen curves for deep-water sea-state noise are probably unrealistic for the shallow water environments in which harbor porpoises live. Second, the CRs in the present study were measured using random Gaussian white noise. However, in reality, background noise levels fluctuate temporally: The acoustic environment consists of noise where the energy across frequency regions is coherently modulated in time. Branstetter and Finneran (2008) showed that bottlenose dolphins have lower masked DTs in temporally fluctuating comodulated noise than in Gaussian white noise with the same spectrum level. This means that, in natural background noise, harbor porpoises can probably hear signals at slightly greater distances than would be calculated using the CRs obtained in the present study. When estimating the audibility of tones in the presence of broadband background noise, the fine structure of the temporal variation patterns of the noise needs to be known to determine if it is Gaussian (continuous) or comodulated (variable). In the case of variable noise levels, it may be necessary to determine the lowest levels that are likely to occur over a period of time equivalent to the integration time of the listener (see Erbe, 2008). Third, it is important to remember that by definition the DTs are for sound levels that are only detected 50% of the time by an animal actively listening for a specific tone.

In their natural environment, harbor porpoises are subjected to various noise levels and spectra. The natural causes of increased noise levels underwater are increased rain and wind (see Wenz curves for sea states in Urick, 1983). Once ambient noise has reached the masking level for a signal of a particular level, the DT for this signal is determined by the ambient noise, and every dB increase in ambient noise means an increase of 1 dB in the signal DT level.

ACKNOWLEDGMENTS

The authors thank students Remona Kerssies, Krista Krijger, Tess van der Drift, Alejandra Vargas, Janna Loot, Eline Berrevoets, and Niek Boonman; and volunteers Menno van den Berg, Jesse Dijkhuizen, Cathy Philipse, Saskia Roose, Joke de Lange, Wijnand de Wolf, Mary Mullaney, Sandra Wisse, and Jaya de Jonge for their help with training and data collection. The authors thank Rob Triesscheijn for making the figures, Bert Meijering, director of sea bait farm Topsy Baits, for providing space for SEAMARCO's Research Institute, and Hein Hermans for technical support at the Research Institute. The authors thank Willem Verboom (JunoBioacoustics) and Erwin Jansen (TNO-Delft) for the acoustic calibration measurements and TNO-Delft for providing the J-11 transducer for this study. The authors thank Dick de Haan (Wageningen IMARES) for technical assistance and Veenhuis Medical Audio (Marco Veenhuis and Herman Walstra) for donating and modifying the audiometer. The authors thank Teun Tollenaar (T.T. Electronics, Harderwijk) and Menno van den Berg for the construction of the sound generator used during the training phase of the project, and Arie Smink for the construction of the mixer. The authors thank the Royal Netherlands Navy for lending the IJkmonitor for the study, and Hans Heiligenberg and Raymond van Elst (Royal Netherlands Navy) for their help in providing the RANA LF transducer. They also thank Nancy Jennings (dotmoth.co.uk, Bristol, UK) and two anonymous reviewers for their valuable constructive comments on the design of the study and on this manuscript. Funding for this project was obtained from The Netherlands Ministry of Defense (Contract No. WO Bruinvissen 235-06-0003-01). The authors thank Erik van Arkel, Vincent Gales, Ronald de Rooij (DRMV, Netherlands Ministry of Defense), and René Dekeling and Frans Jansen (Royal Netherlands Navy) for their guidance during the project. The training and testing of the porpoises were conducted under authorization of the Netherlands Ministry of Agriculture, Nature and Food Quality, Department of Nature Management, with Endangered Species Permit No. FF/75A/2005/047. The authors thank Seppe Raaphorst and Jan van Spaandonk (Ministry of Agriculture, Nature and Food Quality of the Netherlands) for their assistance in making the harbor porpoises available for this project.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).
Au, W. W. L. (**2000**). "Hearing in whales and dolphins: An overview," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 1–42.
Au, W. W. L., and Moore, P. W. B. (**1990**). "Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **88**, 1635–1638.
Branstetter, B. K., and Finneran, J. J. (**2008**). "Comodulation masking release in bottlenose dolphins (*Tursiops truncatus*)," J. Acoust. Soc. Am. **124**, 625–633.
Erbe, C. (**2008**). "Critical ratios of beluga whales (*Delphinapterus leucas*) and masked signal duration," J. Acoust. Soc. Am. **124**, 2216–2223.
Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Data Book* (Hill-Fay Associates, Winnetka, IL).
Finneran, J. J., Schlund, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). "Auditory filter shapes for the bottlenose dolphin (*Tursiops truncatus*) and the white whale (*Delphinapterus leucas*) derived with notched noise," J. Acoust. Soc. Am. **112**, 322–328.
Finneran, J. J., and Schlundt, C. E. (**2007**). "Underwater sound pressure variation and bottlenose dolphin (*Tursiops truncatus*) hearing thresholds in a small pool," J. Acoust. Soc. Am. **122**, 606–614.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
Glasberg, B. R., and Moore, B. C. J. (**2000**). "Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise," J. Acoust. Soc. Am. **108**, 2318–2328.
Hansen, M., Wahlberg, M., and Madsen, P. T. (**2008**). "Low-frequency components in harbor porpoise (*Phocoena phocoena*) clicks: Communication signal, by-products, or artifacts?," J. Acoust. Soc. Am. **124**, 4059–4068.
Hawkins, J. H., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Johnson, C. S. (**1968a**). "Masked tonal thresholds in the bottlenosed porpoise," J. Acoust. Soc. Am. **44**, 965–967.
Johnson, C. S. (**1968b**). "Relation between absolute threshold and duration-

of-tone pulses in the bottlenosed dolphin," J. Acoust. Soc. Am. **43**, 757–763.
Johnson, C. S., McManus, M. W., and Skaar, D. (**1989**). "Masked tonal thresholds in the beluga whale," J. Acoust. Soc. Am. **85**, 2651–2654.
Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D. (**2002**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.
Kastelein, R. A., Janssen, J., Verboom, W. C., and de Haan, D. (**2005**). "Receiving beam patterns in the horizontal plane of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **118**, 1172–1179.
Kastelein, R. A., and Wensveen, P. J. (**2008**). "Effect of two levels of masking noise on the hearing threshold of a harbor porpoise (*Phocoena phocoena*) for a 4.0 kHz signal," Aquat. Mamm. **34**, 420–425.
Ketten, D. R. (**2000**). "Cetacean ears," in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. R. Fay (Springer, New York), pp. 43–108.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York).
Lemonds, D. W., Au, W. W. L., Nachtigall, P. E., and Roitblat, H. L. (**2000**). "High-frequency auditory filter shapes in an Atlantic bottlenose dolphin," J. Acoust. Soc. Am. **108**, 2614.
Lucke, K., Lepper, P. A., Hoeve, B., Everaarts, E., van Elk, N., and Siebert, U. (**2007**). "Perception of low-frequency acoustic signals by a harbour porpoise (*Phocoena phocoena*) in the presence of simulated offshore wind turbine noise," Aquat. Mamm. **33**, 55–68.
Møhl, B., and Andersen, S. (**1973**). "Echolocation: High-frequency component in the click of the harbor porpoise (*Phocoena ph*. L.)," J. Acoust. Soc. Am. **54**, 1368–1372.
Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.
Popov, V. V., Supin, A. Ya., Wang, D., and Wang, K. (**2006**). "Nonconstant quality of auditory filters in the porpoises *Phocoena phocoena* and *Neophocaena phocaenoides* (Cetacea, Phocoenidae)," J. Acoust. Soc. Am. **119**, 3173–3180.
Richardson, W. J., Greene, C. R., Malme, C. I., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, San Diego, CA).
Robinson, D. E., and Watkins, C. S. (**1973**). "Psychophysical methods in modern psychoacoustics," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. **2**, pp. 99–131.
Ross, D. (**1976**). *Mechanics of Underwater Noise* (Pergamon, New York).
Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, San Diego, CA), pp. 159–202.
Southall, B. L., Bowles, A. E., Ellison, W. T., Finneran, J. J., Gentry, R. L., Greene, C. R., Jr., Kastak, D., Ketten, D. R., Miller, J. H., Nachtigall, P. E., Richardson, W. J., Thomas, J. A., and Tyack, P. L. (**2007**). "Marine mammal noise exposure criteria: Initial scientific recommendations," Aquat. Mamm. **33**, 411–521.
Thomas, J. A., Pawloski, J. L., and Au, W. W. L. (**1990**). "Masked hearing abilities in a false killer whale (*Pseudorca crassidens*)," in *Sensory Abilities of Cetaceans: Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 395–404.
Urban, H. G. (**2002**). *Handbook of Underwater Acoustic Engineering* (STN Atlas, Bremen).
Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).
Verboom, W. C., and Kastelein, R. A. (**1995**). "Acoustic signals by harbour porpoises (*Phocoena phocoena*)," in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 1–39.
Verboom, W. C., and Kastelein, R. A. (**1997**). "Structure of click train signals of harbour porpoises (*Phocoena phocoena*)," in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 343–362.
Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.

Hydroacoustic measurements of the behavioral response of arctic riverine fishes to seismic airguns

John K. Jorgenson[a] and Eric C. Gyselman
Fisheries and Oceans Canada, 501 University Crescent, Winnipeg, Manitoba R3T 2N6, Canada

(Received 22 May 2008; revised 7 April 2009; accepted 18 June 2009)

Seismic surveys for hydrocarbon exploration in the Mackenzie River involve the use of airguns. Airguns produce a repetitive, intense, low-frequency sound that has the potential to cause physiological damage and behavioral changes in fishes. Some of these impacts have been documented in marine environments but few studies have been conducted in freshwater systems where the confining nature of the environment produces a different acoustic regime and could constrain possible fish response. In the current study, hydroacoustic surveys are conducted in the presence of airgun firing in the Mackenzie River to determine if fish behavior can mitigate or enhance the potential impact of this sound. It is shown that fish behavioral characteristics measured in this study are generally not changed by the presence of airgun noise. The most likely mechanism to facilitate a severe physiological effect in fishes from a mobile airgun firing is a herding response in front of the airgun, resulting in prolonged exposure to the noise. Analysis of tracked fish directional movement does not indicate that herding behavior occurs. Consequently, no evidence is found to indicate that fishes in this study would sustain severe physiological damage from this airgun seismic survey. [DOI: 10.1121/1.3177276]

PACS number(s): 43.80.Nd, 43.30.Sf [MCH] Pages: 1598–1606

I. INTRODUCTION

Steadily increasing oil and gas prices and market demands over the past few years have renewed interest in hydrocarbon reserves in Canada's Arctic. These reserves are estimated at 7×10^9 barrels of oil and 70×10^{12} ft^3 of gas, 80%–90% of which are in the Mackenzie River basin in the Northwest Territories (National Energy Board of Canada, 2004). The result has been the initiation of a number of major exploration programs since 2000 (Cott *et al.*, 2003). Exploration companies have proposed using water-based surveys on the Mackenzie River and its large tributaries using airguns as an energy source (IMG-Golder Corp., 2002). In this method, an air compressor produces steadily increasing air pressure in the airgun until a threshold is reached. The pressure is then rapidly released, producing an expanding air bubble that then collapses under the pressure of the surrounding water. This causes a sharp concussion termed a "shot" in the lexicon of the industry. The peak sound levels of individual airguns can be as high as 230 dB (re 1 μPa at 1 m) (Popper *et al.*, 2005). Individual airguns are typically grouped into an array, suspended from floats, and towed behind a moving vessel. During the survey, the airguns are fired at regular intervals. The sound waves are directed downwards and reflect off geologic formations. These echoes are measured by hydrophones also towed behind the vessel. Characteristics of the returning signal act as a predictor of the presence of oil or gas below the survey area. This type of survey is common in the offshore marine environment but it has not been commonly used in rivers.

Fisheries and Oceans Canada, tasked with providing information to the regulatory agency licensing these surveys, has concerns about the potential impacts of airgun generated sound on aquatic life in the confined environment created by a river channel. These impacts could range from behavioral changes that could interrupt or alter migration patterns of fish stocks to temporary or permanent hearing loss and tissue damage in fishes along the survey route. A number of studies have examined the potential impacts to marine mammals and fishes in the marine environment, a review of which is presented in Worcester, 2006, but very little information is available for freshwater and none on Arctic waters. Fisheries and Oceans Canada requested one proponent of an airgun survey along the Mackenzie River to carry out studies to verify predictions in their Environmental Impact Statement. The resultant 2002 study (IMG-Golder Corp., 2002) suggested that fishes respond to the seismic sound but the authors felt that there were further questions that needed to be answered.

This project is composed of two distinct studies: (1) a physiological component that looked at short and long-term impacts of seismic sound on hearing structures in species found in the Mackenzie River and (2) this study, which examines the *in-situ* behavioral response of fishes exposed to sound produced by seismic airgun firing.

The physiological study is described in Popper *et al.*, 2005. Their conclusion was that no permanent hearing threshold shift (a measure of hearing damage) occurred for any species tested; broad whitefish (*Coregonus nasus*), northern pike (*Esox lucius*), and lake chub (*Couesius plumbeus*), but a temporary shift was observed for northern pike and lake chub. In both species, recovery occurred within 24 h. While permanent physiological impacts did not appear to be a concern for the species tested, behavioral responses could result in changes to movements and migrations that may have a detrimental effect on stocks, many of which are important subsistence species harvested in the area. For ex-

[a]Author to whom correspondence should be addressed. Electronic mail: john.jorgenson@dfo-mpo.gc.ca

ample, large numbers of juvenile coregonids migrate down the Mackenzie River each summer (Reist and Bond, 1988). Seismic surveys may have a detrimental behavioral impact on these fishes.

The objectives of this study are to determine whether fishes exhibit any predictable behavioral response to airgun noise and, if they do respond, how? Do fishes exhibit avoidance behavior, which mitigates the potential physiological impacts of the sound? Do fishes exhibit "herding" behavior in front of the seismic vessel that could result in an increase in their total noise exposure and more severe physiological effects than reported in the physiological study, such as permanent hearing loss? At what sound levels do fishes respond to the sound? Can fishes locate the direction of the sound and respond to it?

All of these questions are viewed as important in determining the potential impact of seismic sound on fishes in the Mackenzie River. The results can assist in decision-making relating to environmental impact assessments and regulatory requirements for airgun seismic surveys in northern rivers.

II. METHODS

A. Overview

Direct measurement of the movements of individual fish is necessary to accomplish the objective of this study. This is done using a split-beam digital acoustic system, which enables individual fish to be detected and their location within the acoustic beam to be precisely located in three-dimensional space. Sound is output in short bursts called pings and the relative size of the fish is determined by measuring the intensity or target strength (TS) of its echo. Fish in successive pings are identified and linked to their position in previous pings using postprocessing software. The result is a three-dimensional fish "track" vector, which has both a vertical and horizontal heading and speed. The cumulative effect of the approaching airgun on fish swimming behavior cannot be directly observed as individual fish are only in the acoustic beam for relatively short periods of time. Instead, the authors use statistical comparisons of the pre-treatment control measurements to treatment (exposed to sound from seismic airguns) measurements, allowing the determination of any significant difference in the track vector components.

The model for this study predicts that if a fish responds to the seismic sound, they could respond in one of two ways. If they could not determine the direction of the sound source, they would show a generalized reaction by (1) increasing velocity, (2) decreasing tortuosity (a measure of the linearity of the track), and (3) increasing the downward vertical component of the swimming vector. A downward or diving response to anthropogenic sound has been reported in a number of studies (e.g., Handegard and Tjostheim, 2005). If a fish could determine the direction of the sound source, they would respond in the same way but they would show a change in horizontal direction away from the sound source as well.

The study is divided into two experiments. Experiment 1 is designed to measure the behavioral response of fishes to the approach and passing of the seismic barge relative to the stationary acoustic launch. It provides the best opportunity to measure the potential of a herding response. Experiment 2 examines the behavior of individual fish before and after exposure to a seismic airgun shot. The design is intended to measure an immediate startle response.

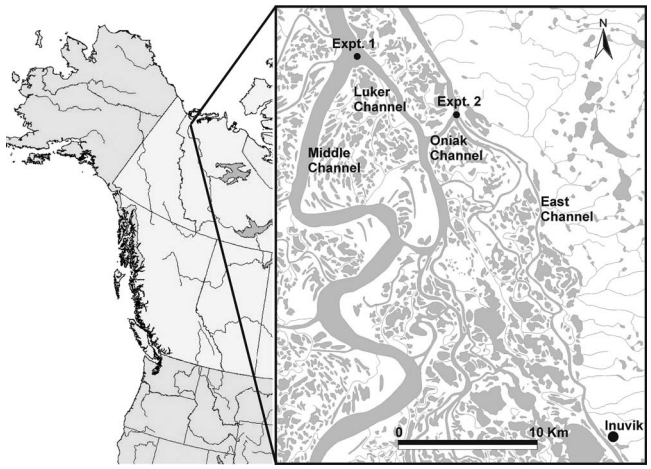


FIG. 1. Location of study site. Mackenzie River Delta, Northwest Territories, Canada.

B. Study site

The study was conducted in late July/early August, 2004, at two adjacent locations on the Mackenzie River, Northwest Territories (Fig. 1): Experiment 1 was conducted at the junction of Middle and Luker Channels (68.60° N, 134.15° W) at a depth of 45 m, and experiment 2 at the junction of East and Oniak Channels (68.56° N, 133.98° W) at depths ranging from 10 to 30 m. Both sites were chosen for their abundance of fishes, absence of a strong or variable surface current (<0.5 m s^{-1}), and relative proximity to the base of operations in Inuvik. Also, the sites satisfied a number of requirements to optimize the use of the acoustic system, including a low ambient noise level (<-137 dB re 1 W as measured by the acoustic system at the transducer face), relatively low debris levels (false targets), and deep enough for an adequate acoustic beam size. At both sites, bottom surface sediments were fine clays, typical of large river deltas.

C. Acoustic launch

The hydroacoustic system installed on the acoustic survey launch (M/V *Wood*) was a Simrad EK60 120 kHz split-beam scientific echosounder running ER60 ver. 2.1.1 software. A Simrad ES120-7C transducer with a 7° beamwidth was hull-mounted in the launch and aimed vertically.

A Novatel RT20 global positioning system (GPS) receiver was connected to the Simrad EK60 on the M/V Wood and an E-Ping DGPS receiver was used to receive differential correction information from the Canada-Wide Differential GPS Correction Service (CDGPS) on the M-SAT satellite system. This equipment provides an absolute position error of less than 2 m and a fix-to-fix position error of less than 0.5 m, a level of precision essential for this study. The refresh rate for the GPS system was 2 Hz. A Garmin Model 76

handheld GPS was used to collect position information on the seismic barge. Wide area augmentation system correction was not available at the latitude of this study but the Garmin unit advertised a position error of less than 7 m, sufficient resolution for the experiments.

D. Seismic airgun array

The seismic sound source consisted of a barge-mounted single air compressor supplying a towed clustered array of eight SGI and SGII type airguns equally spaced 70 cm apart in a 2×4 arrangement and suspended from floats at a depth of 1.8 m. The total volume of the array was 12 000 cc and the dimension was 2.6 m in length and 1.22 m across. Firing was manual, at approximately 13.1 kPa chamber pressure, leading to small variations in the firing pressure (Popper *et al.*, 2005). The barge was towed by a tugboat. Details of the array, its sound spectrum, intensity, and particle velocity are presented in MacGillivray *et al.*, 2004.

E. Experimental design

Experiment 1 was carried out in Middle Channel at the junction of Luker Channel (Fig. 1). The M/V Wood was anchored on the east side of Middle Channel. This provided for a long straight approach and departure for the seismic barge both upstream and downstream, minimizing complicating river bed structure that might affect the seismic sound path. Control data were collected on July 27, July 28, and early August 1 in the absence of the seismic barge in order to measure the natural variability of fish movement at the site. On August 1, the seismic barge arrived on site and the authors carried out experiment 1. The seismic barge began 2 km downstream and approached the anchored M/V Wood at 2 m s^{-1}. The airguns were fired every 45 s. The barge passed within 10 m of the acoustic launch and then steamed upstream for 2 km, continuing to fire the airguns at 45 s intervals. The barge then turned around and approached the anchored M/V Wood from upstream at the same speed and the same firing rate, again passing the launch and continuing downstream for 2 km. The approach from downstream and then from upstream is designed to evaluate the difference in behavior caused by the direction of travel of the seismic barge relative to the river current. This cycle was repeated to create a statistical replicate. Further cycles had to be canceled because of high winds and waves that threatened the seismic barge. The position of the acoustic launch and the seismic barge were collected using onboard GPSs to allow calculation of the relative position of the two vessels and the speed of the seismic barge. The relative position and heading of the two vessels were verified using the radar unit on the M/V Wood. Background noise levels and total noise levels during the seismic shots were collected by JASCO Research Ltd. during the control period and experiment 1 seismic trials. The methodology used and the results are presented in detail in MacGillivray *et al.*, 2004.

Experiment 2 was carried out on August 2 at the junction of Oniak and East Channels (Fig. 1). The seismic barge was anchored at the downstream side of the study area. The M/V Wood was allowed to drift by current and wind toward the seismic barge. All mechanized equipment on the acoustic launch was turned off to minimize extraneous noise. When an individual fish was recorded on the echogram screen aboard the M/V Wood, the crew requested the seismic airgun operator to fire the array. The track of the fish was therefore recorded before and after the shot, representing a control and experimental treatment, respectively. As the acoustic launch drifted toward the seismic barge, the sound intensity increased relative to observed fish as the distance between the two vessels was reduced. Once the M/V Wood came as close as possible to the barge, the engine was started and the launch was piloted slowly back to the starting point, thus completing a cycle. A total of 13 drift cycles were completed. During the first six drift cycles the airgun array was not fired. This allowed the authors to acoustically record the natural movements and behavior of fishes in the absence of airgun noise. During the last seven drift cycles, the airgun was fired on command. This formed the treatment group. The speed of the M/V Wood averaged 0.5 m s^{-1}. The position of the acoustic launch and the seismic barge was recorded by GPS for each vessel. These data are used to calculate the range and heading of the acoustic launch relative to the seismic barge. Verification of range was made by using a handheld laser range finder. Heading was verified using the radar unit on the launch. JASCO Research Ltd. staff were not available to monitor sound levels for this experiment. Therefore, the range/intensity regression calculated for experiment 1 is used to estimate the sound intensity the fishes were exposed to in experiment 2.

For both experiments, the echosounder transmit signal power was set to 200 W and the pulse length set at 256 μs. At this pulse length, the system automatically sets the sample interval to 64 μs and the bandwidth to 8710 Hz. The ping rates were 9 Hz for experiment 1 and 11 Hz for experiment 2.

F. Fish species

Different species of fishes and fish of different sizes of the same species may react differently to seismic sounds. Therefore, attempts were made to capture representative samples of the fishes at the experimental sites using multimesh gillnets. These attempts were not successful because of water depth, current speed, and the presence of intermittent debris. Species composition is instead estimated from historical samples, current scientific knowledge, and traditional ecological knowledge.

The Mackenzie Delta has numerous species typical of large northern sub-arctic rivers. It is somewhat unique in that large numbers of juvenile coregonids migrate out of natal tributaries and through the Delta into various aquatic ecosystems (freshwater, estuarine, and near-shore marine). Broad whitefish, lake whitefish (*C. clupeaformis*), least cisco (*C. sardinella*), Arctic cisco (*C. autumnalis*), and inconnu (*Stenodus leucicthys*) are all common and either utilize or migrate through the delta at various times of the year and at different periods in their life cycle (Reist and Bond, 1988). Other species common in the Delta include northern pike and burbot (*Lota lota*), as well as numerous species from the

Cyprinid, Catostomid, Gasterosteid, and Cottid families (McPhail and Lindsey, 1970; Evans *et al.* 2002; Sawatzky *et al.*, 2007). Most of the small midwater targets seen during this study are likely to be coregonids. Young-of-the-year (YOY) of all coregonid species with perhaps the exception of inconnu are common throughout the summer. Juvenile coregonids (<4 years) are less abundant but still common, particularly lake whitefish. Most of the other species in the study area are either associated with the near-shore (e.g., northern pike) or the bottom (e.g., burbot). The authors believe that the majority of the targets are coregonids; the smaller targets being YOY migrating to rearing areas and the larger targets being residual juveniles or spawning and nonspawning adults. The adults are more likely to be inconnu, lake whitefish, and Arctic cisco.

Target strength information on juvenile freshwater species is limited. YOY coregonids range in size from 40 to 60 mm in length. Ponton and Meng (1990) looked specifically at the TS of young coregonids. They found that mean TS for fish near the surface was −56 dB and fish deeper in the water was −63 dB. For this study the authors assume that fish less than −50 dB are YOY and fish larger than −50 dB are juveniles or adults.

G. Data processing

The raw Simrad data are processed using SonarData's "ECHOVIEW" version 3.25 software. ECHOVIEW is used to identify returns as single fish targets and then to group these targets into single fish tracks based on entered parameters within the "Fish Tracking Module." Fish tracking is a complex mathematical procedure but is not foolproof. All tracks selected by the software are reviewed for accuracy. Some corrections are necessary, usually involving joining two or more tracks of what is obviously the same fish, splitting tracks of obviously different fish, or correcting tracks of intersecting fish. No software-generated tracks are subjectively eliminated in this process. Similarly, the embedded position information from the GPS is edited to ensure anomalous locations or values where the differential signal was lost are corrected. Details of the software settings used in the Fish Tracking Module are available from the authors on request. The Fish Tracking Module output vectors consist of heading and speed in three dimensions and tortuosity.

H. Statistical analysis

Statistical analysis is performed to determine if there is a significant difference in the swimming vectors and tortuosity between the control and treatment observations. In experiment 1, swimming vectors can be used directly because the acoustic launch was anchored and stationary. In experiment 2, all the swimming vectors of the fishes are corrected for the drift vector of the survey launch. This is done using the GPS fix information embedded in the Simrad acoustic data file.

The initial step in performing the statistical analysis is to divide the data into two types, vertical and horizontal. The vertical heading and all the velocity results are treated as conventional continuous linear data, and standard parametric statistics are used. The horizontal heading data are not continuously linear but circular. They are expressed as a degree angle ranging from 0° to 360°. Consequently, circular statistical analyses are used, largely following Batschelet (1981). Each of the vector components (horizontal heading, vertical heading, horizontal velocity, vertical velocity, and tortuosity) are treated as independent variables. Separate analyses are done for each.

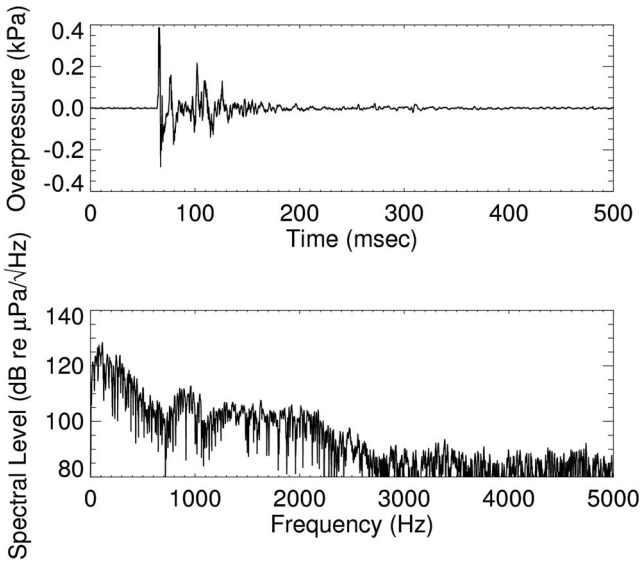


FIG. 2. Acoustic pressure waveform and spectral levels at 80 m range for a single airgun shot as measured from the acoustic survey launch during a pass-by from the airgun array (from MacGillivray *et al.*, 2004).

For experiment 1, the first step in the analyses is to use circular statistical techniques to determine if any significant directionality in the direction of travel can be observed. The control and treatment data are analyzed separately. If there is no directionality to the control or treatment samples, then a comparative analysis is meaningless. A Rayleigh test is used to determine whether the distribution of fish track headings differs significantly from randomness (Batschelet, 1981). The null hypothesis is that the parent population of fish tracks is evenly distributed in direction of travel (randomness). The r value (mean vector length for the sample direction) is calculated for each sample. Critical values for small sample sizes ($n<20$) are from Papakonstantinou (1979). For larger sample sizes ($n\geq 20$), the z-statistic, calculated as $z=nr^2$, is used. The level of significance, α, is set at 0.05 for all tests. Horizontal speed, vertical direction, vertical speed, and tortuosity are all linear variables, which can be tested with parametric statistics. Analysis of variance (ANOVA) techniques are used to compare test results to controls. As in the horizontal direction analyses, the data are subdivided by direction of barge travel and acoustic size. Each complete track in experiment 1 is considered to be the treatment with tracks recorded prior to airgun firing as the control.

The analyses for experiment 2 compare the horizontal direction, horizontal velocity, vertical direction, vertical speed, and tortuosity of fish movement before and after exposure to a single airgun shot. As in experiment 1, the horizontal direction data are analyzed using circular statistical methods and the other components of movement are treated as continuous linear variables. In experiment 2, the track for

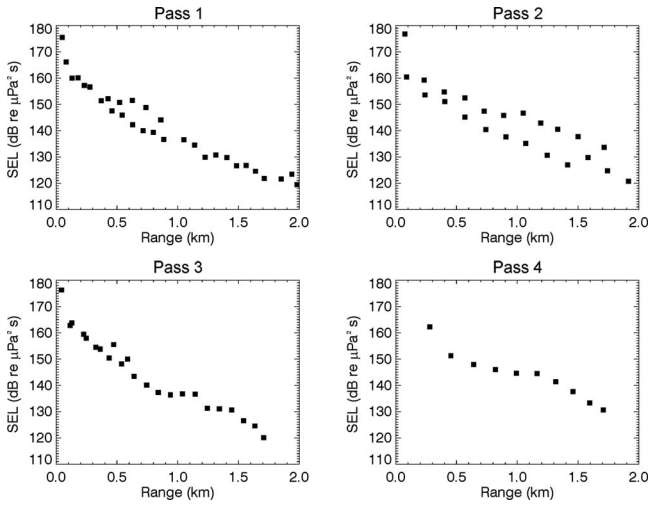


FIG. 3. SELs versus range, for passes of the airgun array. The sound levels presented are mean levels measured between the upper and lower hydrophones for each airgun shot (from MacGillivray *et al.*, 2004).

each fish is divided into two. The first section, before the seismic shot, is considered the control. The second section, after the seismic shot, is considered the treatment. Therefore, the swimming behavior before the shot is compared to that after, within a single track. Small targets (<-50 dB) are analyzed separately from larger ones (≥ -50 dB). Standard *t*-tests are used to compare pre- and post-shot linear variables.

III. RESULTS

The pressure waveform and frequency spectrum for a typical seismic airgun shot from experiment 1 with range =80 m are shown in Fig. 2. The mean sound exposure levels (SELs) at range for the four passes of the seismic barge are shown in Fig. 3. The values are the averages from two hydrophones, one at 36 m and the other at 32 m. The results from these four passes are pooled to calculate an exposure-at-range regression that is used to estimate the exposure level for experiment 2 as follows:

$$\mathrm{SEL} = -22.8\ \mathrm{range(m)} + 163.5; \quad R^2 = 0.87. \tag{1}$$

The summary statistics for the experiments are shown in Table I. During experiment 1, fishes are in the acoustic beam on average for about 4 s. The majority of fishes are in the beam for at least 3 s but rarely more than 5 s. Occasionally, individuals are seen for longer than 15 s. Experiment 2 has 33 tracks long enough to be evaluated. Fishes are in the acoustic beam on average for 3–4 s before the shot and 5–7 s after.

A. Experiment 1

Table II shows the results for the Rayleigh tests from the circular statistical analyses for experiment 1. Figure 4 shows a typical radar plot of some of the directional data. None of the circular statistical tests for non-randomness are significant (Table II). In none of the controls or treatments is there any statistical evidence that the horizontal directional movements are anything but random. Since all of the tests show that horizontal movement is random, further analyses were not conducted.

Comparison of fish depths between control and experiment groups shows no statistical difference ($t=0.510$, P-value$=0.61$). The results for the linear analyses for experiment 1 are presented in Tables III and IV. Horizontal speed, vertical speed, vertical direction, and tortuosity are all significantly different among the three control periods (July 27, July 28, and August 1). Similarly, all four variables are significantly different between the fish less than -50 dB and those larger than -50 dB when the control groups are pooled. These results show that day-to-day variability exists in the four test parameters. This could be because of slight differences in the M/V Wood position, daily variation in current speed, or a host of other variables. Because the authors cannot determine the cause of this variability, they only use the control data collected immediately before the seismic tests on August 1 for comparison to the experimental results. Pooling of the control data for comparison to the experimental data may lead to erroneous false-positive results that are caused by the variability in control data not present in the experimental data.

When the August 1 control data are compared to the experimental results, no difference among the four test parameters (horizontal speed, vertical speed, vertical direction, and tortuosity) for either of the comparisons (acoustic size and seismic barge direction) is significant except for tortuosity in fishes larger than -50 dB. Tortuosity in the control group is 28.85, nearly twice as large as any other observed period for this group. This group also had the highest stan-

TABLE I. Target track summary statistics.

				Targets per track			
						Range (m)	
Experiment	Treatment	Date	No. of tracks	Mean	Std. Err.	Minimum	Maximum
Experiment 1	Control	July 27	2337	46	0.87	5	568
	Control	July 28	365	42	1.73	5	275
	Control	August 1	96	36	2.56	5	109
	Experiment	August 1	154	36	2.18	5	170
Experiment 2	Control	August 2	33	36	3.02	5	72
	Experiment	August 2	33	68	5.27	13	163

TABLE II. Circular statistics for horizontal direction in experiment 1 [n=sample size, z=Z-statistic ($n>20$), r=Rayleigh statistic ($n\leq 20$), and P=probability].

Date	Treatment		n	z	r	P
July 27	Control	Pooled	2337	0.37		0.71
		Size <-50 dB	2038	0.45		0.65
		Size ≥ -50 dB	299	0.04		0.97
July 28	Control	Pooled	365	1.31		0.19
		Size <-50 dB	305	1.11		0.27
		Size ≥ -50 dB	60	0.64		0.52
August 1	Control	Pooled	96	0.51		0.61
		Size <-50 dB	87	0.37		0.71
		Size ≥ -50 dB	9		0.26	0.56
	Experiment	Pooled	154	0.29		0.77
		Size <-50 dB	123	0.74		0.46
		Size ≥ -50 dB	31	0.27		0.79
		Barge dir.=TD [a]	29	0.80		0.42
		Barge dir.=TU [a]	55	0.43		0.67
		Barge dir.=AD [a]	13		0.18	0.67
		Barge dir.=AU [a]	57	0.99		0.32

[a]Barge direction: TD=towards acoustic launch from downstream, TU=towards acoustic launch from upstream, AD=away from acoustic launch downstream, and AU=away from acoustic launch upstream.

dard deviation (SD=39.03) and the smallest sample size (n =9). The reasons for these comparatively high statistical values in the control data are not clear.

B. Experiment 2

A total of 33 tracks are analyzed for experiment 2 (Table I). The circular statistical analysis for non-random direction of horizontal movement for both the control and experimental groups is not significant ($P_{(\text{control})}=0.34$; $P_{(\text{experiment})}=0.21$). Consequently, the direction of movement is considered random. The radar-chart showing the direction of movement for both control (pre-shot) and experimental (post-shot) fishes is shown in Fig. 5. Note that the direction to the seismic barge is 0°. The data are corrected for all drift movements and headings of the M/V Wood. Since there is no significant directivity to either group, further analysis cannot be done. The results of the linear analysis of horizontal velocity, vertical velocity, vertical direction, and tortuosity are summarized in Table V. Paired t-tests are used to analyze each variable. There are no statistically significant differences in the means between the pre-shot and post-shot variables' mean horizontal velocity, vertical velocity, vertical direction, and tortuosity.

IV. DISCUSSION AND CONCLUSIONS

There is a limited body of knowledge on hearing in fishes, especially those native to the Mackenzie Delta (Mann *et al.*, 2007). Studies that have been conducted have shown that fishes react to sound in water and extended exposure to loud noise can result in temporary hearing loss in several, but

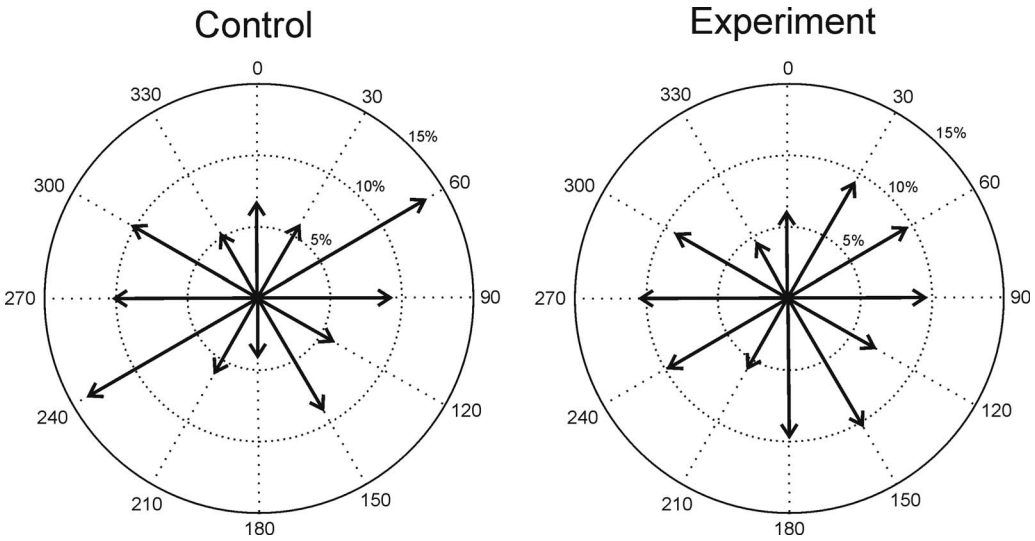


FIG. 4. Radar plots of horizontal direction of movement of fishes for experiment 1 (control and experiment).

TABLE III. Summary statistics for linear variables in experiment 1 (v=velocity, θ=angle, and t=tortuosity).

Date	Treatment		n	Horizontal speed v (m s^{-1})	SD	Vertical direction θ (deg)	SD	Vertical speed v (m s^{-1})	SD	Tortuosity t	SD
July 27	Control	Pooled	2335	0.52	0.24	−4.54	22.88	0.03	0.08	15.47	16.50
		Size < −50 dB	2036	0.53	0.24	−4.22	23.27	0.03	0.08	16.00	16.69
		Size ≥ −50 dB	299	0.47	0.24	−6.72	19.90	0.06	0.09	11.87	14.66
July 28	Control	Pooled	365	0.54	0.21	0.56	20.99	0.00	0.09	19.63	21.81
		Size < −50 dB	307	0.52	0.20	1.20	20.07	0.00	0.08	20.32	22.31
		Size ≥ −50 dB	58	0.61	0.25	−2.81	25.27	0.00	0.14	15.95	18.66
August 1	Control	Pooled	96	0.56	0.55	−1.42	18.95	0.00	0.08	14.53	15.10
		Size < −50 dB	87	0.51	0.18	0.16	17.63	0.00	0.08	13.05	9.30
		Size ≥ −50 dB	9	0.57	0.27	−16.62	25.25	0.03	0.11	28.85	39.03
	Experiment	Pooled	154	0.51	0.20	0.08	17.52	0.01	0.13	13.72	15.32
		Size < −50 dB	123	0.50	0.19	0.90	16.06	0.02	0.09	14.10	16.63
		Size ≥ −50 dB	31	0.59	0.21	−3.17	22.46	0.04	0.21	12.22	8.31
		Barge dir. = TD [a]	61	0.49	0.34	1.51	17.13	0.01	0.11	16.28	20.04
		Barge dir. = TU [a]	28	0.59	0.38	2.79	13.96	0.02	0.11	11.77	6.66
		Barge dir. = AD [a]	13	0.68	0.58	−1.57	10.51	0.04	0.13	10.51	16.73
		Barge dir. = AU [a]	57	0.60	0.39	−2.28	20.52	0.00	0.15	12.89	12.41

[a]Barge direction: TD=towards acoustic launch from downstream, TU=towards acoustic launch from upstream, AD=away from acoustic launch downstream, and AU=away from acoustic launch upstream.

not all, of the few species studied (e.g., Scholik and Yan, 2001; Smith *et al.*, 2004; Popper *et al.*, 2005). There is also some information on reaction of fishes specifically to sound pressures created by seismic airguns, but this is primarily in marine environments (e.g., Engas *et al.*, 1996; Dalen and Knutsen, 1987). Studies on response of fishes to seismic airgun sound in freshwater are much more limited; none exists in the primary literature for arctic rivers. Nonetheless, water-based seismic surveys have the potential of allowing collection of information on oil and gas deposits with much less environmental impact than land-based surveys, as no stream fording or forest clearing for access is required. The key question for regulators is whether sound produced by airguns in river seismic surveys has any significant short-term or long-term impact on fishes.

The results from experiment 1 of this study show no significant directionality to the horizontal aspect of fish movement (Table II) either during the control observation or when the fish were subjected to the seismic shots. The authors use this as an indicator of no herding as fish numbers were not high enough to statistically compare changes in target density as the airgun approached. Similarly, the authors observe no significant difference between control and experimental groups in horizontal speed, vertical speed, vertical direction, and tortuosity, except for tortuosity for fish >−50 dB (Table IV). Tortuosity of fish >−50 dB in the experimental group is significantly lower than the control group. The reason for this high value is not clear but the sample size is small (n=9), less than any other group (Table III). Important incidental observations in experiment 1 are the highly significant variability in horizontal speed, vertical speed, vertical direction, and tortuosity among the control observation days. In fact, comparison of tortuosity between experiment and other control days shows no significant differences. These results clearly demonstrate that fish movements vary from day-to-day and therefore observations made on separate days may not be comparable.

For experiment 2, the results are similar. None of the t-tests for differences in horizontal velocity, horizontal direction, vertical velocity, vertical direction, or tortuosity are significant (Table V). There is no evidence of a startle response in these fishes and the authors see no other behavioral swimming response to exposure to the seismic shot.

TABLE IV. Linear statistics for experiment 1 (F=F-statistic, P=probability, and *=significant at 5%).

Test		Horizontal speed F	P	Vertical speed F	P	Vertical direction F	P	Tortuosity F	P
Control–pooled	July 27, July 28, and August 1	40.93	0.00*	29.36	0.00*	8.62	0.00*	9.53	0.00*
Control × experiment pooled	August 1	0.14	0.70	0.65	0.42	0.40	0.52	0.17	0.68
Control × experiment	<−50 dB	0.48	0.49	3.32	0.07	0.10	0.75	0.28	0.59
Control × experiment	≥−50 dB	0.07	0.79	0.03	0.86	2.37	0.13	5.14	0.03*
Control × experiment	Barge direction	0.14	0.67	0.53	0.72	0.61	0.65	0.74	0.57

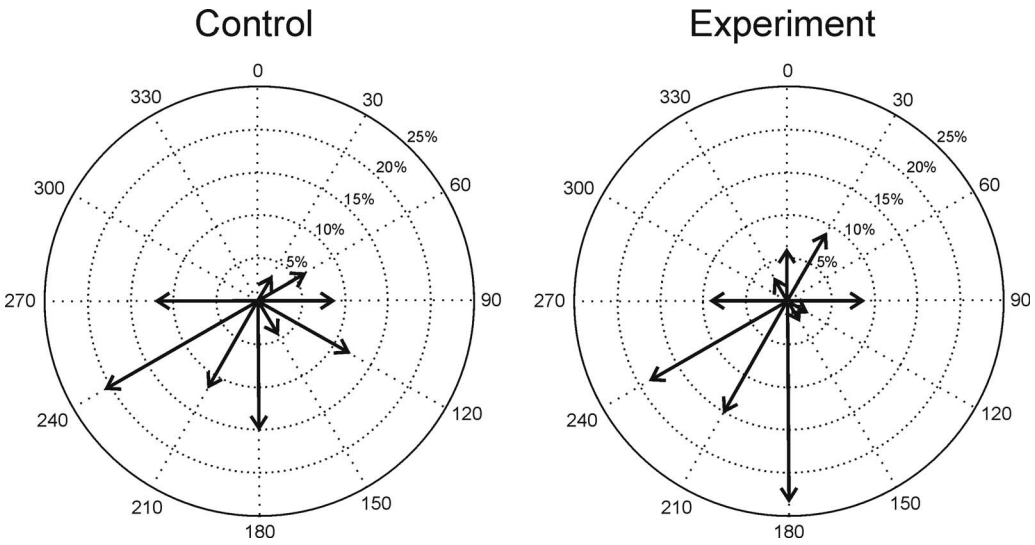


FIG. 5. Radar plots of horizontal direction of movement of fishes for experiment 2 (control and experiment).

The authors' results do not support the original hypothesis that fishes in this study will respond to their exposure to seismic sound by changing their swimming behavior. For small fishes, particularly young coregonids migrating from their natal streams through the delta to the Beaufort Sea, their small size and limited mobility may prevent any sort of meaningful avoidance swimming. Larger fishes such as mature coregonids, northern pike, and burbot are certainly capable of avoidance behavior but do not appear to do so in this study. In a commercial seismic survey, they may not avoid the vessel and would consequently be subjected to whatever sound level was in the river as the survey vessels passed. But it is important to remember that this study is limited to one location, one time of the day, one season, and one species composition. Under different environmental conditions, the same species may react differently. Similarly, different sediment compositions may have a significant impact on the behavior and intensity of the seismic sound, initiating a different response by the fish.

The evidence from the physiological component of this project suggests that test fish exposed to an airgun array used in a riverine seismic survey would at worst suffer a temporary hearing loss, while extended exposure to seismic sound, manifested as a herding response, is the most likely mechanism of possible permanent physiological damage to fishes (Popper *et al.*, 2005). The behavioral study the authors present here suggests that herding in the fishes they examined does not occur and, thus, permanent physiological damage is unlikely.

While the described research indicates fishes in this study exhibit no behavioral changes in the presence of airguns, expansion of this conclusion past the limits of this study is tenuous and caveats exist. The physiological portion of this project could be replicated to test for threshold effects on more species and size classes important in the subsistence fishery. Popper *et al.* (2005) also recommended testing species with other hearing specializations and testing at higher total noise exposures. If physiological effects are found, another behavioral component could be run at different times, locations, or conditions in an attempt to be more inclusive of species of interest and their size ranges. This would entail more positive identification of target species. A similar study that includes determination of riverbed structure and composition at the experimental site would be of value in describ-

TABLE V. Linear summary and statistics for experiment 2 (t=student's t).

Measure	Treatment	Mean	Std. Dev.	t	P-value
Time in beam (s)	Control	3.90	1.99		
	Experiment	3.19	2.58		
Horizontal velocity (m s^{-1})	Control	0.87	0.64	1.41	0.17
	Experiment	0.73	0.16		
Vertical velocity (m s^{-1})	Control	0.18	0.63	0.44	0.67
	Experiment	0.23	0.24		
Vertical direction (deg)	Control	0.05	18.94	−1.02	0.31
	Experiment	2.20	18.07		
Tortuosity	Control	6.66	4.25	0.30	0.76
	Experiment	6.36	5.24		

ing seismic signal propagation. It would also be worthwhile to test for behavioral reactions to airgun noise during specific events such as spawning runs that are critical to the stability of species that are important to subsistence fisheries. To fully sample the fish population, in particular, fish in the upper portion of the water column that are under-represented in this study due to the conical shape of the acoustic beam, the experiment could be replicated using an upward-looking transducer mounted on the riverbed. Use of an acoustic camera-type sonar system to examine fish behavior in the shallow, near-shore environment of the river would also be valuable. Examination of the swimming behavior of caged fish to an approaching airgun may give additional information as to the long-term cumulative effects of an airgun survey on fish swimming behavior. These recommendations emphasize that extrapolation of the results from this project to other species, times, locations, and sound sources should be done with caution and more research on the impacts of seismic generated noise on fish behavior and physiology is warranted.

ACKNOWLEDGMENTS

This study received funding from the Program of Energy Research and Development (PERD), Indian and Northern Affairs Canada, the Inuvialuit Fisheries Joint Management Committee (FJMC), and WesternGeco Ltd. The study was made possible through the support of numerous people and their organizations: Peter Cott, Bruce Hanna, Marty Bergmann, Ron Allen, Steve Ferguson, Blair Dunn, Mark Ouellette, and the Inuvik Area Office, of Fisheries and Oceans Canada; Edward Dillon and Merik Allen for their valuable local knowledge; Andrea Hoyt and Kevin Bill of the FJMC and their mentoring students Gerald Kisoun, Noel Cockney, and Candice Cockney; Alex MacGillivray and Melanie Austin of JASCO Research Ltd.; Dr. Arthur N. Popper of the University of Maryland; Steve Whidden of WesternGeco Ltd.; Terry Jackson and his crew from Conquest Seismic Survey Services Ltd.; Les Harris of the Gwich'in Renewable Resource Board; and Angus Alunik and Keith Rosindell. In addition, the authors would like to thank the reviewers of this paper; Dr. Arthur N. Popper, Peter Cott, and Alex MacGillivray, as well as three anonymous reviewers, for their useful comments.

Batschelet, E. (**1981**). *Circular Statistics in Biology* (Academic, London).

Cott, P. A., Hanna, B. W., and Dahl, J. A. (**2003**). "Discussion on seismic exploration in the Northwest Territories 2000-2003," Can. Manuscr. Rep. Fish. Aquat. Sci. 2648, 42 pages.

Dalen, J., and Knutsen, G. M. (**1987**). "Scaring effects on fish and harmful effects on eggs, larvae and fry by offshore seismic explorations," in *Progress in Underwater Acoustics*, edited by H. M. Merklinger (Plenum, New York), pp. 93–102.

Engas, A., Løkkeborg, S., Ona, E., and Soldal, A. V. (**1996**). "Effects of seismic shooting on local abundance and catch rates of cod and haddock," Can. J. Fish. Aquat. Sci. **53**, 2238–2249.

Evans, C. E., Reist, J. D., and Minns, C. K. (**2002**). "Life history characteristics of freshwater fishes occurring in the Northwest Territories and Nunavut, with major emphasis on riverine habitat requirements," Can. MS Rep. Fish. Aquat. Sci. 2614, 177 pages.

Handegard, N. O., and Tjostheim, D. (**2005**). "When fish meet 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.

IMG-Golder Corp. (**2002**). "Behavioural and physical response of riverine fish to airguns," Report No. 022-2265, WesternGeco, Calgary, AB, Canada.

MacGillivray, A., Austin, M., and Hannay, D. (**2004**). *Underwater Sound Level and Velocity Measurements From 2004 Study of Airgun Noise Impacts on Mackenzie River Fish Species* (JASCO Research Ltd., Victoria, BC, Canada).

Mann, D. A., Cott, P. A., Hanna, B. W., and Popper, A. N. (**2007**). "Hearing in eight species of northern Canadian freshwater fishes," J. Fish Biol. **70**, 109–120.

McPhail, J. D., and Lindsey, C. C. (**1970**). "Freshwater fishes of northwestern Canada and Alaska," Bull.-Fish. Res. Board Can. **173**, 381.

National Energy Board of Canada (**2004**). "Canada's conventional natural gas resources: A status report," National Energy Board of Canada, Calgary, AB, Canada.

Papakonstantinou, V. (**1979**). "Beiträge zur zirkulären statisitik (Contributions to circular statistics)," Ph.D. thesis, University of Zurich, Switzerland.

Ponton, D., and Meng, H. J. (**1990**). "Use of dual-beam acoustic technique for detecting young whitefish, *Coregonus* sp., juveniles: First experiments in an enclosure," J. Fish Biol. **36**, 741–750.

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.

Reist, J. D., and Bond, W. A. (**1988**). "Life history characteristics of migratory coregonids on the lower Mackenzie River, Northwest Territories, Canada," Finn. Fish. Res. **9**, 133–144.

Sawatzky, C. D., Michalak, D., Reist, J. D., Carmichael, T. J., Mandrak, N. E., and Heuring, L. G. (**2007**). "Distributions of freshwater and anadromous fishes from the mainland Northwest Territories, Canada," Can. MS Rep. Fish. Aquat. Sci. 2793, 253 pages.

Scholik, A. R., and Yan, H. Y. (**2001**). "Effects of underwater noise on auditory sensitivity of a cuyprinid fish," Hear. Res. **152**, 17–24.

Smith, M. E., Kane, A. S., and Popper, A. N. (**2004**). "Acoustical stress and hearing sensitivity in fishes: Does the linear threshold shift hypothesis hold water?," J. Exp. Biol. **207**, 3591–3602.

Worcester, T. (**2006**). "Effects of seismic energy on fish: A literature review," DFO Can. Sci. Advis. Sec. Res. Doc. 2006/092, 70 pages.

Erratum: 1aSCb14. Effects of sleep deprivation on nasalization in speech [J. Acoust. Soc. Am. 125, 2499 (2009)]

Xinhui Zhou
Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

Suzanne Boyce
University of Cincinnati, Cincinnati, Ohio 45267

Joel MacAuslan
Speech Technology and Applied Research Corporation, Bedford, MA 01730

Walter Carr, Thomas Balkin, and Dante Picchioni
Walter Reed Army Institute of Medicine, Silver Spring, Maryland 20910

Allan Braun
National Institute of Deafness and Other Communication Disorders, Bethesda, Maryland 20892

Carol Espy-Wilson
University of Maryland, College Park, Maryland 20742

(Received 10 July 2009; accepted 10 July 2009)
[DOI: 10.1121/1.3192342]

PACS number(s): 43.70.Bk, 43.10.Vx

There is a publisher's error in the list of authors of this abstract in the published program. The correct author list is as follows: Xinhui Zhou (Dept. of Elect. and Comp. Eng., Univ. of Maryland, College Park, MD 20742, zxinhui@glue.umd.edu), Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45267), Joel MacAuslan (Speech Technol. and Appl. Res. Corp., Bedford, MA 01730), Walter Carr, Thomas Balkin, Dante Picchioni (Walter Reed Army Inst. of Medicine, Silver Spring, MD 20910), Allan Braun (National Inst. of Deafness and Other Communication Disorders, Bethesda, MD, 20892), and Carol Espy-Wilson (Univ. of Maryland, College Park, MD, 20742).

 0001-4966/2009/126(3)/1607/1/$25.00

Erratum: 1pSC11. Discriminating dysarthria type and predicting intelligibility from amplitude modulation spectra [J. Acoust. Soc. Am. 125, 2530 (2009)]

Susan J. LeGendre
Speech, Language, and Hearing Sciences, University of Arizona, P.O. Box 210071, 1131 East 2nd Street, Tucson, Arizona 85721-0071

Julie M. Liss
Arizona State University, Tempe, Arizona 85287-0102

Andrew J. Lotto
University of Arizona, Tucson, Arizona 85721-0071

(Received 10 July 2009; accepted 10 July 2009)
[DOI: 10.1121/1.3192343]

PACS number(s): 43.70.Dn, 43.10.Vx

There is a publisher's error in the list of authors of this abstract in the published program. The correct author and affiliation list is as follows: Susan J. LeGendre (Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box 210071, 1131 E. 2nd St., Tucson, AZ 85721-0071, sjlegend@email.arizona.edu), Julie M. Liss (Arizona State Univ., Tempe, AZ 85287-0102), and Andrew J. Lotto (Univ. of Arizona, Tucson, AZ 85721-0071).

ACOUSTICAL NEWS

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

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

Report of the Auditor

Published herewith is a condensed version of our auditor's report for calendar year ended 31 December 2008

Independent Auditors's Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America (the "Society") as of December 31, 2008 and 2007 and the related statements of activities and cash flows for the years then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audits.

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

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of December 31, 2008 and 2007 and the changes in its net assets and its cash flows for the years then ended in conformity with accounting principles generally accepted in the United States of America.

O'Connor Davies Munns & Dobbins, LLP
New York, New York
May 22, 2009

ACOUSTICAL SOCIETY OF AMERICA
STATEMENTS OF FINANCIAL POSITION
31 DECEMBER 2008 and 2007

	2008	2007
Assets		
Cash and cash equivalents	$ 2,685,634	$ 2,796,600
Accounts receivable, net	152,679	119,024
Due from related parties	–	181,869
Investments	7,536,938	9,714,086
Property and equipment, net	19,379	34,579
Other assets	507,455	400,688
	$ 10,902,085	$ 13,246,846
Liabilities and Net Assets		
Liabilities		
Accounts payable and accrued expenses	$ 348,576	$ 445,713
Due to related parties	130,334	16,717
Deferred revenue	1,277,087	1,539,161
Total liabilities	1,755,997	2,001,591
Net assets		
Unrestricted	8,188,805	10,035,476
Temporarily restricted	657,283	909,779
Permanently restricted	300,000	300,000
Total net assets	9,146,088	11,245,255
	$ 10,902,085	$ 13,246,846

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF ACTIVITIES
FOR THE YEARS ENDED 31 DECEMBER 2008 AND 2007

	2008	2007
Changes in Unrestricted Net Assets Revenue		
Revenues:		
Dues	$ 734,483	$ 710,461
Publishing—JASA	2,413,348	2,446,387
Standards	370,390	352,458
Spring Meeting	65,034	240,594
Fall Meeting	220,721	214,452
Other member services revenue	33,594	21,319
Other	165,795	156,006
Net assets released from restrictions	62,753	88,269
	4,066,118	4,229,946
Expenses:		
Publishing	1,661,248	1,517,417
Standards	414,203	396,502
Spring Meeting	92,308	237,305
Fall Meeting	374,788	287,469
Member Services	261,210	269,327
Other	604,335	517,701
Administration	683,965	531,828
	4,092,057	3,757,549
Net Income from Operations	($25,939)	$472,397
Non-operating activities:		
Interest and dividends	344,244	359,489
Realized gain on investments	156,769	380,352
Unrealized (loss) gain on investments	(2,321,745)	777,681
	(1,820,732)	1,517,522
Change in Unrestricted Net Assets Before Adoption of		
Recognition Provisions of FASB Statement No. 158	(1,846,671)	1,989,919
Effect of adoption of FASB Statement No. 158	—	51,119
	(1,846,671)	2,041,038
Changes in Temporarily Restricted Net Assets		
Contributions	8,200	22,637
Interest and dividends	27,981	40,799
Realized gain	16,343	64,556
Unrealized (loss) gain	(242,267)	131,994
Net assets released from restrictions	(62,753)	(88,269)
Change in Temporarily Restricted Net Assets	(252,496)	171,717
Change in net assets	(2,099,167)	2,212,755
Net Assets		
Beginning of year	11,245,255	9,032,500
End of year	$ 9,146,088	$11,245,255

Results of the ASA Election

The following candidates were elected Officers and Members of the Executive Council in the 2009 Society election:
George V. Frisk, *President-Elect*
Judy R. Dubno, *Vice President-Elect*
James H. Miller, *Member of the Executive Council*
Scott D. Sommerfeldt, *Member of the Executive Council*

Annual Reports of Technical Committees (See November issue for additional reports)

Acoustical Oceanography

This is my last annual report as Chair of the Acoustical Oceanography Technical Committee. I want to thank all of the members of the AO TC for their support, encouragement, and collegiality. Our new Chair is Martin Siderius of the University of Portland and I am sure that he can depend on the membership like I did. Special thanks go to my colleague, Dr. Gopu Potty, who served as a "Vice Chair" of the Technical Committee. Dr. Potty organized the Best Student Paper competitions in New Orleans, Paris, Miami and Portland, managed the AO web site, and assisted in our AO TC meeting refreshment logistics.

Fall 2008 Meeting (Miami, Florida). The Technical Committee on Acoustical Oceanography (AO) sponsored two special sessions: "Attenuation coefficient of sediments from low- to mid-frequencies" organized by Jim Lynch and "Three-dimensional acoustics and inversions on the Continental Shelf and canyons" also organized by Jim Lynch. Both of these sessions were also sponsored by the Underwater Acoustics Technical Committee. AO also was a co-sponsor of one special session: "Acoustics of harbors, ports and shallow navigable waterways." Special thanks go to Jennie Wylie who served as the AO representative to the Technical Program Organizing Meeting for Miami. The Best Student Paper Awards in Acoustical Oceanography went to Megan Ballard (first prize) of the Pennsylvania State University for her paper "Variability of the water column sound speed profile and its effect on acoustic propagation during the Shallow Water 2006 Experiment," to Lin Wan (Second Prize) of the Georgia Institute of Technology for the paper "Sound speed and attenuation in the sea bottom from broadband sound propagation in the Yellow Sea," and to Theodore F. Argo IV (Third Prize) of the University of Texas for his paper "Laboratory measurements of sound speed and attenuation in water-saturated artificial sediments as a function of porosity." The 2007 Science Writing Award for Professionals in Acoustics was presented in Miami to Kathleen Vigness Raposa, Gail Scowcroft, Christopher Knowlton, and past AOTC Chair Peter Worcester for the website "Discovery of Sound in the Sea."

Spring 2009 Meeting (Portland, Oregon). The 2008 Medwin Prize in Acoustical Oceanography was awarded to Martin Siderius of Portland State University for his contributions to acoustic measurement of bubble, plankton and fish. Dr. Siderius presented the AO Prize Lecture entitled "For imaging the structure of the ocean bottom using ambient sound." AO sponsored two special sessions in Portland: "Temporal and spatial field coherence applied to ocean sensing" organized by Tim Duda, Kyle Becker, and Michael Brown, "Environmental inferences in inhomogeneous ocean environments" organized by Mohsen Badiey. AO also was a co-sponsor of four special sessions: "Autonomous remote monitoring systems for marine animals" with Animal Bioacoustics, "Poroelastic materials: Models, bounds, and parameter estimation" with Signal Processing in Acoustics, "Physics-based undersea clutter model verification and validation" with Underwater Acoustics, and "Session in honor of Ralph Goodman" with Underwater Acoustics. Special thanks go to Jeffrey Nystuen who served as the AO representative to the Technical Program Organizing Meeting for Portland. Winners of the Acoustical Oceanography Best Student Paper Awards will be announced before the meeting in San Antonio. Information on these and related matters is available on the Acoustical Oceanography Technical Committee website. It can be reached through the ASA web page by clicking on "Committees."

JAMES H. MILLER
Chair 2008–2009

Architectural Acoustics

TCAA continued to enjoy strong member numbers and levels of participation in 2007–08. At the Fall 2008 meeting in Miami, TCAA special sessions and their organizers included Dick Godfrey—"Acoustics of single family residences;" Lou Sutherland, David Lubman—"Special session in honor of Mike Nixon;" Brandon Tinianov—"Green Building standards and acoustics;" Angelo Campanella—"Multifamily structures—advances and legal issues;" Jessica Clements—"What went wrong: facilities you designed;" Michael Yantis and Bill Dohn—"Acoustics of small, multipurpose performance spaces;" Dana Hougland and Dave Woolworth—"Acoustics of retrofitted performance spaces;" Bill Cavanaugh—"Knudsen Memorial Lecture—Barry Blesser;" Damian Doria—"Celebrating the works of Russell Johnson;" Molly Norris and Scott Pfeiffer—"Use of innovative materials." Gary Siebien served as the TCAA representative to the Technical Program Organizing Meeting for this meeting.

At the Spring 2009 meeting in Portland, TCAA special sessions and their organizers included Dave Bradley—"Measurements and modeling of scattering effects;" Bob Coffeen—"Computer auralization;" Ken Roy—"Acoustics of health and healing environments;" James Phillips—"Mechanical equipment in multifamily dwellings;" Steve Pettyjohn—"Acoustics of mixed use buildings;" Bob Coffeen—"Outdoor performance spaces;" Dave Woolworth—"SPL, loudness and room acoustics;" Lily Wang—"Indoor noise criteria;" Brandon Tinianov and David Sykes—"Acoustics of green buildings: A 360° panel discussion;" Brigitte Schulte-Fortkamp—"Soundscapes;" Ning Xiang and Allan Pierce—"Preparing JASA and JASA-EL articles;" Boaz Rafaely and Ning Xiang—"Multichannel systems in room acoustics." David Bradley served as the TCAA representative to the Technical Program Organizing Meeting for this meeting.

There was an enormous effort on the part of these chairs and the high quality papers presented within the sessions. Thanks to each of them for the dedication in support of the TC. Achieving good acoustics in healthcare facilities continues to be an important topic that TCAA members are strongly advocating. For almost four years, the joint subcommittee with the Technical Committee on Noise concerning Speech Privacy, chaired by Greg Tocci and David Sykes, has been very active with special sessions at both of the this year's meetings. More information about this subcommittee's activities may be found on the website: www.speechprivacy.org.

Also continuing to draw ongoing activity is the topic of classroom acoustics. We are nearing completion of an addition to the ANSI standard S12.60, with expansion to address the large number of modular or 'portable' classrooms in use today. Resolution of the topic is expected next year.

Our newest subcommittee on Green Building Acoustics continues to influence green building standards and the acoustical issues they sometimes create. Well attended special sessions took place at both the Miami and Portland meetings. This group can be found at the webpage: http://groups.google.com/group/asa-gba.

This year the TCAA again offered Best Student Paper Awards. Two awards were presented for papers given at the Miami meeting. Congratulations to first place co-recipients Robert M. Tanen and Jonathan C. Silver of the University of Hartford and second place recipient Linda Gedemer of Rensselaer Polytechnic Institute. The Portland meeting also featured the student design competition. First honors went to Karl Eriksson, Rikard Olsson, and Victor Gunnarsson of Chalmers University of Technology. Commendations, in no particular order were awarded to: Shane Kanter, Jon Birney, John Hodgson—University of Kansas; Alic Vedad, Hanna Mangs, Daniel Johansson—Chalmers University of Technology; Anna Sandberg, Caroline Werner, Elis Johansson—Chalmers University of Technology; Sebastian Fors, Lisa Kinnerrud, Fredrik Hagman—Chalmers University of Technology.

Congratulations to TCAA members who became ASA Fellows this year. New Fellows include Marshall Long, Russell E. Berger and Trevor R. T. Nightingale. We are also delighted to announce the awarding of the ASA Silver medal to John Bradley.

Many thanks to the following individuals for their hard work this past year in ASA on behalf of TCAA: Dana Hougland is our representative on the Medals and Awards Committee. Ron Freiheit serves on the Membership Committee, and George Winzer is TCAA representative to the ASA Committee on Standards. Lauren Ronsse from the University of Nebraska has served as our Student Council representative and will next serve as the

Student Council Chair. The Associate Editors in Architectural Acoustics are Lily Wang and Ning Xiang for JASA, Ning Xiang for JASA Express Letters. Ralph Muehleisen is the editor of POMA. Tony Hoover serves on the Editorial Board for Acoustics Today. Finally, thanks to Alex Case for his service as TCAA Secretary.

BRANDON TINIANOV
Chair

Biomedical Ultrasound/Bioresponse to Vibration

I would first like to acknowledge my predecessor, Michael Bailey, for his hard work the past three years. The Biomedical Ultrasound/Bioresponse to Vibration Technical Committee (BB) is one of the smaller Technical Committees, but we are very active in the society and Mike did an outstanding job keeping things organized during his tenure.

As is normally the case for Fall meetings, the 156th ASA meeting in Miami was fairly calm. There were no BB special sessions, but cosponsored special sessions included "High precision acoustical measurements" and "Recent developments in coded signals in acoustics." There were two BB contributed sessions, "Ultrasound interaction with tissues" and "Microbubble response and modeling" chaired by Michael Oelze and Tyrone Porter, respectively. Saurabh Datta was the TPOM representative. One of our members, Arman Sarvazyan, was awarded fellowship at the Plenary Session.

Special sessions (and organizers) at the 157th ASA meeting in Portland included 1) "Image Enhancement and Targeted Drug and Gene Delivery" (Azzdine Ammi and Saurabh Datta), 2) "Biomedical Applications of Standing Waves" (Armen Sarvazyan), 3) "Medical Ultrasound Imaging with High Frequency and/or Contrast Agents" (John Allen and Jonathan Lindner), 4) "Shock Wave Therapy" (Michael Bailey and Thomas Matula), 5) "Metrology and Calibration of High Intensity Focused Ultrasound" (Peter Kaczkowski), and 6) "Biomedical Applications of Acoustic Radiation Force" (Mostafa Fatemi). Azzdine Ammi and Peter Kaczkowski were the TPOM representatives for the Portland meeting. The Student Paper Award was given to Kelley Garving for her paper "Ultrasound standing wave fields control the spatial distribution of cells and protein in three-dimensional engineered tissue."

A few items of business arose in Portland concerning BB. First, Tyrone Porter is Chair of a joint Ad Hoc Committee to revise the format of the student paper competition for BB and Physical Acoustics. There has been concern about feedback to students and uniformity of judging and Tyrone will propose a revised format that will address these issues. Mark Wochner, Kim Lefkowitz, and Matt Poese will also be serving on the committee. The plan is to implement the new format at the 159th ASA Meeting in Baltimore. Second, the official "scope" or description of BB is very much out of date and needs to be revised. Christy Holland, Charlie Church, Doug Mast and Ron Roy will serve on an Ad Hoc Committee to revise the scope and bring it into line with current research activity within BB. A draft will be discussed at the BB Technical Meeting at the 158th ASA Meeting in San Antonio and the draft will be submitted for final approval, after discussion, in Baltimore.

In addition to everyone who has volunteered to help this past year, we would like to acknowledge the contribution of some specific individuals. The associate editors for BB are Charles Church, Floyd Dunn, and Douglas Miller. Shira Broschat maintains the BB website http://moab.eecs.wsu.edu/.shira/asa/bubv.html. Lawrence Crum is the representative to the Medals and Awards Committee. Christy Holland is the representative to the Membership Committee. Peter Kaczkowski and Vera Khokhlova are the representatives to ASACOS. Michael Canney is the Chair of the Student Council and Lucie Somaglino is the BB representative to the Student Council.

JEFFREY KETTERLING
Chair

Engineering Acoustics

At the Fall, 2007 meeting in Miami, Engineering Acoustics organized two special sessions. Victor Nedzelnitsky organized "High precision acoustical measurements" and Michael Scanlon organized "Acoustics for battlefield operations and homeland security." Both sessions were well attended and generated an interesting set of papers across diverse viewpoints. Dave Brown presented a Hot Topics paper, "Single crystal piezoelectric materials and acoustic devices."

At every meeting, the EATC sponsors a student paper competition. At the Fall meeting, the winners of this competition were George Lewis for "Development of a portable therapeutic ultrasound system for military, medical, and research use" and Scott Porter for "Verification of a method for measuring magnetostrictive parameters for use in transducer design modeling."

At the Spring, 2008 meeting in Portland, our TC organized three special sessions. Tom Howarth organized "Acoustic engineering of wind turbines," Ken Walsh organized "Lasers in underwater acoustics," and Tom Howarth organized "Piezoelectric energy harvesting." Once again, these sessions demonstrated the strength of Engineering Acoustics to draw interest and authors across several other Technical Committees and encourage interdisciplinary discussion.

The winners of the student paper competition were Timothy Marston for "Infrasonic microphones" and Scott Porter for "Rapid identification of candidate materials for tonpilz head-mass design." The student contributions have been excellent; however, we have noticed that many student presenters who are qualified to compete are not involved in the contest simply because they have not requested participation when they submit their abstracts. I would like to encourage all students and their advisors to make sure that they submit their abstracts for the student competition. There is very little additional work on the part of the student, and the prize money and resume material is easily worthwhile. I have seen several excellent papers go unrecognized by our committee; I strongly encourage all students to participate.

The EATC is represented on the Student Council by Scott Porter, on Metals and Awards by Kim Benjamin, on Membership by Steve Thompson, on the Journal by Allan Zuckerwar, and on ASACOS by Robert Drake.

All TC's have been asked to rewrite their scope statement. We have created a Subcommittee for EATC Scope Revision, with the members Steve Thompson, Daniel Warren, Beth McLaughlin, Victor Nedzelnitsky, Mark Sheplak, and Hasson Tavossi. Please forward and suggestions or comments for the revised scope statement to daniel.warren@knowles.com. Finally, the committee held elections for a new chairman this year. Mike Scanlon has begun his three year term as of the end of the Portland meeting. The outgoing chair would like to thank the Acoustical Society and the Engineering Acoustics Technical Committee for a rewarding and productive term.

DANIEL M. WARREN
Chair 2006–2009

Musical Acoustics

During 2008–2009 the Technical Committee on Musical Acoustics (TCMU) was chaired by Paul Wheeler. Representatives to the committee were: James P. Cottingham, membership; Uwe J. Hansen, Medals and Awards; Diana Deutsch, ASACOS; Eric A. Dieckman, Student Council; and Paul A. Wheeler, Technical Council. Associate Editors were Diana Deutsch and Neville H. Fletcher (JASA), Thomas D. Rossing (Express Letters), and James W. Beauchamp (POMA). Technical Program Organizing Committee (TPOM) representatives were Edward Large (Miami) and Paul Wheeler (Portland). Those appointed or reappointed as TCMU members for 2009–2012 were Rolf Bader, Xavier Boutillon, Jonas Braasch, Murray D. Campbell, Andrey R. DaSilva, Nicholas J. Giordano, William M. Hartmann, William L. Martens, Andrew C. H. Morrison, James M. Pyne, Sten O. Ternstrom, and Shigeru Yoshikawa.

At the 156th Meeting in Miami, held November 10–14, 2008, TCMU presented four special sessions: 1aMU/1pMU/2aMU "Dynamical Approaches in the Study of Music Perception and Performance I-III" organized and chaired by Edward W. Large; 2pMU "Telematic Music Technology" organized and chaired by Jonas Braasch; 3aMU "Structural Vibration in Musical Instruments" organized and chaired by Uwe J. Hansen; and 4pMU "Statistical Approaches for Analysis of Music and Speech Audio Signals" organized and chaired by Paris Smaragdis and George Tzanetakis. At the Miami meeting Gabriel Weinreich received the Silver Metal award and Murray D. Campbell received the Rossing Prize in Acoustics Education. The winners of the Best Student Papers were: Summer Rankin "Tempo fluctuation and perceptive synchrony in music" and Hiroko Terasawa "A hybrid model of timbre perception."

At the 157th Meeting in Portland, held May 18–22, 2009, TCMU presented four special sessions: 1pMU, "Microphone Array Techniques in Musical Acoustics" organized and chaired by Rolf Bader; 2aMU/2pMU "Wind Instruments I & II" organized and chaired by Thomas D. Rossing and Dean R. Ayers; 3pMU "Acoustics of Bagpipes" organized and chaired by Murray D. Campbell; and 4aMU "Musical Perception and Modeling" organized and chaired by Diana Deutsch. In addition, TCMU presented two lectures/mini concerts: Members of the Edinburgh Renaissance Band and friends organized by Murray D. Campbell and a bagpipe demonstration presented by Kevin Carr which was arranged by Thomas D. Rossing. During the Portland meeting Thomas D. Rossing received the Gold Medal Award. The Best Student Papers were: Nicholas Goodweiler "Acoustics of single reed duck calls" and David Krueger "Acoustic and vibrometry analysis of beating in a large Balinese gamelan gong."

James Cottingham continues to maintain a very useful website for TCMU (http://www.public.coe.edu/~jcotting/tcmu/). It includes links to future meetings, minutes to previous TCMU meetings, annual reports, student paper award winners, as well as useful links to teaching websites and musical acousticians.

PAUL WHEELER
Chair

Noise

TC Noise was represented at the 156th meeting of ASA in fall 2008 in Miami with a Tutorial Lecture, 2 workshops and 4 special sessions. The Tutorial was on Aircraft Noise Prediction by Joe Posey. For the workshops: It was the joint workshop with Animal Bioacoustics on Advances in measurement and noise and noise effects on humans and non-human animals in the environment I and II by Ann Bowles and Brigitte Schulte-Fortkamp as a follow up to the workshop in New Orleans. Another follow up will be held in the meeting with Noise–Con 2010 in Baltimore. The second workshop joint with Architectural Acoustics and ASA Committee on Standards was on Standardization for soundscape techniques: Soundscape and sound quality-—Measurement and lexicon by Brigitte Schulte-Fortkamp, Bennett Brooks, and Bob Kull. Noise and Architectural Acoustics had 3 joint special sessions: Fire codes and acoustics by Matthew V. Golden and Ralph T. Muehleisen, and the Acoustical issues of Green Buildings by Brandon D. Tinianow; Sound levels and acoustical characteristics of modular classrooms by Paul Schomer. A further session was on Topics in Noise-Active noise, product noise, and community noise by Erica Ryherd.

Noise was proud to announce Joe W. Posey as a new Fellow and the Young presenters who received awards at the Paris meeting: Sarah Gourlie, The University of Texas, USA, Brice Lafon (Renault, France, and Sarah R. Payne, The University of Manchester, UK Erica Ryherd and Rich W Peppin were the Noise representatives at the Technical Program Organizing Meeting for Miami.

For final conclusions on this meeting it should be highlighted that the joint session "Noise and Noise Effects on Animals and Humans in the Environment I+II" was again a success and understood as a next step in collaboration with respect to measures and methodologies in research in both areas of Animal Bioacoustics and Noise. The Soundscape workshop brought new results needed for the application in city planning and was continued at the ASA meeting in Portland, May 2009. Moreover, the workshops over the last few years on Soundscapes were the basis for the successful application for the innovative project competition being conducted by the Acoustical Society of America.

The efforts of several volunteers should be recognized: Nancy Timmerman is the Noise representative on the Medals and Awards Committee, Bennett Brooks is the representative on the Membership Committee and Richard Peppin is the representative on the ASA Committee on Standards. Cole Duke is the Student Council representative and, following the practice initiated a few meetings ago, the scribe for the TC-Noise meetings. The Noise web page _http://www.nonoise.org/quietnet/tcn/ is maintained by Les Blomberg. Joe Posey coordinated the Noise Young Presenter Awards.. Serving as Associate Editor for JASA Express Letters is Mike Stinson and JASA Associate Editors are: Kenneth A. Cunefare, Kirill V. Horoshenkov and Brigitte Schulte-Fortkamp.

At the 157th ASA meeting in Portland Noise was represented with the Short Course on Outdoor noise estimation and mapping by Bob Putnam, Ken Kalinski, Brigitte Schulte-Fortkamp, and Klaus Genuit. Also, a one day symposium took place together with the City of Portland "Urban design with soundscape in mind. A symposium on urban planning with the consideration of noise impacts and the people concerned" is an outcome of the application for the innovative project competition being conducted by the Acoustical Society of America—organized by Kerrie Standlee. More than 80 people from the City of Portland and ASA attended this event.

As "Urban Design with Soundscape in Mind" is an innovative way to connect ASA with the community at large and will pull together urban noise impact information discussed on a regular basis at ASA conferences and place it in the hands of those that can immediately use it to improve the lives of citizens across the nation, it will have a follow up in Baltimore.

Furthermore, Noise had 2 joint special sessions with Architectural Acoustics: Prediction and control of noise related to buildings by James E. Phillips; Acoustics of green buildings. A 360 degree perspective by David M. Sykes and Brandon Tinianow; 3 special sessions joint with Architecture and ASA Committee on Standards: Hospital noise and health care facilities by Erica Ryherd; Soundscape techniques and applications—community and urban environments by Brigitte Schulte-Fortkamp and Bennett Brooks: Soundscape techniques and applications—wilderness and park soundscapes by Nancy Timmerman and Paul Schomer. Furthermore, one special session with Animal Bioacoustics and ASA Committee on Standards: Bioacoustic metrics and the impact of noise on the natural environment by Michael Stocker. Additionally: Noise litigation by John Erdreich; Road vehicle and construction noise measurement, modelling and control by Kerrie Standlee.

Congratulations to Clinton Francis, Dept. of Ecology and Evolutionary Biology, University of Colorado who received the Young presenter award at the Miami meeting and thanks to Kerrie Standlee who was the Noise representative at the Technical Program Organizing Meeting for Portland.

Also, the efforts of several volunteers should be recognized: Nancy Timmerman served a last time as the Noise representative on the Medals and Awards Committee, Bennett Brooks is the representative on the Membership Committee and Richard Peppin is the representative on the ASA Committee on Standards. Serving as Associate Editor for JASA Express Letters is Mike Stinson and JASA Associate Editors are: Kenneth A. Cunefare, Kirill V. Horoshenkov and Brigitte Schulte-Fortkamp. The Noise web page is maintained by Les Blomberg. Cole Duke is the Student Council representative and the scribe at the TC-Noise meetings.

Let me close my report with my best wishes for Nancy Timmerman who is now in charge to chair TC Noise. As past chair I would like to take this opportunity to thank all the members of Noise and ASA who have been so supportive during my three years as Chair. It has been an interesting and rewarding experience.

BRIGITTE SCHULTE-FORTKAMP
Chair 2006–2009

Psychological and Physiological Acoustics

Reflecting P&P's policy of emphasizing spring meetings, the fall meeting in Miami, Florida was a small one. The P&P section sponsored two sessions. We thank Gail Donaldson for being our representative at the Technical Program Organizing Meeting.

The spring meeting in Portland, Oregon was a great success, with many exciting sessions. P&P's own Marjorie Leek held it all together as the Technical Chair for the whole meeting, and Erick Gallun was P&P's representative at the Technical Program Organizing Meeting. Erick Gallun, along with Nat Durlach, also organized P&P's special session on Theory Construction in the Domain of Auditory Perception, the stellar speaker line-up which included no fewer than five ASA silver medal winners, and Barbara Shinn-Cunningham was invited to give one of the interdisciplinary "Hot Topics" talks. The technical committee meeting was well attended, and the current Associate Editors (AEs) presented their annual reports to the membership. Rich Freyman is stepping down after three years of outstanding service as AE, and Ruth Litovsky has very kindly agreed to serve a second three-year term. We thank our other current AEs, Brenda Lonsbury-Martin, John Middlebrooks, Brian Moore, Bill Shofner, and Magdalena Wojtczak for their continuing service, and we are pleased to welcome two new AEs, Chris Plack and Michael Akeroyd. Erick Gallun, whose name is beginning to appear suspiciously often in this column, remains AE for POMA, and Qian-Je Fu continues to be AE for JASA-EL. The technical committee meeting was used as the venue to elect new members to the P&P technical committee. The terms of Amy Horwitz, Bert Schlauch, Kathy Arehart, Gle-

nis Long, Enrique Lopez-Poveda and Stan Sheft came to an end, the terms of Bernhard Seeber, Dan Tollin, Emily Buss, Erick Gallun (again), Jose Alcantara, and Sridhar Kalluri began, and the following members were newly elected to begin their terms in 2010: John Grose, Ervin Hafter, Walt Jesteadt, Lori Leibold, Nicole Marrone, and Pamela Souza. Thanks to Kathy Arehart and Bert Schlauch for recruiting the candidates and organizing the election.

We thank Bill Hartmann, our representative to the Medals and Awards Committee, Lynne Werner, our representative to the Membership Committee, and Brent Edwards, our representative to ASACOS. The P&P Technical Initiatives continue unchanged. The initiatives include limited travel support for invited speakers, student receptions, and homepage maintenance. Suggestions for uses of funds, including innovations such as workshops, satellite meetings, etc., are welcome and should be sent to oxenham@umn.edu.
ANDREW J. OXENHAM
Chair

Signal Processing in Acoustics

The Signal Processing Technical Committee (SPTC) has been quite active during the period of this report. We organized 3 sessions for the 156th ASA meeting in Miami, November 10–14, 2007; and 3 sessions for the 157th ASA meeting in Portland, May 18–22, 2009. We wish to thank our Technical Program Organizing Meeting (TPOM) representatives, David Chambers for Miami and Paul Hursky for Portland, for their work in organizing the SP sessions for the meetings.

The special sessions for Miami were "Recent developments in coded signals in acoustics" organized by David Waddington, "Signal processing for high clutter environments" organized by Ron Wagstaff and Joal Newcomb, and "Autonomous system acoustic sensors and processors" organized by Juan Arvelo. There were a total of 21 invited papers and 18 contributed papers presented at the meeting.

In Portland the special sessions were "Detection and classification of underwater targets" organized by Patrick Loughlin, Maya Gupta, and Jack McLaughlin, "Poroelastic materials: models, bounds, and parameter estimation" organized by Max Deffenbaugh, and "Pattern recognition in acoustic signal processing" organized by Grace Clark. There were 22 invited papers and 38 contributed papers presented at the meeting. In addition, David Chambers and Ning Xiang gave an invited talk on "Hot topics in signal processing." The Best Paper by a Young Presenter Award was given to Hui Ou from the University of Hawaii at Manoa for the paper "Automatic classification of underwater targets using fuzzy-cluster-based wavelet signatures."

We are proud that Jim Candy was selected for the Helmholtz-Rayleigh Interdisciplinary Silver Medal in Signal Processing in Acoustics and Underwater Acoustics.

This year an election was held for the next Chair of the Signal Processing Technical Committee, term beginning after the Portland meeting. This was won by Lee Culver from the Applied Research Laboratory at Pennsylvania State University.

We wish to thank Edmund J. Sullivan and William Carey, our signal processing Associate Editors of the *Journal of the Acoustical Society of America*, and Jim Candy, our signal processing Associate Editor for JASA Express Letters for the selfless service to the Society and SPTC.

We appreciate David Havelock's efforts to maintain the SP web site, and Sean Lehman's work organizing the Gallery of Acoustics. Colin Jemmott has been our representative to the Student Council.
DAVID H. CHAMBERS
Chair 2006–2009
LEE CULVER
Chair 2009–2012

Speech Communication

The Speech Technical Committee (TC) supports the activities, meetings, publications, etc., for the largest technical area in the Society. This report covers the ASA meeting in Miami, Florida and Portland, Oregon. The current members of the TC are Augustine Agwuele, Jean E. Andruski, Patrice S. Beddor, Fredericka Bell-Berti, Tessa C. Bent, Suzanne E. Boyce, Ann R. Bradlow, Kate E. Bunton, Dani M. Byrd, Rebeka Campos-Astorkiza, Bruce R. Gerratt, Helen M. Hanson, Rachel Frush Holt, Susan G. Guion, Diane Kewley-Port, Jody E. Kreiman, Jelena Krivokapic, Andrew J. Lotto, Conor T. McLennan, Shrikanth S. Narayanan, Dwayne Paschall, Christine H. Shadle, Rahul Shrivastav, Rajka Smiljanic, Mitchell S. Sommers, Brad H. Story, Joan E. Sussman, Scott L. Thomson, Jennel C. Vick and Zhaoyan Zhang. Many of these members are new to the TC. I want to welcome them and thank them for their support.

Other STC members who assisted us by serving on committees were Freddie Bell-Berti and Diane Kewley-Port-Executive Council, Jody Kreiman-Membership Committee, Shrikanth Narayanan-ASACOS, Anders Loqvist -Medals and Awards Committee, Corine Bickley and Ann Syrdal-Standards Committee on TTS Systems, Chritian Stilp-Student Representative, and Brad Story who maintains our web page. The continuing Associate Editors are: Speech Production—David Berry, Anders Lofqvist and Christine Shadle, Speech Perception—Kenneth Grant, Paul Iverson, Mitchell Sommers, Joan Sussman, Rochelle Newman and Allard Jongman; and Speech Processing—Douglas O'Shaughnessy. I am happy to welcome two new associate editors, Mark Hasegawa-Johnson and Shrikanth Narayanan, in the area of speech processing.

The two ASA meetings went very smoothly this year because of the particular efforts of several members. First, we are grateful to our Technical Program Organizing Meeting (TPOM) representatives, who sorted papers, arranged the technical programs, and determined presentation rooms. They were Catherine Rogers and Stefan Frisch for the Miami meeting and Susan Guion and Melissa Redford for the Portland meeting. Second, we greatly appreciate the effort of the coordinators for student judging at the meetings: Terry Gottfried and Corine Bickley in Miami and Terry Gottfried in Portland.

We applaud Winifred Strange for receiving the Silver Medal in Speech Communication at the Portland meeting. We also are very happy to welcome Dave Berry, Suzanne Boyce, Ann Bradlow, Bruce Gerratt, Frank Guenther, Keith Kluender, Rich McGowan, Luc Mongeau, D. Lloyd Rice, Christine Shadle, Ann Syrdal, and Doug Whalen as new fellows of the ASA in the Speech Communication TC.

Student Activities

In our continuing effort to promote student participation in ASA meetings, the Speech Technical Committee sponsored two student activities at each meeting, a competition with a cash award for best student presentation and an evening reception in Miami and Portland. The student reception, which is joint with other technical committees, is intended to allow students to meet more senior ASA members informally. The receptions were well attended. The student papers were judged by STC members and the winners were awarded $300 for first prize and $200 for second prize. I was not able to announce the winners for the Acoustics '08 in last years reports, so I will include them here. The first place winner was Carol Mettigan, University of London, for "Investigating the Perception of Noise Vocoded Speech and Individual Difference Approach." The second place winner was Thomas Hueber, ESPCI Telecom Paris, for "Ultrasound-based Silent Speech Interface." At the Miami meeting, the first place winner was Elizabeth Hunt, MIT, for "Acoustic characteristics of glides /j/ and /w/: Interactions with phonation. The second place winner was Joseph Toscano, University of Iowa, for "Online processing of acoustic cues used in speech perception: Comparing statistical and neural network models." At the Portland meeting, the first place winner was first place: Matias Zanartu, Purdue University, for "An impedance-based inverse filtering scheme for glottal coupling." The second place winner was Rachel Miller, University of California at Riverside, for "Investigating perceptual measures of speech alignment: Do AXB matching tasks make the grade?"

A special thanks to all of the reviewers of the student papers.

Special Sessions, Special Reception and Special Workshop

To create stimulating and focused sessions, we sponsor special sessions every year, which focus on themes of interest to the speech community. In Miami, there were two special sessions: "A quantal transition: Ken Stevens in "retirement" chaired by Helen Hanson and "James J. Jenkins: Teacher, mentor, researcher" chaired by Winifred Strange. There was a special joint reception held for Ken Stevens and James Jenkins. In Portland, there were three special sessions: "Vowel inherent spectral change" co-chaired by Geoffrey Morrison and Peter Assmann, "Source/filter interaction

in biological sound production" chaired by Ingo Titze, "Cognitive and psychological processes in speech perception" chaired by Amee Shah.

The Second ASA Special Workshop on Speech: Speech Cross-language Speech Perception and Variations in Linguistic Experience was held in honor of Winifred Strange at the end of the Portland meeting. The organizers for this workshop were Catherine Best, Ann Bradlow, Susan Guion and Linda Polka. The keynote address was given by Winifred Strange and the closing remarks were given by James Jenkins. Invited speakers were Catherine Best, Louis Goldstein, Michael Tyler & Hosung Nam; Ocke Bohn and Linda Polka; Laura Bosch and Marta Ramon-Casas; Susanne Curtin, Janet Werker & Krista Byers-Heinlein; Patricia Kuhl; Murray Munro; Valerie Shafer; Megha Sundara and Adrienne Scutellaro; Andrea Weber; and Reiko Akahane-Yamada (unable to attend due to flu epidemic in Japan). Additionally, there were posters, many of them given by students. The workshop was a great success.

I would like to thank the Technical Council for all of their support, for getting up early for our breakfast meeting held at each conference and for being such an agreeable group with which to work.

CAROL ESPY-WILSON
Chair

Structural Acoustics and Vibrations

The Structural Acoustics and Vibration Technical Committee (SAVTC) saw significant improvement in activities in the past year. At the 155th ASA/EAA Joint Meeting in Paris, France, the SAVTC sponsored nine special sessions: 1) General Topics in Structural Acoustics and Vibrations organized by Wolfgang Kropp and Sean F. Wu; 2) Vibration and Radiation from Complex Structural Systems organized by David Feit and Jean-Louis Guyader; 3) Source Characterization in Structure Borne Noise Problems organized by Evan Davis and Charles Pezerat; 4) Acoustic Imaging in Confined Space organized by Earl G. Williams and Alexandre Garcia; 5) Efficient Boundary Element Methods organized by Ramani Duraiswami and Lothar Gaul; 6) Fluid-Structure Interaction organized by Noureddine Atalla, Vicente Cutanda Henriquez, and Stefan Schneider; 7) Distributed Active Noise and Vibration Control organized by Kenneth Cunefare and Manuel Collet; 8) Ground Vehicle Noise and Vibration organized by Donald B. Bliss and Paul De Vos; and 9) Active Noise Control: New Strategies and Innovative Concepts organized by Alain Berry and Marie-Annick Galland. The TPOM representative was Yves Berthelot. The attendees were exceptionally large, most of which were from European countries however. The Best Students Papers Competition was conducted by James Phillips and the winners were René Christensen (1st Place, from Denmark), Danielle Moreau (tie 2nd Place, from Australia) and Kerem Ege (tie 2nd Place, from France). Mauro Pierucci gave a summary of the guidelines of the ASA Medals and Awards Committee. Allan D. Pierce reported that the revision of PACS SVATC numbers submitted by the subcommittee chaired by Earl Williams looked good, and wanted to go through all PACS and move all papers on solids and vibration, elastic waves, and porous media to SAV. Jerry Ginsberg reported that Rona Ginsberg is currently organizing *101 Careers in Acoustics*. Those who have stories or helpful information in this regard should contact Rona Ginsberg.

At the 156th ASA Meeting in Miami, Florida, SAVTC sponsored five special sessions on: 1) Emerging Applications of Structural Acoustics in Energy and Power Generation sponsored by Sean Wu and K. M. Li; 2) Wind Turbine Vibration and Sound Radiation (Joint with Engineering Acoustics) sponsored by Sean Wu; 3) Concepts of new vibration sensors (Joint with Engineering Acoustics) sponsored by Steve Shepard; 4) Computational Structural Acoustics sponsored by Kuangchen Wu; and 5) Vibro-acoustic Diagnosis and Prognosis of Complex Structures sponsored by Wen Li. The TPOM representative was Joe Cuschieri. James Phillips also conducted the Best Student Papers Competition and the winners were Jon La Follett (1st Place, from Washington State University) and Micah Shepherd (2nd Place, from Brigham Young University). At this meeting, the Committee discussed Technical Initiatives and Sean Wu, Allan D. Pierce and Jerry H. Ginsberg volunteered to work on a proposal on "Develop computer and actual test demonstration apparatus for illustrating structural acoustics and vibration phenomena," that can be shown to high school students. Sabih Hayek volunteered to prepare a proposal on "Case studies of acoustic related catastrophes and damages." Jerry Rouse volunteered to update the SAV TC web page. On behalf of the Student Council, Jon La Follett reported that it is very important for student members to have senior members to talk to them. Courtney commented that Allan Pierce and other Editors talked to student members about paper submissions to refereed journals, and it was very well received.

At the 157th ASA Meeting in Portland, Oregon, SAVTC sponsored five special sessions: 1) Distinguished Lecture on Structural Acoustics and Vibrations organized by Sean Wu; 2) Emerging Applications of Structural Acoustics in Energy and Power Generation organized by Sean Wu; 3) Concepts of new vibration sensors organized by Daniel W. Warren; 4) Computational Structural Acoustics organized by Kuangchen Wu; and 5) Vibro-acoustic Diagnosis and Prognosis of Complex Structures organized by Wen Li. The TPOM representative was Philip Marston. James Phillips conducted the Best Student Papers Competition and the winners were Nicholas O'Donoghue (1st Place, from Carnegie Mellon University) and Na Zhu (2nd Place, from Wayne State University). At this meeting, a sub-committee was formed to review and amend the "scope" of the SAV TC and the results are to be reported at the next SAV TC to be held in San Antonio. The members of this sub-committee include Karl Grosh (Chair), Earl Williams, Dave Feit, Jerry Ginsberg. Brian Thornock reported that two students expressed interest in the position of student representative to the SAV TC and that the student council will select the representative. As of May 1, 2009, there are 58 members in the SAVTC. At the end of the meeting, Sean Wu nominated eight additional members for the SAVTC.

SEAN WU
Chair

Underwater Acoustics

The Fall 2008 meeting at Miami Doral Resort was a very good meeting for the UWTC. Our very own Harry DeFerrari of the University of Miami was the General Chair, with Altan Turgut of NRL-DC serving as the UWTC TPOM representative. Steve Finette of NRL-DC gave a well attended UW Hot Topics talk on "Uncertainty in Ocean Acoustics." Two UW special sessions were organized, one by Kyle Becker of ARL-PSU on "Acoustics of harbors, ports and shallow navigable waterways," and a second by Claire Debever of SIO on "Robust array processing." Both special sessions were very well organized and attended. In terms of society recognition for our members, the meeting was a good one, with the announcements of society fellowship for John Fawcett of DRDC and the award of the IOA A.B. Wood medal to Karim Sabra of Georgia Tech. Our student paper award winners from the meeting were Aubrey Espana from Washington State with her talk "Excitation of low-frequency modes of solid cylinders by evanescent and ordinary propagating waves" and Lin Wan from Georgia Tech with the talk "Three-dimensional spatial coherence measurements: Vertical, longitudinal/transverse horizontal coherence"

The Spring 2009 meeting in Portland, OR, was also very good meeting for the TC. Many thanks to DJ Tang of APL-UW who served as our representative to the paper sorting meeting. We had a large number of successful special sessions, as well as a very well organized and attended named session for Ralph Goodman, organized by Jerry Caruthers and Ken Gilbert of the University of Mississippi and Steve Stanic of NRL Stennis. Our special sessions were "Waveguide invariant principles for active and passive sonars" organized by Altan Turgut and Lisa Zurk of Portland State, "Mid to high-frequency propagation and scattering with application to underwater communications" organized by Moshen Badiey of the University of Delaware and Dan Rouseff of APL-UW, "Physics-based undersea clutter model verification and validation" organized by Juan Arvelo of APL-JHU, and Tim Stanton and Ken Foote of WHOI, and "Monostatic and bistatic detection of elastic objects near boundaries: methodologies and tradeoffs" organized by Mario Zampolli of NURC and TNO and Karim Sabra. Three new fellows in our TC were named, Martin Siderius of Portland State, Aaron Thode of SIO, and Lisa Zurk of Portland State. In addition Martin was named the winner of the Medwin Award in our sister TC, Acoustical Oceanography.

The Underwater Acoustics Technical Committee has the pleasure to announce Lisa Zurk's election as chair of the UWTC for the 2009–2012 term. We are confident that the TC is in excellent hands. We would also like to thank Lisa for her efforts as General Chair of the Portland meeting, which was an outstanding success with approximately 1200 papers submitted.

The UWTC would like to highlight the service of the following technical committee members serving on ASA committees and councils: Eric Thorsos of APL-UW who is serving a second term on the Medals and Awards committee, Henrik Schmidt of MIT who is likewise serving for a

second term on the Membership committee, Peter Dahl of APL-UW who is serving on the Executive Council, Bob Drake of NUWC who is the UWTC representative on the ASACOS, and Megan Ballard of ARL-PSU who is our Student Council representative. We also congratulate former AOTC chair Jim Miller of URI on his election to the Executive Council. Finally we would like to congratulate George Frisk of FAU for his election to the post of President-Elect for the 2009–2010 term.

KEVIN LE PAGE

Chair 2006–2009

Calendar of Meetings and Congresses

2009

06–10 September	Brighton, UK. InterSpeech 2009 Conference. Web:http://www.interspeech2009.org
07–11 September	Dresden, Germany. 9th International Conference on Theoretical andComputational Acoustics. Web: http://ictca2009.com
14–18 September	Kyoto, Japan. 5th Animal Sonar Symposium. Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–17 September	Koriyama, Japan. Autumn Meeting of the Acoustical Society of Japan. Web: http://www.asj.gr.jp/index-en.html
19–23 September	Rome, Italy. IEEE 2009 Ultrasonics Symposium. E-mail:pappalar@uniroma3.it
21–23 September	Beijing, China. Western Pacific Acoustics Conference (WESPAC). Web: http://www.wespacx.org
23–25 September	Xi'an, China. Pacific Rim Underwater Acoustics Conference (PRUAC). E-mail: lfh@mail.ioa.ac.cn
23–25 September	Cádiz, Spain. TECNIACUSTICA'09. Web: www.-sea-acustica.es
05–07 October	Tallinn, Estonia. International Conference on Complexity of Nonlinear Waves. Web: www.ioc.ee/cnw09
18–21 October	New Paltz, NY, USA. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2009).
26–28 October	Edinburgh, UK. Euronoise 2009. Web: http://www.euronoise2009.org.uk
26–30 October	San Antonio, TX, USA. 158th Meeting of the Acoustical Society of America. Web: http://asa.aip.org/meetings.html
05–06 November	Dübendorf, Switzerland. Swiss Acoustical Society Autumn Meeting. Web: www.sga-ssa.ch
18–20 November	Kyoto, Japan.. 30-th Symposium on Ultrasonics Electronics. Web: www.use-jp.org/USE2009/en/index.html
23–25 November	Adelaide, Australia. Australian Acoustics Society National Conference. Web: www.acoustics.asn.au/joomla

2010

08–11 March	Berlin, Germany. Meeting of the German Association for Acoustics DAGA 2010. Web: www.daga-tagung.de/2010
15–19 March	Dallas, TX, USA. International Conference on Acoustics, Speech, and Signal Processing. Web: http://icassp2010.org
19–23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society of America and Noise Con 2010. Web: http://asa.aip.org/meetings.html
27–30 April	Ghent, Belgium. Institute of Acoustis/Belgian Acoustical Association Joint Meeting. Web: www.ioa.org.uk/viewupcoming.asp
09–11 June	Aalborg, Denmark. 14th Conference on Low Frequency Noise and Vibration. Web: http://lowfrequency2010.org
13–16 June	Lisbon, Portugal. INTERNOISE2010. Web: www.internoise2010.org
23–27 August	Sydney, Australia. International Congress on Acoustics 2010. Web: www.ica2010sydney.org
14–18 September	Kyoto, Japan. 5th Animal Sonar Symposium. Web: http://cse.fra.affrc.go.jp/akamatsu/AnimalSonar.html
15–18 September	Ljubljania, Slovenia. Alp-Adria-Acoustics Meeting joint with EAA. E-mail: mirko.cudina@fs.uni-lj.si
26–30 September	Makuhari, Japan. Interspeech 2010 - ICSLP. Web: www.interspeech2010.org
14–16 October	Niagara-on-the Lake, Ont., Canada. Acoustics Week in Canada. Web: http://caa-aca.ca/E/index.html
11–14 October	San Diego, California, USA. IEEE 2010 Ultrasonics Symposium. E-mail: bpotter@vectron.com
15–19 November	Cancun, Mexico. Second Pan-American/Iberian Meeting on Acoustics (Joint meeting of the Acoustical Society of America, Iberoamerican Congress of Acoustics, Mexican Congress on Acoustics. Web: http://asa.aip.org/meetings.html
19–20 November	Brighton, UK. Reproduced Sound 25. Web: www.ica.org.uk/viewupcoming.asp

2011

27 June–01 July	Aalborg, Denmark. Forum Acusticum 2011. Web: www.fa2011.org
27–31 August	Florence, Italy. Interspeech 2011. Web: www.interspeech2011.org
05–08 September	Gdansk, Poland. International Congress on Ultrasonics. Web: TBA
04–07 September	Osaka, Japan. Internoise 2011. Web: TBA

2013

02–07 June	Montréal, Canada. 21st International Congress on Acoustics (ICA 2013) (Joint meeting: International Congress on Acoustics, Acoustical Society of America, Canadian Acoustical Association). Web: www.ica2013montreal.org

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

The Technical and Administrative Committees listed below have been appointed by the President with the approval of the Executive Council. These appointments, with such changes as may be made by the President from time to time, will be in effect until the Spring meeting of the Society in 2010.

Technical Committees 2009–2010

Acoustical Oceanography

Martin Siderius, Chair to 2010

Term to 2012

Mohsen Badiey
Michael J. Buckingham
Dezhang Chu
John A. Colosi
Christian de Moustier
Stan E. Dosso
Kenneth G. Foote
D. Vance Holliday
Andone C. Lavery
Zoi-Heleni Michalopoulou
Jeffrey A. Nystuen
David R. Palmer
Simon D. Richards
Timothy K. Stanton
Aaron M. Thode
Mark V. Trevorrow

Term to 2011

Kelly J. Benoit-Bird
Daniela Di Iorio

Acoustical Oceanography (*Continued.*)

Gerald L. D'Spain
Timothy F. Duda
David M. Farmer
Gary J. Heald
Jean-Pierre Hermand
Paul C. Hines
David P. Knobles
Timothy G. Leighton
Ying-Tsong Lin
Jennifer Miksis-Olds
Daniel Rouseff
Emmanuel K. Skarsoulis
Dajun Tang
Altan Turgut
Joseph D. Warren
Thomas C. Weber
Term to 2010
Kyle M. Becker
N. Ross Chapman
Grant B. Deane
Brian D. Dushaw
Matt A. Dzieciuch
Peter Gerstoft
Oleg A. Godin
John K. Horne
Bruce M. Howe
James F. Lynch
Anthony P. Lyons
Gopu Potty
Ralph A. Stephen
Alexandra I. Tolstoy
Kathleen E. Wage
Peter F. Worcester
Ex officio:
Peter F. Worcester, member of Medals and Awards Committee
Kenneth G. Foote, member of Membership Committee
Anthony P. Lyons, member of ASACOS
James Traer, member of Student Council

Animal Bioacoustics

David K. Mellinger, Chair to 2012

Term to 2012
Elizabeth F. Brittan-Powell
William Cummings
Seth S. Horowitz
Dorian S. Houser
Jennifer L. Miksis-Olds
Peter M. Narins
Susan E. Parks
Hiroshi Riquimaroux
Andrea M. Simmons
Term to 2011
Judith C. Brown
Sheryl L. Coombs
Edmund R. Gerstein
Charles R. Greene
Darlene R. Ketten
Bertel Mohl
Cynthia F. Moss
Paul E. Nachtigall

Animal Bioacoustics (*Continued.*)

Peter L. Tyack
David G. Zeddies
Term to 2010
Micheal L. Dent
James J. Finneran
Mardi C. Hastings
Charlotte W. Kotas
Marc O. Lammers
David A. Mann
Marie A. Roch
Gary J. Rose
Joseph A. Sisneros
Ex officio:
James M. Simmons, member of Medals and Awards Committee
Richard R. Fay, member of Membership Committee
Ann E. Bowles, member of ASACOS
Mary E. Bates, member of Student Council

Architectural Acoustics

Brandon D. Tinianov, Chair to 2010

Term to 2012
Nils-Ake Andersson
C. Walter Beamer, IV
Leo L. Beranek
Sergio Beristain
Jim X. Borzym
David T. Bradley
David Braslau
Todd L. Brooks
Courtney B. Burroughs
Paul T. Calamia
Alexander U. Case
William J. Cavanaugh
Dan Clayton
Jessica S. Clements
David A. Conant
Damian J. Doria
John Erdreich
Robin S. Glosemeyer
Timothy E. Gulsrud
Byron W. Harrison
Robert D. Hellweg
Murray R. Hodgson
Ian B. Hoffman
Jin Yong Jeon
Jian Kang
Bertram Y. Kinzey, Jr.
Mendel Kleiner
Alexis D. Kurtz
Timothy W. Leishman
Jerry G. Lilly
Edward L. Logsdon
Peter A. Mapp
David E. Marsh
Gregory A. Miller
Matthew A. Nobile
Christian Nocke
Bruce C. Olson
Cornelius H. Overweg
Richard J. Peppin

Stephen D. Pettyjohn
Scott D. Pfeiffer
Norman H. Philipp
James E. Phillips
Jens Holger Rindel
Carl J. Rosenberg
Kenneth P. Roy
Erica Ryherd
Hiroshi Sato
Melvin L. Saunders
Ron Sauro
Paul D. Schomer
Kevin P. Shepherd
Yasushi Shimizu
Gary W. Siebein
Christopher A. Storch
Jason E. Summers
Louis C. Sutherland
Jiri Tichy
Nancy S. Timmerman
Gregory C. Tocci
Lily M. Wang
Alfred C.C. Warnock
George P. Wilson
Ning Xiang

Term to 2011

Wolfgang Ahnert
Robert B. Astrom
Seth E. Bard
Christopher N. Blair
John S. Bradley
Christopher N. Brooks
Angelo J. Campanella
Zhixin Chen
Quinsan Ciao
Robert C. Coffeen
Russell A. Cooper
David B. Copeland
Peter D'Antonio
Felicia M. Doggett
Bill Dohn
Timothy J. Foulkes
Ronald R. Freiheit
Richard D. Godfrey
Matthew V. Golden
Tyrone Hunter
Clare M. Hurtgen
J. Christopher Jaffe
Yun Jing
Thomas E. Kaytt
Jeffrey P. Kwolkoski
Stephen J. Lind
David Lubman
Ralph T. Muehleisen
Matthew L. Nickerson
Boaz Rafaely
Jack E. Randorff
Jonathan Rathsam
H. Stanley Roller
Steven R. Ryherd
B. Schulte-Fortkamp
Kerrie G. Standlee

Architectural Acoustics (Continued.)

Noral D. Stewart
Michelle C. Vigeant
Michael Vorländer

Term to 2010

Russell L. Altermatt
Russ Berger
Warren E. Blazier
Joseph F. Bridger
Norm Broner
Bennett M. Brooks
Steven M. Brown
Richard H. Campbell
Andrew C. Carballeira
Emily L. Cross
F. M. del Solar Dorrego
Erin L. Dugan
Edward C. Duncan
Jesse J. Ehnert
Ronald T. Eligator
Michael Ermann
Adam R. Foxwell
Ronald R. Freiheit
Klaus Genuit
Kenneth W. Good, Jr.
Lewis S. Goodfriend
Bradford N. Gover
Pamela J. Harght
Mark A. Holden
K. Anthony Hoover
Jerald R. Hyde
Jodi Jacobs
Basel H. Jurdy
David W. Kahn
Brian F.G. Katz
Michael P. Kerr
Brian J. Landsberger
Martha M. Larson
Jonathan S. Leonard
Gary S. Madaras
Benjamin E. Markham
David L. Moyer
Paul B. Ostergaard
Dennis A. Paoletti
Stephen W. Payne
Richard F. Riedel
Lauren Ronsse
Hari V Savitala
Benjamin C. Seep
Neil A. Shaw
J. Michael Spencer
Rose Mary Su
Jeff D. Szymanski
Richard H. Talaske
Michelle Vigeant
Alicia J. Wagner
Ewart A. Wetherill
George E. Winzer

Ex-officio:

Dana S. Hougland, member of Medals & Awards Committee
Ronald R. Freiheit, member of Membership Committee
Angelo J. Campanella, member of ASACOS
Lauren Ronsse, member of Student Council

Biomedical Ultrasound/Bioresponse to Vibration

Jeffrey A. Ketterling, Chair to 2011

Term to 2012
Iwaki Akiyama
John S. Allen
Azzdine Y. Ammi
Paul E. Barbone
Shira L. Broschat
Jean-Yves Chapelon
Charles C. Church
Gregory Clement
Diane Dalecki
Jan R. D'Hooge
Floyd Dunn
Stanislav Emelianov
Guillaume Haiat
Wayne E. Kreider
Franck Levassort
Subha Maruvada
Douglas L. Miller
Claire Prada
Purnima Ratilal
Shin-ichiro Umemura
Michel Versluis
Kendall R. Waters
Term to 2011
Michalakis A. Averkiou
Timothy A. Bigelow
Parag V. Chitnis
Saurabh Datta
Sara Davis
Caleb H. Farny
Jonathan Mamou
Stuart B. Mitchell
Neil R. Owen
Oleg A. Sapozhnikov
Armen Sarvazyan
Mark E. Schafer
Eleanor P. Stride
Jahan Tavakkoli
Gail R. ter Haar
Matthew W. Urban
Keith A. Wear
Mark S. Wochner
Term to 2010
Constantin-C. Coussios
Sheryl M. Gracewski
Seyed H.R. Hosseini
Ronald E. Kumon
Pascal P. Laugier
T. Douglas Mast
Thomas J. Matula
James A. McAteer
Robert J. McGough
James G. Miller
Todd W. Murray
Michael L. Oelze
Ronald A. Roy
Preston S. Wilson
Suk Wang Yoon
Evgenia A. Zabolotskaya

Biomedical Ultrasound/Bioresponse to Vibration (Continued.)

Ex officio:
Lawrence A. Crum, member of the Medals and Awards Committee
Christy K. Holland, member of Membership Committee
Peter J. Kaczkowski, member of ASACOS
Vera A. Khokhlova, member of ASACOS
Lucie Somaglino, member of Student Council

Engineering Acoustics

Michael V. Scanlon, Chair to 2012

Term to 2012
Kim C. Benjamin
Richard D. Costley
Stanley L. Ehrlich
Gary W. Elko
Guillermo C. Gaunaurd
Thomas R. Howarth
Dehua Huang
Lixi Huang
Sung-Hwan Ko
William J. Marshall
Victor Nedzelnitsky
Scott P. Porter
P. K. Raju
Stephen C. Thompson
Daniel M. Warren
James E. West
Allan J. Zuckerwar
Term to 2011
Steven R. Baker
David A. Brown
Stephen C. Butler
Robert D. Corsaro
Robert M. Drake
Stephen E. Forsythe
Brian H. Houston
W. Jack Hughes
Robert M. Koch
Christopher C. Lawrenson
L. Dwight Luker
Arnie L. Van Buren
Kenneth M. Walsh
Joseph F. Zalesak
Term to 2010
Mahlon D. Burkhard
Fernando Garcia-Osuna
Charles S. Hayden
Jan F. Lindberg
Elizabeth A. McLaughlin
Alan Powell
Roger T. Richards
Kenneth D. Rolt
Neil A. Shaw
James F. Tressler
Ex officio:
Kim C. Benjamin, member of Medals and Awards Committee
Stephen C. Thompson, member of Membership Committee
Mahlon D. Burkhard, member of ASACOS
Scott P. Porter, member of Student Council

Musical Acoustics

Paul A. Wheeler, Chair to 2011

Term to 2012
Rolf Bader
Xavier Boutillon
Jonas Braasch
Murray D. Campbell
Andrey R. DaSilva
Nicholas J. Giordano
William M. Hartmann
William L. Martens
Andrew C. H. Morrison
James M. Pyne
Sten O. Ternstrom
Shigeru Yoshikawa
Term to 2011
James W. Beauchamp
George A. Bissinger
Annabel J. Cohen
James P. Cottingham
Evan B. Davis
Diana Deutsch
Uwe J. Hansen
Peter L. Hoekje
James H. Irwin
Ian M. Lindevald
Robert W. Pyle
Gary P. Scavone
Brad H. Story
George Tzanetakis
Christopher E. Waltham
Term to 2010
R. Dean Ayers
Judith C. Brown
Courtney B. Burroughs
John R. Buschert
Joel Gilbert
Thomas M. Huber
Bozena Kostek
Barry Larkin
Daniel O. Ludwigsen
Thomas R. Moore
Thomas D. Rossing
David B. Sharp
Julius O. Smith
William J. Strong
Joe Wolfe
Ex officio:
Uwe J. Hansen, member of Medals and Awards Committee
James P. Cottingham, member of Membership Committee
Diana Deutsch, member of ASACOS
to be appointed, member of Student Council

Noise

Nancy S. Timmerman, Chair to 2012

Term to 2012
Sergio Beristain
Susan B. Blaeser
Bennett M. Brooks
Ilene J. Busch-Vishniac

Noise (Continued.)

Angelo J. Campanella
William J. Cavanaugh
Gilles A. Daigle
Patricia Davies
Damian J. Doria
Connor R. Duke
Jesse J. Ehnert
Tony F. W. Embleton
John Erdreich
David J. Evans
Hugo Fastl
Bradford N. Gover
Robert D. Hellweg
Lixi Huang
Tyrone Hunter
Robert C. Kull
William W. Lang
Richard H. Lyon
Luigi Maffei
Alan H. Marsh
Ralph T. Muehleisen
William J. Murphy
Kenneth P. Roy
Erica E. Ryherd
Brigitte Schulte-Fortkamp
Kevin P. Shepherd
Scott D. Sommerfeldt
Kerrie G. Standlee
Michael R. Stinson
Term to 2011
Seth E. Bard
Elliott H. Berger
Ann E. Bowles
Frank H. Brittain
Steven M. Brown
Mahlon D. Burkhard
Robert D. Collier
Lawrence S. Finegold
Samir N. Y. Gerges
Richard D. Godfrey
Matthew V. Golden
Murray R. Hodgson
Jerry G. Lilly
Stephen J. Lind
David Lubman
George A. Luz
Matthew L. Nickerson
Matthew A. Nobile
Richard J. Peppin
Robert A. Putnam
Jack E. Randorff
Stephen I. Roth
Paul D. Schomer
Michelle E. Swearingen
Brandon D. Tinianov
Gregory C. Tocci
Lily M. Wang
Term to 2010
Martin Alexander
Brian E. Anderson
Keith Attenborough
John P. Barry

Noise (Continued.)

Leo L. Beranek
Arno S. Bommer
Dick B. Botteldooren
Giovanni Brambilla
James O. Buntin
John C. Burgess
Jim R. Cummins
Kenneth A. Cunefare
Paul R. Donavan
Andre Fiebig
Ronald R. Freiheit
Klaus Genuit
David C. Haser
Gerald C. Lauchle
George C. Maling
Thomas R. Norris
John P. Seiler
Noral D. Stewart
Louis C. Sutherland
Jiri Tichy
D. Keith Wilson
Ning Xiang
Yuzhen Yang
Ex officio:
Brigitte Schulte-Fortkamp, member of Medals and Awards Committee
Bennett M. Brooks, member of Membership Committee
Richard J. Peppin, member of ASACOS
Cole R. Duke, member of Student Council

Physical Acoustics

Ronald A. Roy, Chair to 2010

Term to 2012
John M. Allen
Anthony A. Atchley
James P. Chambers
Charles C. Church
Kenneth G. Foote
Veerle M. Keppens
Robert G. Leisure
Philip L. Marston
Stuart B. Mitchell
Andrew A. Piacsek
Tyrone M. Porter
Peter H. Rogers
James M. Sabatier
D. Keith Wilson
Evgenia A. Zabolotskaya
Term to 2011
Robin O. Cleveland
Lawrence A. Crum
E. Carr Everbach
Kenneth E. Gilbert
Robert A. Hiller
R. Glynn Holt
Bart Lipkens
Thomas J. Matula
Ralph T. Muehleisen
John S. Stroud
Richard L. Weaver
Preston S. Wilson

Physical Acoustics (Continued.)

Term to 2010
David T. Blackstock
David A. Brown
John A. Burkhardt
Kerry W. Commander
Bruce C. Denardo
Kent L. Gee
Logan E. Hargrove
Julian D. Maynard
Albert Migliori
James G. Miller
Lev A. Ostrovsky
Andrea Prosperetti
Neil A. Shaw
Victor W. Sparrow
Richard Stern
Michelle E. Swearingen
Roger M. Waxler
Ex officio:
E. Carr Everbach, member of Medals and Awards Committee
Steven L. Garrett, member of Membership Committee
Richard Raspet, member of ASACOS
Jon R. La Follett, member of Student Council

Psychological and Physiological Acoustics

Andrew J. Oxenham, Chair to 2011

Term to 2012
Jose I. Alcantara
Emily Buss
Frederick J. Gallun
Sridhar Kalluri
Gerald D. Kidd
Bernhard U. Seeber
Daniel J. Tollin
Term to 2011
Huanping Dai
Christian Lorenzi
Christophe D. Micheyl
Roy D. Patterson
Daniel Pressnitzer
Brian Roberts
Term to 2010
Sid P. Bacon
Qian-Jie Fu
Ruth Y. Litovsky
Robert A. Lutfi
Kim S. Schairer
Christopher Shera
Edward J. Walsh
Beverly A. Wright
Ex officio:
William M. Hartmann, member of the Medals and Awards Committee
Lynne A. Werner, member of Membership Committee
Brent W. Edwards, member of ASACOS
Dorea R. Ruggles, member of Student Council

Signal Processing in Acoustics

Richard Lee Culver, Chair to 2012

Term to 2012

James V. Candy
William M. Carey
David H. Chambers
Grace Clark
Geoffrey S. Edelson
Brian Ferguson
Charles F. Gaumond
Paul J. Gendron
Peter Gerstoft
William M. Hartmann
William S. Hodgkiss
Paul D. Hursky
Jens M. Meyer
James C. Preisig
Brian Rapids
Edmund J. Sullivan
Lisa M. Zurk

Term to 2011

Joseph A. Clark
David I. Havelock
Jean-Pierre Hermand
George E. Ioup
Juliette W. Ioup
Colin W. Jemmott
Sean K. Lehman
William L. Martens
Zoi-Heleni Michalopoulou
Natalia A. Sidorovskaya
Kevin B. Smith
David C. Waddington
Ning Xiang

Term to 2010

Max Deffenbaugh
Alireza A. Dibazar
Gary W. Elko
Kassiani Kotsidou
Patrick J. Loughlin
Alan W. Meyer
Daniel J. Sinder
David C. Swanson
Robert C. Waag
Preston S. Wilson
Lixue Wu

Ex officio:

William J. Carey, member of Medals and Awards Committee
Ning Xiang, member of Membership Committee
Charles F. Gaumond, member of ASACOS
Colin W. Jemmott, member of Student Council

Speech Communication

Carol Espy-Wilson, Chair to 2010

Term to 2012

Jean E. Andruski
Suzanne E. Boyce
Bruce R. Gerratt
Susan G. Guion
Jelena Krivokapic
Conor T. McLennan
Rajka Smiljanic
Mitchell S. Sommers

Brad H. Story
Joan E. Sussman
Scott L. Thomson
Jennel C. Vick

Term to 2011

Fredericka Bell-Berti
Ann R. Bradlow
Kate E. Bunton
Dani M. Byrd
Rebeka Campos-Astorkiza
Rachel Frush Holt
Zhaoyan Zhang

Term to 2010

Augustine Agwuele
Patrice S. Beddor
Tessa C. Bent
Helen M. Hanson
Diane Kewley-Port
Jody E. Kreiman
Andrew J. Lotto
Shrikanth S. Narayanan
Dwayne Paschall
Christine H. Shadle
Rahul Shrivastav

Ex officio:

Anders Lofqvist, member of Medals and Awards Committee
Jody E. Kreiman, member of Membership Committee
Shrikanth S. Narayanan, member of ASACOS
Christian Stilp, member of Student Council

Structural Acoustics and Vibration

Dean E. Capone, Chair to 2009

Term to 2012

Noureddine Atalla
Yves H. Berthelot
Donald B. Bliss
Gerard P. Carroll
David H. Chambers
Kenneth A. Cunefare
Joseph M. Cuschieri
David Feit
Kenneth G. Foote
Guillermo C. Gaunaurd
Joseph R. Gavin
Thomas L. Geers
Jerry H. Ginsberg
Gary M. Glickman
Karl Grosh
Steve A. Hambric
Sabih I. Hayek
Robert Hickling
Yun-Fan Hwang
Jon R. La Follett
Michael Leamy
Kai Ming Li
Wen L. Li
Robert S. Luttrell
Philip L. Marston
John J. McCoy
James G. McDaniel

Jon W. Mooney
Koososh Naghshineh
Andrew N. Norris
Stephen D. O'Regan
James E. Phillips
Allan D. Pierce
Mauro Pierucci
Donald G. Pray
Thomas D. Rossing
Jerry W. Rouse
Vyacheslav M. Ryaboy
Henry A. Scarton
W. Steve Shepard
Kevin P. Shepherd
Micah Shepherd
Scott D. Sommerfeldt
Hasson M. Tavossi
Lonny L. Thompson
Eric E. Ungar
Nicholas E. Vlahopoulos
Kuangcheng Wu

Term to 2011
Joel M. Garrelick
Peter C. Herdic
Teik C. Lim
Thomas J. Royston
Angie Sarkissian
Jeffrey S. Vipperman
Richard L. Weaver

Term to 2010
Jeffrey E. Boisvert
Stephen C. Conlon
Linda P. Franzoni
Robert C. Haberman
Rudolph Martinez
Koorosh Naghshineh
Donald G. Pray
Michael F. Shaw

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

Underwater Acoustics

Lisa M. Zurk, Chair to 2012

Term to 2012
Ralph N. Baer
John R. Buck
Michael J. Buckingham
David R. Dowling
Roger C. Gauss
David J. Goldstein
Kevin D. Heaney
Frank S. Henyey
Paul C. Hines
Marcia J. Isakson
Adrian D. Jones
James H. Miller
Michael B. Porter

John R. Preston
Purnima Ratilal
Karim G. Sabra
Hee Chun Song
Ralph A. Stephen
Alexander G. Voronovich
Kevin L. Williams

Term to 2011
Michael A. Ainslie
Juan I. Arvelo, Jr.
Pierre-Philippe Beaujean
Shira L. Broschat
Nicholas P. Chotiros
Richard L. Culver
Claire Debever
Geoffrey F. Edelmann
Peter Gerstoft
John H. Glattetre
Brian T. Hefner
Jean-Pierre Hermand
Charles W. Holland
Shawn F. Johnson
David P. Knobles
Bruce K. Newhall
John C. Osler
Kevin B. Smith
Alexandra I. Tolstoy
Joshua D. Wilson

Term to 2010
Kyle M. Becker
David C. Calvo
Jee Woong Choi
Christian P. de Moustier
Stan E. Dosso
Steven I. Finette
Kenneth G. Foote
David Fromm
Christohper H. Harrison
Zoi-Heleni Michalopoulou
Tracianne B. Neilsen
Robert I. Odom
Marshall H. Orr
Gregory J. Orris
James C. Preisig
Martin Siderius
William L. Siegmann
Jixun Zhou

Ex officio:
Eric I. Thorsos, member of Medals and Awards Committee
Henrik Schmidt, member of Membership Committee
Robert M. Drake, member of ASACOS
Megan Ballard, member of Student Council

Administrative Committees 2009–2010

Archives and History

Julian D. Maynard, Chair to 2010

Term to 2012
Jont B. Allen
David I. Havelock

Wesley L. Nyborg
Richard J. Peppin
William J. Strong
Term to 2011
Anthony A. Atchley
Leo L. Beranek
William J. Cavanaugh
Steven L. Garrett
Logan E. Hargrove
Allan D. Pierce
Victor W. Sparrow
Term to 2010
Carol Y. Espy-Wilson
E. Carr Everbach
William W. Lang
Richard Stern
David C. Swanson

Audit Committee

Michael R. Stinson, Chair to 2010

Term to 2011
Peter H. Dahl
Term to 2010
Scott D. Sommerfeldt

Books+

David L. Bradley, Chair to 2011

Term to 2012
Juan I. Arvelo
James V. Candy
Jerry H. Ginsberg
Term to 2011
James P. Cottingham
Jeffrey A. Nystuen
Neil A. Shaw
Term to 2010
Richard Stern
Brandon D. Tinianov
Ex officio:
Allan D. Pierce, Editor-in-Chief

College of Fellows Steering

Judy Dubno, Cochair to 2010
K. Anthony Hoover, Cochair to 2011

Term to 2012
Keith R. Kluender
Thomas J. Matula
Scott D. Sommerfeldt
Stephen C. Thompson
Term to 2011
Aaron M. Thode
Brandon D. Tinianov
Lisa M. Zurk
Term to 2010
Peter G. Cable
M. David Egan
Uwe J. Hansen

Diane Kewley-Port
Thomas D. Rossing
William J. Cavanaugh, ex officio as past Chair
Richard H. Lyon, ex officio as past Chair
Janet M. Weisenberger, ex officio as past Chair
Lauren Ronsse, ex officio, representative from Student Council

Education in Acoustics

Preston S. Wilson, Chair to 2012

Term to 2012
William A. Ahroon
Takayuki Arai
Juan I. Arvelo
Anthony A. Atchley
Fredericka Bell-Berti
Suzanne E. Boyce
Robert D. Celmer
Annabel J. Cohen
Gilles A. Daigle
Geoffrey F. Edelmann
E. Carr Everbach
Thomas B. Gabrielson
Steven L. Garrett
Kent L. Gee
Uwe J. Hansen
Katherine S. Harris
Elizabeth S. Ivey
Joie P. Jones
Amy T. Neel
P. K. Raju
Deborah M. Rekart
M. Roman Serbyn
Victor W. Sparrow
Emily A. Tobey
Wayne M. Wright
Term to 2011
David T. Blackstock
Courtney B. Burroughs
James P. Chambers
Robin O. Cleveland
Kenneth A. Cunefare
D. Michael Daly
Mary Florentine
R. Glynn Holt
Murray S. Korman
Luc Mongeau
Peggy B. Nelson
Tyrone M. Porter
Neil A. Shaw
Roger Waxler
James E. West
Douglas Wilcox
Term to 2010
George A. Bissinger
David A. Brown
Stanley A. Cheyne
Robert D. Collier
Lawrence A. Crum
Corinne M. Darvennes
Bruce C. Denardo

Education in Acoustics (Continued.)

Margaritis S. Fourakis
Carl K. Frederickson
Carole E. Gelfer
Daniel O. Ludwigsen
Sharon Y. Manuel
Philip L. Marston
Ralph T. Muehleisen
Andrew A. Piacsek
Thomas D. Rossing
Ronald A. Roy
Dawn R. Schuette
Scott D. Sommerfeldt
Robert A. Walkling

Ethics and Grievances

Janet M. Weisenberger, Chair to 2011

Term to 2012
William J. Cavanaugh
Gerald L. D'Spain
Barbara G. Shinn-Cunningham

International Research and Education

David T. Blackstock, Chair to 2010

Term to 2012
Mark F. Hamilton
Erica E. Ryherd
Michelle C. Vigeant
Suk Wang Yoon
Term to 2011
Keith Attenborough
Sergio Beristain
Vera A. Khokhlova
Brigitte Schulte-Fortkamp
Term to 2010
Malcolm J. Crocker
Samir N.Y. Gerges
Konstantin A. Naugolnykh
Oleg Sapozhnikov
Michael Vorländer

Investments

William A. Yost, Chair to 2010

Term to 2012
Mark F. Hamilton
Term to 2011
Anthony A. Atchley
William M. Hartmann
David Feit, Treasurer, ex officio

Medals and Awards

David L. Bradley, Chair to 2010

Term to 2012	
Lawrence A. Crum	Biomedical Ultrasound/Bioresponse to Vibration
E. Carr Everbach	Physical Acoustics

Medals and Awards (Continued.)

William M. Hartmann	Psychological and Physiological Acoustics
Brigitte Schulte-Fortkamp	Noise
Term to 2011	
Dana S. Hougland	Architectural Acoustics
Mauro Pierucci	Structural Acoustics and Vibration
James A. Simmons	Animal Bioacoustics
Eric I. Thorsos	Underwater Acoustics
Term to 2010	
Kim C. Benjamin	Engineering Acoustics
William M. Carey	Signal Processing in Acoustics
Uwe J. Hansen	Musical Acoustics
Anders Lofqvist	Speech Communication
Peter F. Worcester	Acoustical Oceanography

Meetings

Clark S. Penrod, Chair to 2011 and Fall 2009, San Antonio
Sergio Beristain, Fall 2010, Cancun
Harry A. DeFerrari, Fall 2008, Miami
Judy R. Dubno, Vice President-Elect
Samir N. Y. Gerges, Fall 2010, Cancun
Mardi C. Hastings, Spring 2010, Baltimore
Diane Kewley-Port, Vice President
Thomas J. Matula, Spring 2011, Seattle
Elaine Moran, ASA Office Manager, ex officio
Charles E. Schmid, Executive Director, ex officio
James E. West, Fall 2010, Cancun
Lisa M. Zurk, Spring 2009, Portland

Membership

Peter H. Rogers, Chair to 2012

Term to 2012	
James P. Cottingham	Musical Acoustics
Richard R. Fay	Animal Bioacoustics
Jody E. Kreiman	Speech Communication
Stephen C. Thompson	Engineering Acoustics
Lynne A. Werner	Psychological and Physiological Acoustics
Term to 2011	
Bennett M. Brooks	Noise
Courtney B. Burroughs	Structural Acoustics and Vibration
Christy K. Holland	Biomedical Ultrasound/Bioresponse to Vibration
Term to 2010	
Kenneth G. Foote	Acoustical Oceanography
Ronald R. Freiheit	Architectural Acoustics
Steven L. Garrett	Physical Acoustics
Henrik Schmidt	Underwater Acoustics
Ning Xiang	Signal Processing in Acoustics

Panel on Public Policy

Edward J. Walsh, Chair to 2010

Term to 2012
Mardi C. Hastings
David Lubman
William A. Yost
Term to 2011
William M. Hartmann

Panel on Public Policy

Sabih I. Hayek
Peggy B. Nelson
Victor W. Sparrow
Term to 2010
Angelo J. Campanella
Lawrence A. Crum
Ellen S. Livingston
James H. Miller
Mark F. Hamilton, ex officio as immediate Past President
Diane Kewley-Port, ex officio as Vice President
Charles E. Schmid, ex officio as Executive Director
Paul D. Schomer, ex officio as Standards Director

Prizes and Special Fellowships

Anthony A. Atchley, Chair to 2011

Term to 2012
Fredericka Bell-Berti
Charles C. Church
Term to 2011
Uwe J. Hansen
Term to 2010
Judy R. Dubno
Wayne M. Wright

Public Relations

Andrew A. Piascek, Chair to 2012

Term to 2012
Ann E. Bowles
Geoffrey F. Edelmann
Paul D. Hursky
Jack E. Randorff
Lora J. Van Uffelen
Kathleen E. Wage
Term to 2011
William M. Carey
Katherine H. Kim
Bart Lipkens
Ellen S. Livingston
Brigitte Schulte-Fortkamp
Term to 2010
Kelly J. Benoit-Bird
Diana Deutsch
E. Carr Everbach
Christy K. Holland
Brenda L. Lonsbury-Martin
James H. Miller
Joe W. Posey
Barbara Shinn-Cunningham
Stephen C. ThompsonAllan D. Pierce, Editor-in-Chief, ex officio
Elaine Moran, ASA Office Manager, ex officio
Charles E. Schmid, Executive Director, ex officio
Thomas D. Rossing, Echoes Editor, ex officio

Publication Policy

Brenda L. Lonbsury-Martin, Chair to 2012

Term to 2012
David I. Havelock

Publication Policy (Continued.)

Andrew N. Norris
D. Keith Wilson
Term to 2011
James F. Lynch
Philip L. Marston
Term to 2010
Charles C. Church
Mardi C. Hastings
George V. Frisk, President-Elect, ex officio
Allan D. Pierce, Editor-in-Chief, ex officio

Regional Chapters

Juan I. Arvelo, Cochair to 2011
Erica E. Ryherd, Cochair to 2012

Kent L. Gee	Brigham Young Univ. Student Chapter
Angelo J. Campanella	Central Ohio
Robert M. Keolian	Central Pennsylvania
Sharon Hepfner	Cincinnati
Steven A. Grigoletti	Columbia College Chicago Student Chapter
Rebecca Mercuri	Delaware Valley
Catherine L. Rogers	Florida
Shaun D. Anderson	Georgia Institute of Tech. Student Chapter
Timothy J. Foulkes	Greater Boston
Robert D. Celmer	Univ. of Hartford Student Chapter
Robert C. Coffeen	Univ. of Kansas Student Chapter
Neil A. Shaw	Los Angeles
Hari S. Paul	Madras, India
Richard F. Riedel	Metropolitan New York
Sergio Beristain	Mexico City
William V. Slaton	Mid-South
David A. Brown	Narragansett
Lauren Ronsse	Univ. of Nebraska Student Chapter
George A. Bissinger	North Carolina
Peter F. Assmann	North Texas
James R. Angerer	Northwest
David Lubman	Orange County
Nicholas A. Block	Purdue University Student Chapter
Paul A. Baxley	San Diego
David Braslau	Upper Midwest
Jeffrey Dunne	Washington, D.C.
Thomas M. Disch	Wisconsin

Preston S. Wilson, Chair, Education in Acoustics, ex officio
David Feit, Treasurer, ex officio
Cole Duke, Student Council representative, ex officio

Rules and Governance Committee

William M. Hartmann, Chair to 2011

Term to 2012
Elaine Moran
Charles E. Schmid
Term to 2011
Anthony A. Atchley
Donna L. Neff
Term to 2010
William J. Cavanaugh
Floyd Dunn

Committee on Standards

Executive Committee
Paul D. Schomer, Chair (Standards Director)
Robert D. Hellweg, Vice Chair
Susan B. Blaeser, Standards Manager, ex officio
S1 Representation
Philip J. Battenberg, Chair S1
Richard J. Peppin, Vice Chair S1
Alan H. Marsh, ASA rep. on S1
Paul D. Schomer, ASA alternate rep. on S1
S2 Representation
Ali T. Herfat, Chair S2
Charles Gaumond, Vice Chair S2, ASA rep. on S2
Bruce E. Douglas, ASA alternate rep. on S2
S3 Representation
Craig A. Champlin, Chair S3 and ASA rep. on S3
David A. Preves, Vice Chair S3
Mahlon D. Burkhard, ASA alternate rep. on S3
S3/SC1 Representation
David Delaney, Chair S3/SC1
Mardi C. Hastings, Vice Chair S3/SC1 and ASA rep. on S3/SC1
vacant, ASA alternate rep. on S3/SC1
S12 Representation
William J. Murphy, Chair S12
Robert D. Hellweg, Vice Chair S12 and ASA rep. on S12
David Lubman, ASA alternate rep. on S12
International TAGs (ex officio)
Paul D. Schomer, Chair, U.S. TAG for ISO/TC 43; Chair, ISO/TC 43/SC1; ASA rep. on U.S. TAG for ISO/TC 43
Robert D. Hellweg, ASA alternate rep. on U.S. TAG for ISO/TC 43; ASA rep. on U.S. TAG for ISO/TC 43/SC1
vacant, ASA alternate rep. on U.S. TAG for ISO/TC 43/SC1
David J. Evans, Chair, U.S. TAG for ISO/TC 108
Bruce E. Douglas, ASA rep. on U.S. TAG for ISO/TC 108
Sabih I. Hayek, ASA alternate rep. on U.S. TAG for ISO/TC 108
Victor Nedzelnitsky, U.S. Technical Advisor for IEC/TC 29
Alan H. Marsh, ASA rep. on U.S. TAG for IEC/TC 29
Rufus L. Grason, ASA alternate rep. on U.S. TAG for IEC/TC 29
ASA Technical Committee Representatives
Diane Kewley-Port, Chair of ASA Technical Council, ex officio
Anthony P. Lyons, Acoustical Oceanography
Ann E. Bowles, Animal Bioacoustics
Angelo J. Campanella, Architectural Acoustics
Peter J. Kaczkowski and Vera A. Khokhlova, Biomedical Ultrasound/Bioresponse to Vibration
Mahlon D. Burkhard, Engineering Acoustics
Diana Deutsch, Musical Acoustics
Richard J. Peppin, Noise
Richard Raspet, Physical Acoustics
Brent W. Edwards, Psychological and Physiological Acoustics
Charles F. Gaumond, Signal Processing in Acoustics
Shrikanth S. Narayanan, Speech Communication
Sabih I. Hayek, Structural Acoustics and Vibration
Robert M. Drake, Underwater Acoustics
ASA Officers (ex officio)
David Feit, Treasurer, ex officio
Charles E. Schmid, Executive Director, ex officio
Past Chair of ASACOS (ex officio)
Tony F. W. Embleton
Associate Editors for Standards News—JASA (ex officio)
Susan B. Blaeser
Paul D. Schomer

Student Council

Lauren Ronsse	Chair
Megan S. Ballard	Underwater Acoustics
Colin W. Jemmott	Signal Processing in Acoustics
Dorea R. Ruggles	Psychological/Physiological
to be appointed	Musical Acoustics
Cole R. Duke	Noise and Regional Chapters Committee Liaison
Jon R. La Follett	Physical Acoustics
Lauren Ronsse	Architectural Acoustics
Lucie Somaglino	Biomedical/Bioresponse
Christian E. Stilp	Speech Communication
Mary E. Bates	Animal Bioacoustics
to be appointed	Structural Acoustics and Vibration
James Traer	Acoustical Oceanography
Scott P. Porter	Engineering Acoustics

Tutorials Committee

Lily Wang, Chair to 2012

Term to 2012
Ann R. Bradlow
David H. Chambers
Subha Maruvada
Term to 2011
Kenneth A. Cunefare
David R. Dowling
Barbara G. Shinn-Cunningham
Term to 2010
Paul E. Barbone
Micheal L. Dent
Michelle E. Swearingen
Charles E. Schmid, Executive Director, ex officio

Women in Acoustics

Marcia J. Isakson, Chair to 2012

Term to 2012
Sarah Ferguson
Kathryn W. Hatlestad
Carolyn J. Richie
Term to 2011
Judy R. Dubno
Mardi C. Hastings
Andone C. Lavery
Tracianne Neilsen
Lily M. Wang
Term to 2010
Helen M. Hanson
Jennifer L. Miksis-Olds
Peggy B. Nelson
Judy R. Dubno, ex officio as Vice President-Elect

JASA Editorial Board

A.D. Pierce, Chair

Term to June 2012
M.A. Akeroyd, Psychological Acoustics
S.L. Broschat, Underwater Sound
B.S. Cazzolato, Noise
N.P. Chotiros, Undewater Sound

J.A. Colosi, Underwater Sound
D.R. Dowling, Underwater Sound
N.A. Gumerov, Computational Acoustics
M.A. Hasegawa-Johnson, Speech Processing
M.C. Hastings, Bioacoustics-Animal
R.Y. Litovsky, Psychological Acoustics
A. Lofqvist, Speech Production
B.L. Lonsbury-Martin, Physiological Acoustics
T.D. Mast, Ultrasonics and Physical Acoustics
J.J. McCoy, Mathematical Acoustics
E. Moran, Acoustical News
S.S. Narayanan, Speech Processing
M.J. Owren, Bioacoustics-Animal
C.J. Plack, Psychological Acoustics
T.D. Rossing, Education in Acoustics
C.H. Shadle, Speech Production
R.K. Snieder, Ultrasonics and Physical Acoustics
V.W. Sparrow, Education in Acoustics
R. Stern, Electronic Archives/References/Forum
E.J. Sullivan, Acoustic Signal Processing
A.I. Tolstoy, Underwater Sound
J.A. Turner, Ultrasonics and Physical Acoustics
D.K. Wilson. Atmospheric Acoustics and Aeroacoustics
S.F. Wu, General Linear Acoustics

Term to June 2011

D.A. Berry, Speech Production
S.B. Blaeser, Acoustical News-Standards
D.S. Burnett, Computational Acoustics
W.M. Carey, Signal Processing in Acoustics
K.A. Cunefare, Noise, Its Effects and Control
D. Deutsch, Music and Musical Instruments
D. Feit, Structural Acoustics and Vibration
N.H. Fletcher, Music and Musical Instruments
K.G. Foote, Underwater Sound
K.W. Grant, Speech Perception
K.V. Horoshenkov, Noise
P.E. Iverson, Speech Perception
P.L. Marston, Acoustical Reviews-Books
V.E. Ostashev, Atmospheric Acoustics and Aeroacoustics
J.W. Posey, Atmospheric Acoustics and Aeroacoustics
L.D. Rice, Acoustical Reviews-Patents
B. Schulte-Fortkamp, Noise: Effects & Controls
W.P. Shofner, Physiological Acoustics
M. Sommers, Speech Perception
R.A. Stephen, Underwater Sound
J.E. Sussman, Speech Perception
A.J. Szeri, Ultrasonics and Physical Effects of Sound
L.M. Wang, Architectural Acoustics
R.M. Waxler, General Linear Acoustics
A.J. Zuckerwar, Applied Acoustics; Transduction; Acoustical Measurements

Term to June 2010

W.W. L. Au, Bioacoustics-Animal
P.E. Barbone, Ultrasonics; Physical Effects of Sound
Y. H. Berthelot, Ultrasonics, Physical Effects of Sound
C.C. Church, Bioacoustics
R.O. Cleveland, Nonlinear Acoustics
A.J.M. Davis, General Linear Acoustics
L.P. Franzoni, Structural Acoustics and Vibration
S.A. Fulop, Patents
J.H. Ginsberg, Structural Acoustics & Vibration
A. Hirschberg, Atmospheric Acoustics and Aeroacoustics

JASA Editorial Board (Continued.)

J.G. McDaniel, Structural Acoustics & Vibration
J.C. Middlebrooks, Psychological Acoustics
D.L. Miller, Bioacoustics
B.C.J. Moore, Psychological Acoustics
R.S. Newman, Speech Perception
D.D. O'Shaughnessy, Speech Processing and Communication Systems
R. Raspet, Ultrasonics, Physical Effects of Sound
O.A. Sapozhnikov, Nonlinear Acoustics
P.D. Schomer, Acoustical News Standards
W.L. Siegmann, Underwater Acoustics
J.A. Simmons, Bioacoustics
L.C. Sutherland, Atmospheric Acoustics and Aeroacoustics
L.L. Thompson, General Linear Acoustics
R. L. Weaver, Structural Acoustics and Vibration
E. G. Williams, Structural Acoustics and Vibration
M. Wojtczak, Psychological Acoustics
N. Xiang, Architectural Acoustics

Associate Editors of JASA Express Letters (JASA-EL)

Term to 30 June 2012

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

Term to 30 June 2011

D. Deutsch, Musical Acoustics
Q.-J. Fu, Psychological Acoustics
V.E. Ostashev, Aeroacoustics and Atmospheric Acoustics
M.D. Sheplak, Transduction

Term to 30 June 2010

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

Acoustics Today Editorial Board

Richard Stern, Chair
Elliott H. Berger
Carol Espy-Wilson
K. Anthony Hoover
Allan D. Pierce
Thomas D. Rossing
Brigitte Schulte-Fortkamp

Ad-Hoc Committees 2009–2010

Foundation Transition

Peggy Nelson, Chair
Judy R. Dubno, Vice Chair
Anthony Atchley
Mahlon D. Burkhard
Patricia K. Kuhl
Richard H. Lyon
Paul B. Ostergaard
Carl J. Rosenberg
George P. Wilson
Ex officio: David Feit, ASA Treasurer

Innovation

Donna L. Neff, Chair
Bennett M. Brooks
E. Carr Everbach
George V. Frisk
Mardi C. Hastings
Diane Kewley-Port
Mauro Pierucci
Lily M. Wang
Janet M. Weisenberger
Ex-officio: David Feit, ASA Treasurer

Intra-Society

Courtney B. Burroughs, Chair
James W. Beauchamp
Alexander U. Case
N. Ross Chapman
Brenda L. Lonsbury-Martin
Rahul Shrivastav
Natalia V. Sizov
Kerrie G. Standlee
Paul A. Wheeler
Charles E. Schmid, *Ex-officio*

Diversity in Acoustics

Juan Arvelo, Chair
Anthony A. Atchley
David T. Bradley
Nandini Iyer
Elaine Moran
Tyrone M. Porter
Sonya T. Smith
Laura Tejada
James E. West

International Meetings

William A. Yost, Chair
David T. Blackstock
Lawrence A. Crum
Clark S. Penrod

Posters for Education

Victor W. Sparrow, Chair
Fredericka Bell-Berti
Kelly J. Benoit-Bird
Robin O. Cleveland
Judy R. Dubno
Geoffrey F. Edelmann
Pamela J. Harght

Website for Kids

Preston S. Wilson, Chair
Kelly J. Benoit-Bird
Ann R. Bradlow
Uwe J. Hansen
Pamela J. Harght
Andrew A. Piacsek
James M. Sabatier

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]

Paul D. Schomer, Standards Director
Schomer and Associates, 2117 Robert Drive, Champaign, IL 61821 [Tel.: (217) 359-6602; Fax: (217) 359-3303; e-mail: Schomer@SchomerAndAssociates.com]

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

Standards Meetings Calendar—National

- In conjunction with the 158th ASA meeting, ASA Committee on Standards (ASACOS) will meet in San Antonio, Texas, **27 October 2009** at 7:30 a.m.
- ASACOS; Accredited Standards Committees S1, S2, S3, S3/SC1, and S12; and the Standards Plenary Group will meet during the Joint 159th ASA Meeting and Noise-Con 2010 in Baltimore, Maryland, **19–23 April 2010**. Specific dates and times will be announced.

Standards Meeting Calendar—International

• 16–20 November 2009—Seoul, South Korea

– ISO/TC 43, Acoustics

– ISO/TC 43/SC 1, Noise

– ISO/TC 43/SC 2, Building acoustics

• 9–13 November 2009, Tokyo, Japan

– IEC TC 29, Electroacoustics

Thanks to our Volunteers!

The ASA Standards Secretariat would like the thank the following people for volunteering their time and expertise to coordinate comments and formulate recommendations for the U.S. vote on numerous ISO and IEC documents throughout the first half of this year. Their efforts are greatly appreciated and extremely valuable. The following is a list of volunteers through the end of June 2009.

Michael Bahtiarian
Raymond Bankert
Kenneth Culverson
Bruce Douglas
Ronald Eshleman
Dave Evans
Sanford Fidell
William Foiles
R. Lee Grason
Arthur Kilcullen
Walter Madigosky
Alan Marsh
Eric Maslen
Macinissa Mezache
William Murphy
Victor Nedzelnitsky
John Niemkiewicz
Larry Pater
Dave Preves
Stephen Roth
Paul Schomer
Richard Taddeo
Dave Vendittis
Michael J. White
Laura Wilber
David Wilder

STANDARDS NEWS FROM THE UNITED STATES

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

American National Standards Call for Comment on Proposals Listed

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

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

New Standards

BSR/AHRI 350-200x, Sound Performance Rating of Non-Ducted Indoor Air-Conditioning Equipment (new standard)

Applies to the indoor portions of factory-made non-ducted air-conditioning equipment. (August 10, 2009)

ATIS (Alliance for Telecommunications Industry Solutions)

New Standards

BSR ATIS 0600010.02-200x, Equipment Handling, Transportation Vibration and Rail Car Shock Requirements for Network Telecommunications Equipment (new standard)

Provides evaluation criteria to industry to ensure that the effects of Transportation and Installation Handling Requirements on network telecommunications equipment are minimized. (June 29, 2009)

ASA (ASC S12) (Acoustical Society of America)

New Standards

BSR/ASA S12.60-200x/Part 2, Acoustical Performance Criteria, Design Requirements, and Guidelines for Schools—Part 2: Relocatable Classroom Factors (new standard)

Provides a relocatable-classroom-specific supplemental version of ANSI S12.60. Includes siting requirements, acoustical performance criteria & design requirements for relocatable classrooms. Annex A provides commentary info on this standard. Annex B provides procedures for determining compliance with background sound requirements. Seeks to provide design flexibility without compromising goal of obtaining adequate speech intelligibility for students and teachers in learning spaces within the standard's scope. (August 3, 2009)

Call for Members (ANS Consensus Bodies)

Directly and materially affected parties who are interested in participating as a member of an ANS consensus body for the standards listed below are requested to contact the sponsoring standards developer directly and in a timely manner.

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

BSR/AHRI 275-200x, Application of Sound Rating Levels of Outdoor Unitary Equipment (new standard)

BSR/AHRI 300-200x, Sound Rating and Sound Transmission Loss of Packaged Terminal Equipment (new standard)

BSR/AHRI 350-200x, Sound Performance Rating of Non-Ducted Indoor Air-Conditioning Equipment (new standard)

BSR/AHRI 370-200x, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (revision of ANSI/AHRI Standard 370-2001)

ASA (ASC S1) (Acoustical Society of America)

BSR/ASA S1.26-200x, Method for Calculation of the Absorption of Sound by the Atmosphere (revision of ANSI/ASA S1.26-1995 (R2009))

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.

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

BSR/AHRI 260-200x, Sound Rating of Ducted Air Moving and Conditioning Equipment (new standard)

Applies to all ducted air moving and conditioning equipment containing fans as defined in this standard. Project Need: To establish a method of sound rating the indoor portions of ducted air moving and conditioning equipment. Stakeholders: Manufacturers, engineers, installers, contractors and users.

BSR/AHRI 275-200x, Application of Sound Rating Levels of Outdoor Unitary Equipment (new standard)

Applies to the outdoor sections of factory-made air-conditioning and heat-pump equipment, as defined in Section 3 and AHRI Standard 210/240, when rated in accordance with AHRI Standard 270. Project Need: To establish for outdoor unitary equipment: definitions, procedures for estimating A-Weighted sound pressure levels, and recommended application practices. Stakeholders: Manufacturers, engineers, installers, contractors, and users.

BSR/AHRI 300-200x, Sound Rating and Sound Transmission Loss of Packaged Terminal Equipment (new standard)

Applies to the indoor and outdoor sections of factory-made Packaged Terminal Equipment, as defined in AHRI Standard 310/380. Project Need: To establish, for packaged terminal equipment: Definitions; test requirements; rating requirements; minimum data requirements for Published Ratings; and conformance conditions. Additionally, this standard establishes a method to determine sound transmission loss for Packaged Terminal Equipment. Stakeholders: Manufacturers, engineers, installers, contractors and users.

BSR/AHRI 350-200x, Sound Performance Rating of Non-Ducted Indoor Air-Conditioning Equipment (new standard)

Applies to the indoor portions of factory-made non-ducted air-conditioning equipment, as defined in AHRI Standards 210/240, 340/360, 310/380, and 440. Project Need: To establish for non-ducted indoor air-conditioning equipment: Definitions; test requirements; rating requirements; minimum data requirements for Published Ratings; marking and nameplate data; and conformance conditions. Stakeholders: Manufacturers, engineers, installers, contractors and users.

BSR/AHRI Standard 370-200x, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (revision of ANSI/AHRI Standard 370-2001)

Applies to the outdoor portions of factory-made commercial and industrial Large Outdoor Refrigerating and Air-Conditioning Equipment, including heat pumps, used for refrigerating or air-conditioning of spaces, as defined in Section 3 of this standard. Project Need: To establish methods for determining the sound ratings of the outdoor portions of factory-made commercial and industrial large outdoor refrigerating and air-conditioning equipment. Stakeholders: Manufacturers, engineers, installers, contractors, and users.

ASA (ASC S1) (Acoustical Society of America)

BSR ASA S1.26-200x, Method for Calculation of the Absorption of Sound by the Atmosphere (revision of ANSI/ASA S1.26-1995 (R2009))

Provides the means to calculate atmospheric absorption losses of sound from any source, over a wide range of meteorological conditions. Attenuation coefficients for pure-tone sounds are calculated by means of equations (or a table) for the frequency of the sound, and the humidity, pressure, and temperature of the atmosphere. For sounds analyzed by fractional-octave-band filters, alternative methods to calculate the attenuation caused by atmospheric absorption are provided. Project Need: To resolve comments arising from a reaffirmation ballot regarding the unit of distance to be used in the calculations, the role of the relative humidity, and possible errors in the tabulations of atmospheric attenuation coefficients. The WG will consider these comments. Stakeholders: Industry, government agencies (Federal and State), U.S. Military agencies, acoustical consultants.

SCTE (Society of Cable Telecommunications Engineers)

BSR/SCTE DSS 02-14-200x, IPCablecom 1.5 Part 17: Audio Server Protocol (revision and redesignation of ANSI/SCTE 24-17-2007)

Defines a set of signaling protocols that are used to provide announcement services within a cable network. For one of these protocols, the PacketCable Network Call Signaling (NCS) protocol, this specification defines two new event packages: a Base Audio Package and an Advanced Audio Package. Project Need: To update the standard to include current technology. 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. Technical Reports have been registered in accordance with ANSI's *Procedures for the Registration of Technical Reports with ANSI*.

AHRI (Air-Conditioning, Heating, and Refrigeration Institute)

New Standards

ANSI/AHRI 1140-2006, Sound Quality Evaluation Procedures for Air-Conditioning and Refrigeration Equipment

ASA (ASC S1) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S1.11-2004 (R2009), Specification for Octave-Band and Fractional-Octave-Band Analog and Digital Filters (reaffirmation and redesignation of ANSI S1.11-2004)

ANSI/ASA S1.26-1995 (R2009), Method for Calculation of the Absorption of Sound by the Atmosphere (reaffirmation and redesignation of ANSI S1.26-1995 (R2004))

ASA (ASC S2) (Acoustical Society of America)

New Standards

ANSI/ASA S2.62-2009, Shock Test Requirements for Equipment in a Rugged Shock Environment

Reaffirmations

ANSI/ASA S2.25-2004 (R2009), Guide for the Measurement, Reporting, and Evaluation of Hull and Superstructure Vibration in Ships (reaffirmation and redesignation of ANSI S2.25-2004)

ASA (ASC S3) (Acoustical Society of America)

Revisions

ANSI/ASA S3.2-2009, Method for Measuring the Intelligibility of Speech over Communication Systems (revision and redesignation of ANSI S3.2-1989 (R1999))

ANSI/ASA S3.22-2009, Specification of Hearing Aid Characteristics (revision and redesignation of ANSI S3.22-2003)

ASA (ASC S12) (Acoustical Society of America)

Reaffirmations

ANSI/ASA S12.18-1994 (R2009), Outdoor Measurement of Sound Pressure Level (reaffirmation and redesignation of ANSI S12.18-1994 (R2004))

CEA (Consumer Electronics Association)

New Standards

ANSI/CEA 2006-B-2009, Testing and Measurement Methods for Mobile Audio Amplifiers

HI (Hydraulic Institute)

Revisions

ANSI/HI 9.6.4-2009, Rotodynamic Pumps for Vibration Measurements and Allowable Values (revision of ANSI/HI 9.6.4-2000)

Newly Published ISO and IEC Standards

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

ISO Standards

Acoustics (TC 43)

ISO 389-9:2009, Acoustics—Reference zero for the calibration of audiometric equipment—Part 9: Preferred test conditions for the determination of reference hearing threshold levels

ISO 3382-1:2009, Acoustics—Measurement of room acoustic parameters—Part 1: Performance spaces

ISO 13474:2009, Acoustics—Framework for calculating a distribution of sound exposure levels for impulsive sound events for the purposes of environmental noise assessment

EARTH-MOVING MACHINERY (TC 127)

ISO 6394/Cor1:2009, Acoustics—Measurement of airborne noise emitted by earth-moving machinery—Operators position Stationary test condition—Corrigendum

ISO 6396/Cor1:2009, Acoustics—Measurement at the operators position of noise emitted by earth-moving machinery—Dynamic test conditions—Corrigendum

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO 18431-1/Cor1:2009, Mechanical vibration and shock—Signal processing—Part 1: General introduction—Corrigendum, FREE

Draft International Standards

This section lists proposed standards that the International Organization for Standardization (ISO) is considering for approval. The proposals have received substantial support within the technical committees or subcommittees that developed them and are now being circulated to ISO members for comment and vote. Standards Action readers interested in reviewing and commenting on these documents should order copies from ANSI.

ISO Drafts

ERGONOMICS (TC 159)

ISO/DIS 24501, Ergonomics—Accessible design—Sound pressure levels of auditory signals for consumer products (August 22, 2009)

TRACTORS AND MACHINERY FOR AGRICULTURE AND FORESTRY (TC 23)

ISO/DIS 22868, Forestry and garden machinery—Noise test code for portable hand-held machines with internal combustion engine—Engineering method (Grade 2 accuracy) (August 15, 2009)

IEC Drafts

29/683/FDIS, IEC 60318-1 Ed.2: Electroacoustics—Simulators of human head and ear—Part 1: Ear simulator for the measurement of supra-aural and circumaural earphones (July 10, 2009)

45A/754/FDIS, IEC 60988 Ed.2: Nuclear Power Plants—Instrumentation important to safety—Acoustic monitoring systems for detection of loose parts: Characteristics, design criteria and operational procedures (July 24, 2009)

86B/2862/FDIS, IEC 61300-2-1 Ed. 3.0: Fibre optic interconnecting devices and passive components—Basic test and measurement procedures—Part 2-1: Tests—Vibration (sinusoidal)

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.

History of Russian Underwater Acoustics

Oleg A. Godin and David R. Palmer

World Scientific Publishing Company, 2008. 1232 pp. Price: $216 (hardcover). ISBN-13; 978-981-256-825-0

History of Russian Underwater Acoustics (HRUA) is an English translation of a Russian book with a similar title published in St. Petersburg in 1999. With 90 articles by 100 authors, the original Russian version was published to commemorate the 300th anniversary of the Russian Navy in 1998. The current English version provides easy access by Western readers to a wealth of underwater acoustic science and technology that has been classified for security reasons or otherwise unavailable before now, owing to the Cold War between global superpowers that dominated the last half of the 20th century. The editors believe that the book will provide fascinating reading to engineers and scientists who were (and are now) engaged in similar work in the West, to Soviet and Cold War historians, to members of the world's navies (serving and retired), and to present-day researchers.

Relative to the Russian version, HRUA has been broadened in coverage by the inclusion of a new chapter, *The Physics of Underwater Sound*, comprising three extensive review articles on Russian contributions to the science, contributed by Russian experts. There is also an extra article on dolphin acoustics. On the other hand, some 10% of the original Russian book was not included in the translation.

Preparation of HRUA was supported in part by the U.S. Office of Naval Research, to complement an ONR-sponsored series of monographs on state-of-the-art underwater acoustics as practiced in the United States.[1,2] However, the book reviewed here and the first two volumes of the monograph series are quite different in style: Where HRUA is mostly essays, reviews, and personal memoirs, the ONR monographs present substantial experimental data and technical mathematics.

Following the Prefaces to the English and Russian Editions, the chapters of HRUA are given as follows.

I. Introduction: Underwater Acoustics and the Ocean
II. Hydroacoustics in Russia from the 19th Century to the Present Time
III. The Physics of Underwater Sound
IV. Laying the Scientific and Practical Foundation for Home Hydroacoustics
V. Submarine Hydroacoustic Equipment
VI. Sonar Systems for Surface Ships
VII. Stationary Sonar Systems
VIII. Specialized Hydroacoustic Systems
IX. Sonar Arrays
X. The Role of the Radio Engineering Department and the Naval RI [Research Institute] in the Creation of Hydroacoustic Equipment
XI. Organization of Hydroacoustic Equipment Development
XII. Training of Hydroacoustics Engineering and Research Personnel
XIII. Veterans Remember

The first 200 pages or so, comprising 6 essays in 3 chapters, are closer to what is normally considered history, including contributions by authors that should be familiar to Western researchers: Brekhovskikh, Goncharov, Lysanov, and Kuryanov, for example. The content is mostly scientific and to some extent the cited sources have been available in translation in unclassified publications; however, having it cited in one place is convenient. This portion perhaps would be of more interest to those studying the science of underwater acoustics. The remaining 10 chapters are dominated by technology and systems, typically regarded as classified information at the time, in the form of personal memoirs by the individuals directly involved. This portion would be of more interest to engineers engaged in similar developments. It is possible to sample the actual content of this book online by searching the internet site books.google.com using the search terms "Russian underwater."

My personal impression is that HRUA is interesting, although not at all light reading. I found it interesting because I have worked in this area in the West (Canada) myself from 1977 to 2008, half-in and half-out of the Cold War years. At the outset, the Soviets were a clear "enemy," in defense science parlance, and I never dreamed of someday being able to read material such as contained in this book. (Things had changed dramatically by the end of my career: I overheard two fellow Defense Scientists casually conversing in Russian in a hallway of our establishment!) Much of the material is a bit dry: a chronicle of names, dates, places, and projects; but this in part is what a history is for. The review articles provide a convenient bibliography of the key contributions by Russians to studies of propagation, noise, scattering, and reverberation. Probably as a result of the memoir-style of most of the writing, there is a slight hyperbolic tendency: All managers were wise and efficient, all experts were brilliant, all machinists were inventive, all laborers were hard-working, and so on. It is easy to compensate for this tendency while reading. Particular mention should be made of the memoir by V. Z. Krants, which I found to be an amusing and refreshing departure from the norm. Finally, HRUA's usefulness to both scientific and historical researchers would be enhanced if there were subject and name indices.

In conclusion, if it seems that the entire 1200+ pages of HRUA would be a challenging read, the book's format allows you to explore and discover those gems of special interest, assuming that you have some interest in the subject matter to begin with. The book is certainly worthy of joining other general books on underwater acoustics science and technology on the shelves of a technical library.

DAVID M.F. CHAPMAN
8 Lakeview Avenue,
Dartmouth,
Nova Scotia B3A3S7, Canada

[1]M. D. Richardson, *High-Frequency Seafloor Acoustics* (Springer, New York, 2007).
[2]C. H. Sherman and J. L. Butler, *Transducers and Arrays for Underwater Sound* (Springer, New York, 2008).

Elastic Waves in Composite Media and Structures: With Applications to Ultrasonic Nondestructive Evaluation

Subhendu K. Datta and Arvind H. Shah

CRC Press, 2009, 318 pp. Price: $149.95 (hardcover) ISBN: 978-1-4200-5338-8.

For many years there have been books about the structural and material properties of composite laminates and other books about wave propagation in anisotropic media, but there have been essentially no books about wave propagation specifically in composites. This new monograph by Subhendu K. Datta and Arvin H. Shah begins to fill that gap. At just over 300 pages this new book covers a broad range of topics in wave propagation, generally closely aligned with the extensive research contributions of its authors.

After a historical discussion of waves in composites and a scoping of the book's reach, the authors set the stage for an in-depth treatment of wave mechanics. The first chapter is devoted to a presentation of the fundamental equations of elastic waves in anisotropic, but still homogeneous, media. The treatment here is rather heavily reliant on theoretical mechanics; there are

 0001-4966/2009/126(3)/1633/2/$25.00

but a few figures, and the ones that do occur appear only at the end. The presentation overall is a model of mathematical efficiency. In other words, the chapter is wonderfully economical in its style and presentation, and the knowledgeable reader cannot help but to be impressed. For the novice, however, it may not be the easiest place to learn the subject. In truth, this chapter has the character of a review for workers already quite familiar with the topic.

The second technical subject examined in this book is waves in periodic layered media. This subject is appropriate for such a book because composite laminates are often composed of some basic building block, such as a ply of fibers in the 0° direction, then a ply in the +45° direction, then −45°, and finally 90°. This basic unit cell is then repeated many times throughout the laminate. Periodic layering is, in fact, one of the hallmarks of composite materials that distinguishes them from other structural materials. The discussion on this topic offered in this book is again very competently done, and quite streamlined from a mathematical standpoint. The exposition leads naturally to a discussion of Floquet waves, although the argument is almost too brief to give the uninformed reader any idea of the essence of this concept. Following on the heels of this discussion is an examination of effective medium theories, an important milestone in the applied mechanics of composites under dynamic load. Although the reader is warned that this topic applies to "long wavelengths compared to the unit cell," there is little further interpretation of what this means to someone trying to use the theory. There follows a section devoted entirely to numerical results, calculated on the basis of the theory presented at the head of the chapter. These graphs or plots are shown as normalized frequency versus the real part of the propagation constant, with imaginary or complex branches shown as dashed curves. Relatively little interpretation is included, except to note that dispersion is greatest when waves propagate normal to the layering. Effective medium theory results are also presented, but the only comparisons shown are between competing types of theories.

The next chapter is devoted to the discussion of guided waves in plates. The familiar theoretical analysis, known to most researchers in this area, is absent here. The approach of this book is based on wave potentials, a concept usually associated with partial waves consisting of pure modes, for which the method was invented. Instead, the partial waves here, in fully anisotropic media, are all possibly quasimodes, and so the intuitive appeal of the potential theory approach seems to be lost. Formally, of course, it is all quite valid; it is just that the intuitive association of a single potential with a unique pure mode no longer holds. It goes on this way until the potentials are eventually used to compute displacements and stresses. The readers never actually gets to see a guided wave dispersion equation, only a matrix of formal elements whose definition must be looked up. Then, the reader is told to simply take a determinant and set the results equal to zero. There follows, as before, a series of calculated curves based on various hypothetical material situations. The curves are presented in such a way as to be comfortable for a theorist to consider: imaginary branches plotted next to real ones, pseudo-three-dimensional representation of complex branches, and plots of normalized wavenumber versus normalized frequency. One wonders, however, whether engineers in industry who design and perform composite inspections would find such plots at all congenial.

The following section on material characterization does a much better job of making the connection between theory and practice. Actual material constants are listed, and experimental data are shown an compared with the calculations. Surprisingly, the data chosen for exposition are measurements on layered aramind-aluminum laminate. Every other layer and the outside layers of this material are polycrystal aluminum (and therefore isotropic). These aluminum layers alternate with layers of aramid (Kevlar is the brand name) fibers in epoxy. Indeed, this laminate qualifies as a composite material but because this series of some half dozen plots is the only one in the whole book, perhaps a graphite-epoxy lay-up might have been the choice more anticipated. For whatever reason, it is not. The subject of waves in thin layers is taken up next, together with more analysis, including more familiar sets of equations. The results presented use the example of guided waves in paper. The seminal work in the area of guided waves in webs of moving paper was performed decades ago by M. Luukkala and collaborators in Finland, and P. H. Brodeur at Georgia Tech, but their work seems not to have been mentioned. There follows a section on waves in plates with thin layers, either one the surfaces or in the interior. This topic is evidently a set-up for the presentation of quite a bit of data on layered high-temperature superconductors. To find this topic so extensively represented in a chapter title "Guided waves in fiber-reinforced composite plates" is frankly a little jarring. The chapter ends with a detailed calculation of laser ultrasound in the thermoelastic regime. Results of model calculations are presented, but without experimental data.

A related chapter is on waves guided by cylinders of various cross section, primarily circular, but including rectangular. Graphs are of theoretical predictions and comparisons between competing calculational approaches.

The book's last chapter is devoted to scattering of guided waves in plates and cylinders, a complicated problem under the best of circumstances. Here, the model calculations are carefully explained and many plots of predictions based on the models illustrate the range of the theory. Both finite element and boundary element methods are presented as means to solve the difficult geometrical problems posed by the candidate defects. In the graphical material, comparisons are almost exclusively between various mathematical means of solution, seldom between measurements and predictions.

The volume ends with a series of descriptions of computer programs contained on an included CD-ROM. The programs are divided into three categories: waves in plates, waves in cylinders, and material properties and bulk waves. For waves in plates, the programs can compute frequency spectra, phase, and group velocities, and Green's functions for several categories of wave types. The same selection is available for waves in cylinders also. Then, for bulk waves the programs can compute material property transformations and slowness surfaces. The programs, which must be installed on the host computer, are executables designed for the Microsoft Windows© operating system. After installation, the programs can be run from a command icon automatically installed into the *Start Menu*. Clicking on the icon brings up a *Table of Contents* with a choice of 17 selectable programs, each with a *Help* guide. The operator enters material properties (or uses embedded examples), number of layers, and other physical parameters. The calculations proceed quickly, and the results are presented in both tabular and graphical formats. Both kinds of data can be saved for further use to standard files named by the user (Excel© for tabular data, and bitmap or JPEG for graphical).

Finally, this book does several things very well. It is extensively (almost exhaustively) referenced, the mathematical exposition is very competently done, and it addresses many of the most difficult problems in wave propagation and scattering. The reader is left wishing, however, that the book contained more experimental data, a more thorough description of real composite laminates—complete with their imperfections, and some guidance on how to use the theory when the idealizations assumed in the calculations inevitably collide with the non-ideal nature of real-world composite materials. Because the book's subtitle is "With Applications to Ultrasonic Nondestructive Evaluation," a reader might reasonably expect to see those blanks filled in. Nonetheless, I would recommend this book enthusiastically to the more experienced engineers among us, who may be in a better position to mine its many glories.

DALE E. CHIMENTI
Department of Aerospace Engineering,
Center for Nondestructive Evaluation,
Iowa State University,
Ames, IA 50011

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

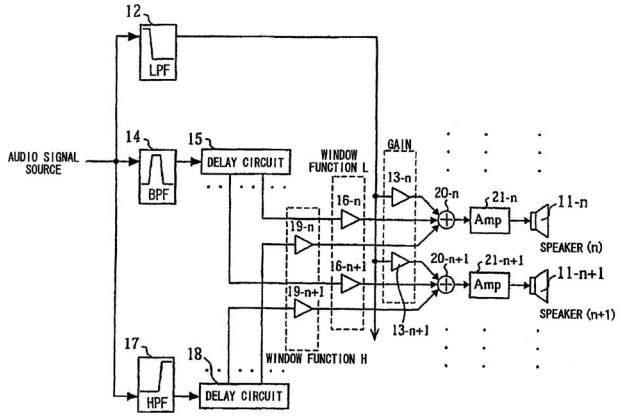
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *Department of Electrical and Computer Engineering, University of Rochester, Rochester, New York 14627*

7,519,187

43.38.Hz ARRAY SPEAKER SYSTEM

Yusuke Konagai, assignor to Yamaha Corporation
14 April 2009 (Class 381/98); filed in Japan 2 June 2003

To maintain uniform directivity from a loudspeaker line array, frequency-dependent shading can be employed so that the effective length of the array is inversely proportional to frequency. Another possibility is to use arrays of different lengths for low, mid, and high frequencies. This patent adapts the second approach to a single array. The input signal is first divided



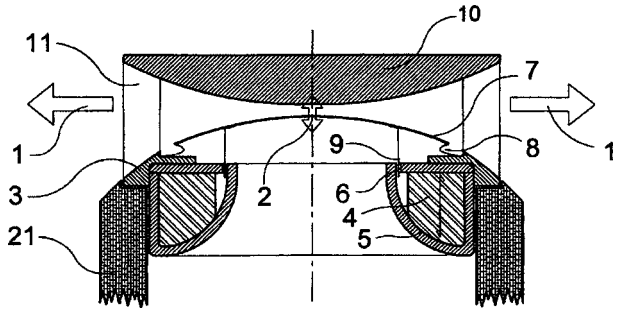
into three bands of frequencies—low, mid, and high. Each band has its own set of weighting factors, creating three virtual arrays. There is a lot of prior art in this field, which may be why the patent claims limit the scope of the invention to symmetrical, delay-steered arrays.—GLA

7,426,281

43.38.Ja ELECTRODYNAMIC ACOUSTIC TRANSDUCER

Patrick Hoffmann, assignor to Rötelzeichnung Holding A.G.
16 September 2008 (Class 381/339); filed in France 10 February 2003

An omnidirectional transducer to generate sound at low to medium audio frequencies is claimed. Magnet 4 creates steady magnetic flux in iron 5 and across gap 6. Backer 10 is stationary, being fixed by struts 11 to mount 21. Audio frequency current is passed through a winding on armature 9



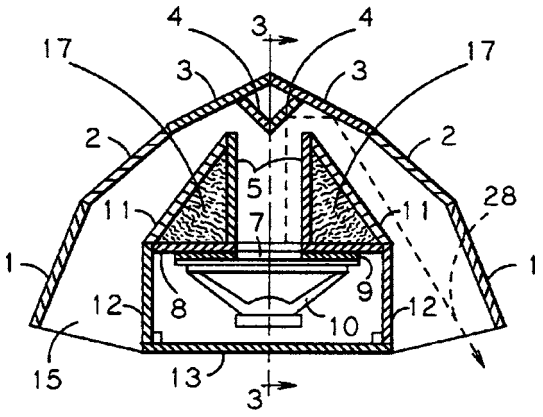
causing motion 2 of stiff membrane 7, driving air 1 which proceeds outward (arrows) as sound radiating in all directions around the axis. Other analogous arrangements to radiate sound at the higher audio frequencies are also claimed.—AJC

7,506,721

43.38.Ja CONVERTIBLE FOLDED HORN ENCLOSURE

Dana A. Moore, Bothell, Washington
24 March 2009 (Class 181/156); filed 10 November 2006

As can be seen from the illustration, this is a straightforward folded corner horn design, no better or worse than a dozen others. However, the geometry allows the front panel to be replaced by a loudspeaker mounting panel, thus converting a front-loading horn to a rear-loading horn should the need arise. A variant is covered separately in U.S. Patent No. 7,513,332.—GLA

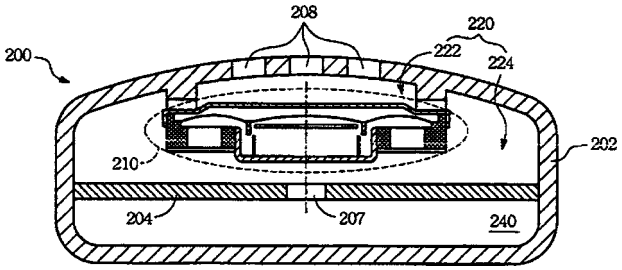


7,508,933

43.38.Ja RESONANCE CHAMBER OF A CELLULAR PHONE

Tsung-Lung Yang, assignor to BenQ Corporation
24 March 2009 (Class 379/433.02); filed in Taiwan 3 February 2004

The "Background of the Invention" section of this patent starts out with a fairly detailed but substantially erroneous explanation of how a Helmholtz resonator loads a loudspeaker. Nonetheless, venting back chamber 224 to otherwise unused chamber 240 can be expected to extend the low frequency range of the system. Opening 207 can be sized to suppress a mid-frequency peak that might otherwise occur. Response curves are included that show a usable low frequency extension of 1/3-octave to 1 octave.—GLA

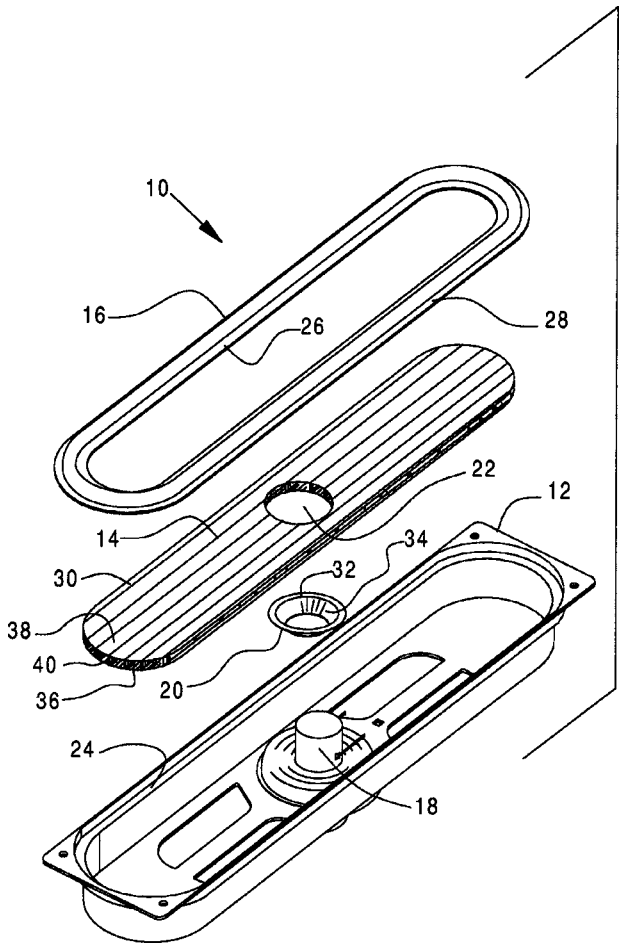


7,510,047

43.38.Ja SPEAKER EDGE AND RESONATOR PANEL ASSEMBLY

Keiko Muto and Mayuki Yanagawa, both of Marina Del Rey, California
31 March 2009 (Class 181/173); filed 4 October 2007

An elongated flat-panel loudspeaker tends to have irregular and unpredictable response characteristics. Normal manufacturing tolerances may result in substantial changes in performance. This patent describes a design that is said to provide extended frequency response along with less critical assembly requirements. Panel 14 is made up of top and bottom membranes 36, 38 separated by ribs 40. The ribs may be spaced and angled to achieve optimum performance. The inventors have also discovered (as did Olson more than 50 years ago) that the panel edge termination is an important factor in achieving smooth frequency response. Accordingly, outer suspension 16 includes provisions for non-uniform damping. Frequency response curves are included in the patent.—GLA

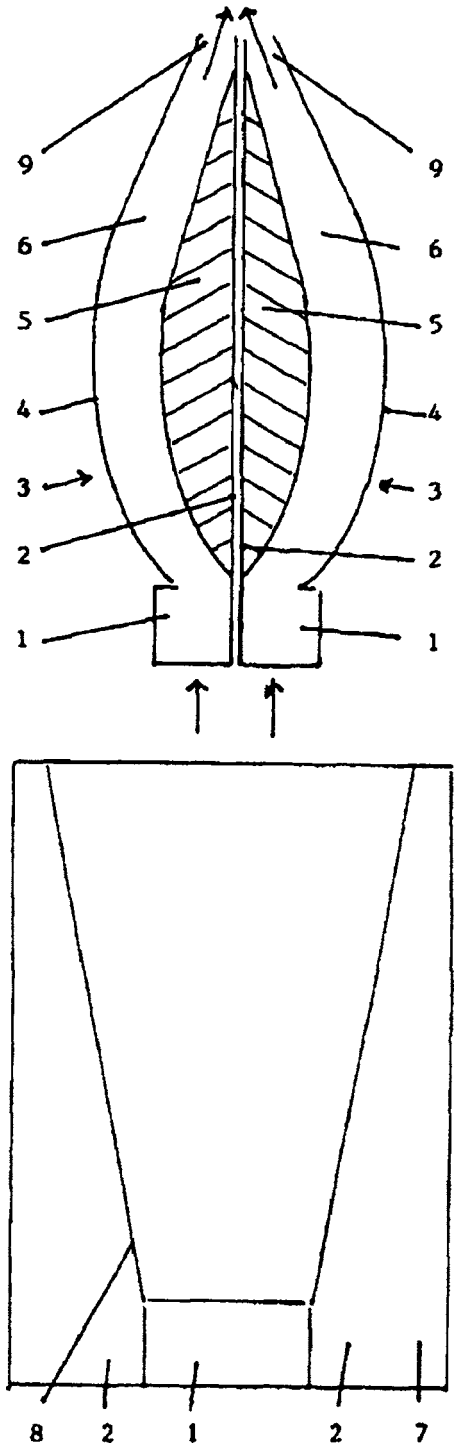


7,510,049

43.38.Ja ACOUSTIC TRANSFORMER AND METHOD FOR TRANSFORMING SOUND WAVES

Martin Kling, 30179 Hannover, Germany
31 March 2009 (Class 181/187); filed in Germany 27 October 2005

Each module of a modern line array is intended to produce a cylindrical wave front. This is difficult to achieve at high frequencies. One approach is to use several high frequency sources stacked vertically. Another is to couple a single high frequency driver to a vertical slot through some kind of waveguide that maintains a constant path length from the driver to any point on the vertical slot. The first such waveguide was patented by Heil in 1992. Since then a number of other geometries have been patented, most of which use multiple sound channels bent into various shapes. The waveguide pictured here consists of symmetrical sections on either side of a planar vertical septum. With the exception of the throat coupler, all of the external surfaces can be formed from sheet stock if desired. The internal "displacement body" 5 has the form of a tulip. The well-written patent claims specify only the basic geometry; there are no restrictions as to size, proportions, flare rate, or path lengths.—GLA

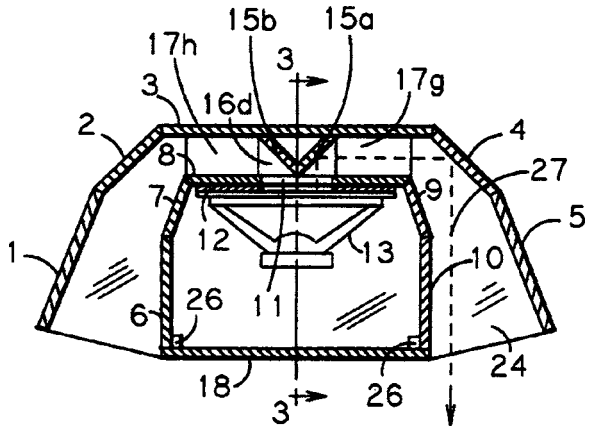


7,513,332

43.38.Ja CONVERTIBLE FOLDED HORN ENCLOSURE WITH IMPROVED COMPACTNESS

Dana A. Moore, Bothell, Washington
7 April 2009 (Class 181/155); filed 12 September 2007

This is a companion to U.S. Patent No. 7,506,721. The geometry is similar but the shape of the cabinet is more nearly rectangular, making it a better choice for non-corner placement.—GLA

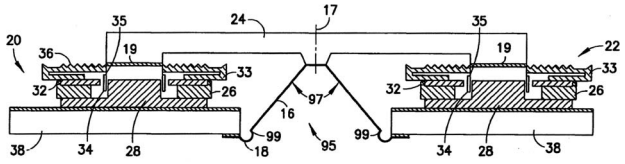


7,515,724

43.38.Ja LOUDSPEAKER DRIVER

Kourosh Salehi, New York, New York
7 April 2009 (Class 381/182); filed 3 April 2007

The operation of this inside-out, spread-apart loudspeaker is actually quite simple but difficult to visualize from the illustration. If it were drawn to scale, the diameter of cone 97 would be much larger in relation to driving motors 19. The two or more driving motors are conventional magnet and voice coil assemblies mounted to the back side of baffle board 38. The voice



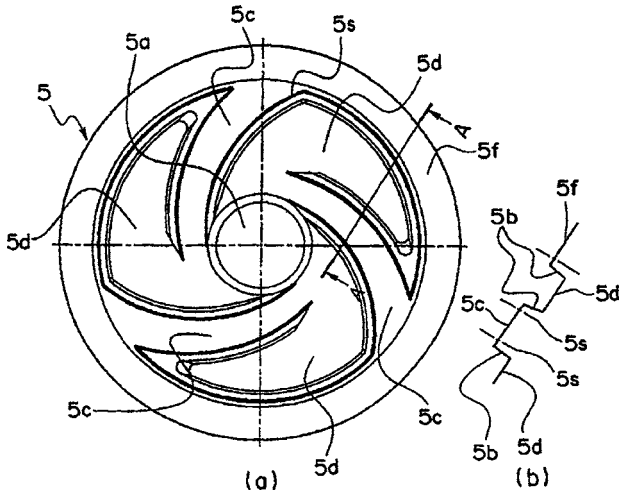
coils are mechanically connected to the cone through lightweight yoke 24. In comparison with other multi-drive loudspeakers, the novel feature is that the motors are all located outside the perimeter of the diaphragm. This is said to result in a shallower assembly.—GLA

7,418,109

43.38.Kb DIAPHRAGM STRUCTURE OF LIGHT SOUND CONVERTER

Yoshio Sakamoto and Takahiro Imai, assignors to Kabushiki Kaisha Kenwood
26 August 2008 (Class 381/423); filed in Japan 19 June 2001

A light modulating microphone diaphragm 5 with a better frequency response and sensitivity is claimed where 1.3 mm diameter dome reflector 5a is positioned over a laser-receiver pair, not shown. Arc shaped slits 5s that are 40 μm wide are laser-cut into the peripheral diaphragm area to



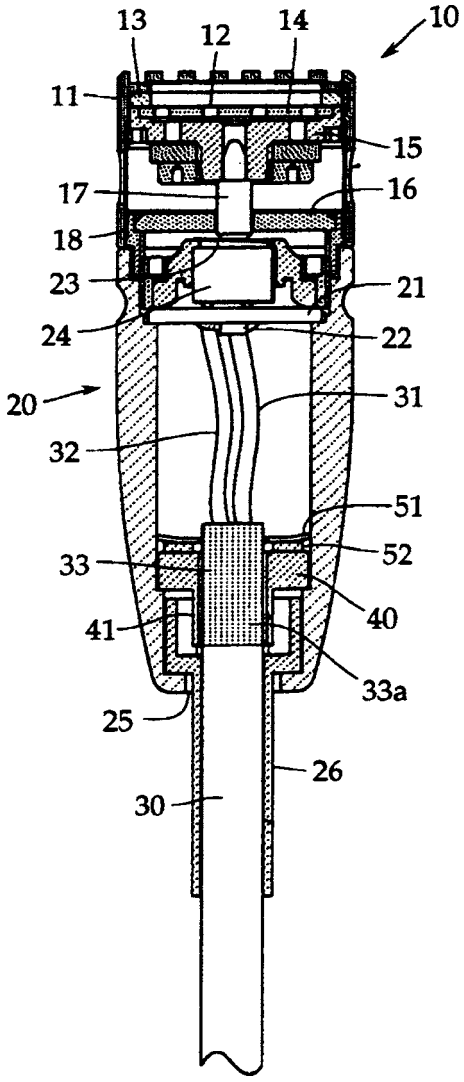
define cantilever supports 5c connecting dome 5a to support ring 5f. Remaining sails 5d increase acoustic velocity reception area. Corrugations 5bstiffen sails 5d. Resonance frequency is controlled by the width of cantilevers 5c. Measurements at 1 kHz indicate that the dome vibration amplitude with slits is 12–27 times that of a diaphragm without slits.—AJC

7,483,542

43.38.Kb CONDENSER MICROPHONE

Hiroshi Akino *et al.*, assignors to Kabushiki Kaisha Audio-Technica
27 January 2009 (Class 381/189); filed in Japan 31 August 2004

The proliferation of hand-held radio frequency devices has increased the possibility that these devices' electromagnetic emissions can interfere with audio equipment. One piece of equipment that may be susceptible to



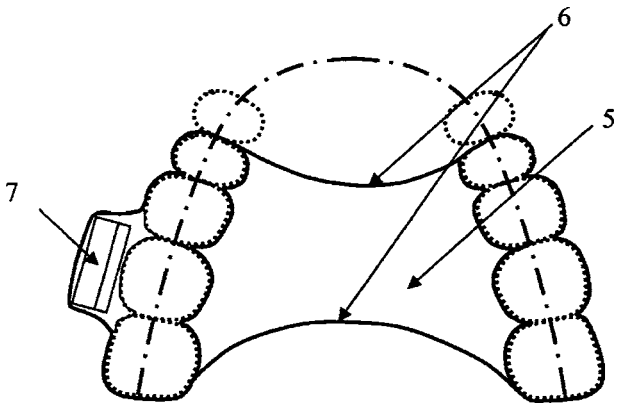
this interference is the electret microphone. Previous designs terminated the shield from cable 30 to the field-effect transistor circuit board 21. By use of sleeve 40 and associated parts, the shield of cable 30 is terminated so that it is bonded to the metal case of microphone 20.—NAS

7,486,798

43.38.Kb METHOD AND APPARATUS FOR TOOTH BONE CONDUCTION MICROPHONE

Muniswamappa Anjanappa *et al.*, assignors to Mayur Technologies, Incorporated
3 February 2009 (Class 381/151); filed 6 April 2005

High sensitivity tooth microphone and associated circuitry 7 is held in place by retainer 5, which features cut-outs 6 so that "impediment for free tongue movement in speech critical areas" are "eliminated." The device may also have a rf transmitter for wireless communication mounted similarly on the other side of retainer 5, and further may also have a tooth receiver and associated circuitry for two-way communication. The device was developed for use in high noise environments to enhance a speech signal, such as those encountered when using a Phraselator (which converts speech from one language to another) and similar devices.—NAS

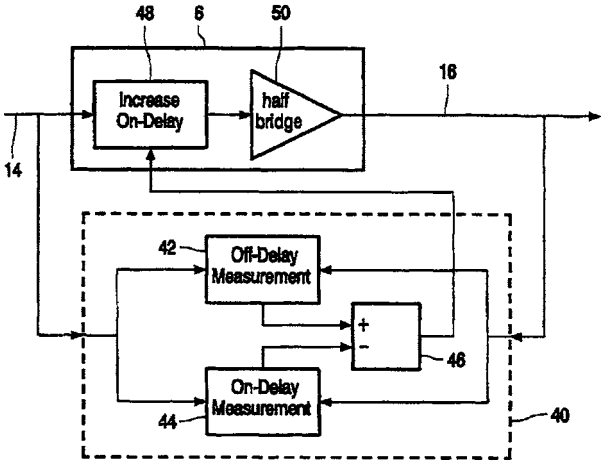


7,518,450

43.38.Lc DIGITAL AMPLIFIER

Matthias Wendt *et al.*, assignors to NXP B.V.
14 April 2009 (Class 330/251); filed in the European Patent Office 7 April 2003

This patent should be read by anyone interested in audio amplifier design. The performance of class D (pulse width modulation) power amplifiers has improved dramatically in the past few years, and class D amplifiers are now routinely used for sound reinforcement and other demanding applications. According to the patent, however, an important source of distortion in class D amplifiers has been largely ignored. Although the output transistors are simply switched on or off, the transition is not instantaneous. "On" delay may differ from "off" delay, causing distortion. Moreover, switching delay may vary depending on the particular device, aging, current, and temperature. What has been patented is the basic concept of measuring on and off delays and applying appropriate compensation in the driving stage.—GLA

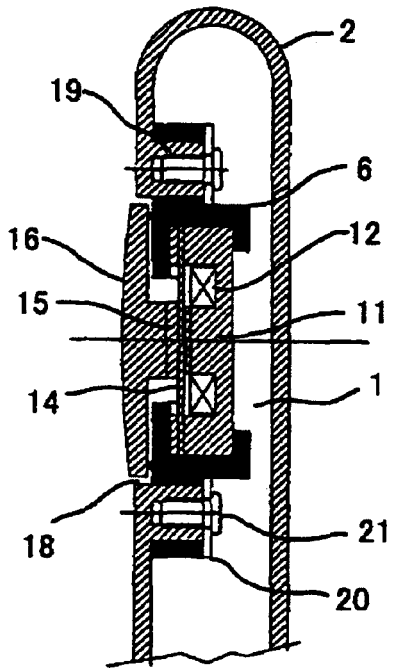


7,519,189

43.38.Lc PROCESSING OF AN AUDIO SIGNAL FOR PRESENTATION IN A HIGH NOISE ENVIRONMENT

Anthony Bongiovi, Port St. Lucie, Florida
14 April 2009 (Class 381/106); filed 18 June 2007

This short patent is a continuation of U.S. Patent No. 7,254,243 filed in 2004. The goal is to provide subjectively pleasing music reproduction in a high noise environment such as a motor vehicle. The signal is first equalized to attenuate low frequencies and boost high frequencies. The equalized signal is then compressed. Finally, reverse equalization is applied to the compressed signal. The result is said to be "…a sound presentation of compressed volume range and a bass-rich sound spectrum."—GLA

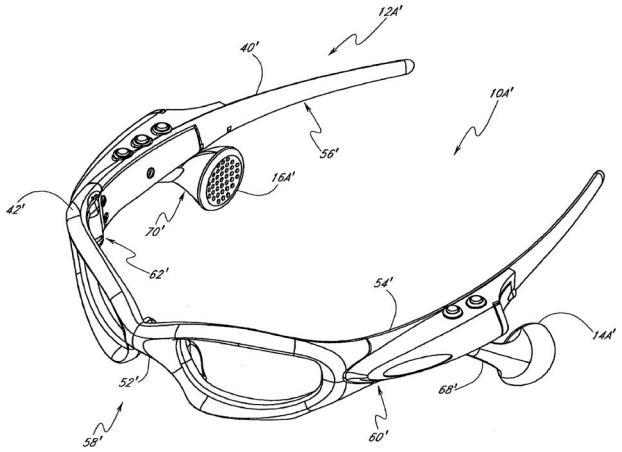


7,512,414

43.38.Si WIRELESS INTERACTIVE HEADSET

James Jannard *et al.*, assignors to Oakley, Incorporated
31 March 2009 (Class 455/556.1); filed 28 July 2003

With the miniaturized components available today, a combination cellular telephone and stereo FM receiver can be contained within an eyeglass frame. The assembly shown includes small loudspeakers positioned near the user's ears rather than sealed headphones.—GLA



7,512,425

43.38.Si PORTABLE TELEPHONE USING BONE CONDUCTION DEVICE

Mikio Fukuda, assignor to Temco Japan Company, Limited
31 March 2009 (Class 455/575.1); filed in Japan 16 January 2004

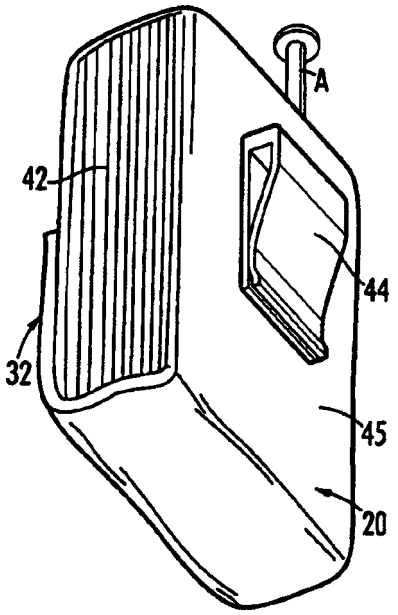
In this cellular telephone design, bone conduction transducer 11 is used for both transmission and reception. Resilient mounting material 6 provides mechanical isolation from case 2. In comparison with prior art, the geometry shown is said to allow thinner overall construction.—GLA

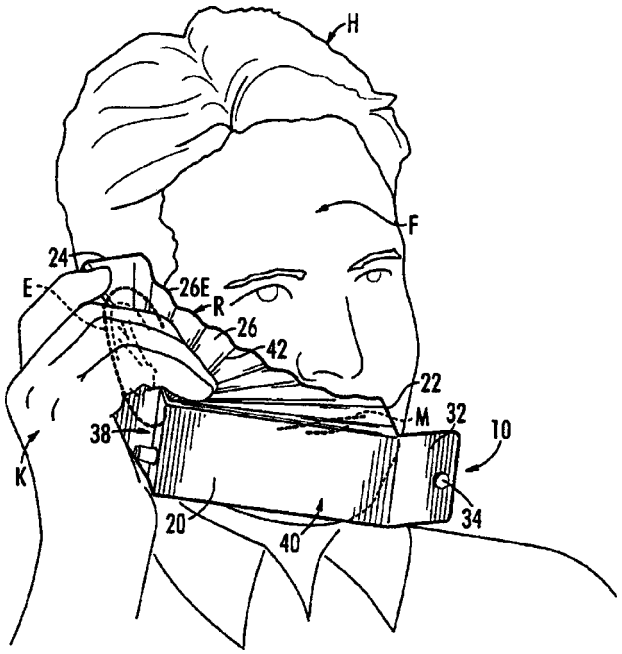
7,515,708

43.38.Si PRIVACY CASE FOR TELEPHONES

Arthur F. Doty III, Edisto, and Robert M. Turkewitz, Charleston, both of South Carolina
7 April 2009 (Class 379/440); filed 29 January 2004

This collapsible cellular phone case expands into a kind of privacy chamber when the phone is in use. Flexible walls 42 are intended to gently seal against the user's mouth and cheek.—GLA



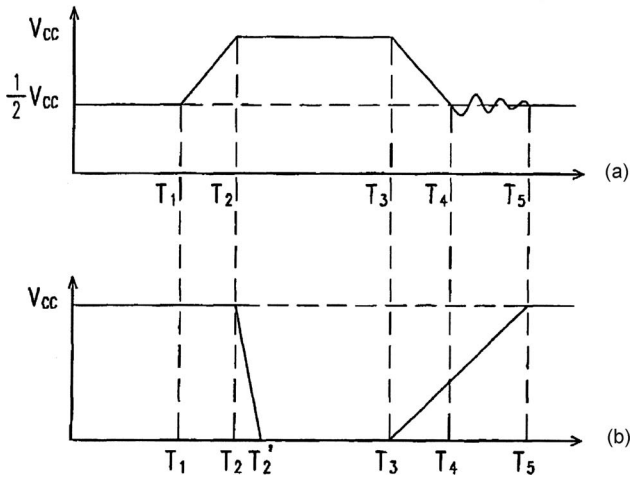


7,519,185

43.38.Si EARPHONE DETECTION CIRCUIT

Chan-Li Liang, assignor to HTC Corporation
14 April 2009 (Class 381/74); filed in Taiwan 15 February 2002

We have reached the stage where a fairly complicated electronic circuit may be smaller, cheaper, and more reliable than a mechanical switch or relay. According to this patent, when you plug a headphone into your TV receiver or camcorder the action does not simply disconnect the loudspeakers but reroutes the audio to a separate channel. To provide a dc path for the detection circuitry, bulky blocking capacitors must be included in the audio path (progress never comes cheaply). The circuit shown is said to provide a simple, inexpensive alternative to the blocking capacitors. The one remaining weak point would seem to be mechanical contact point 322. With a little additional ingenuity, it should be possible to get rid of that as well.—GLA

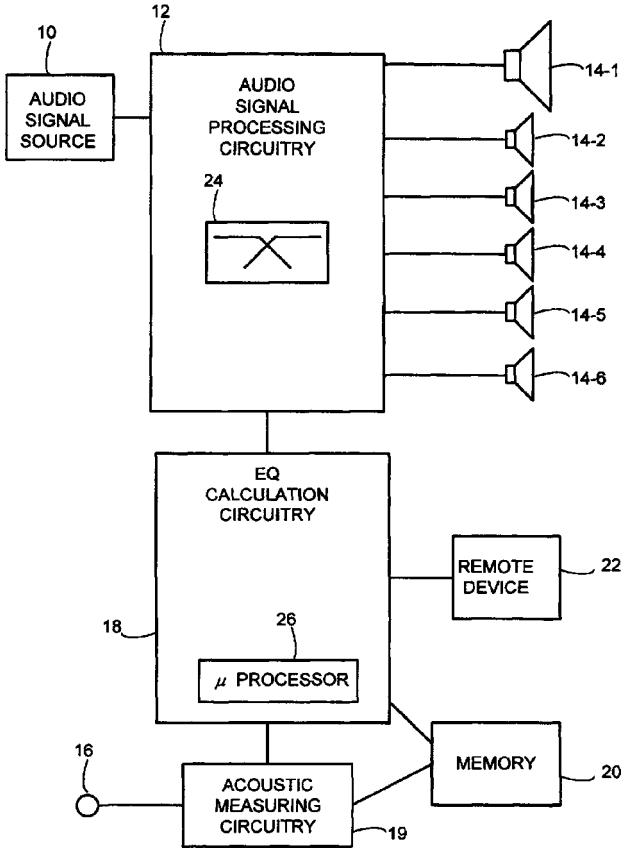


7,483,540

43.38.Tj AUTOMATIC AUDIO SYSTEM EQUALIZING

William M. Rabinowitz *et al.*, assignors to Bose Corporation
27 January 2009 (Class 381/103); filed 25 March 2002

This entry in the auto-equalization category allows for a microphone 16 which is mounted in a headphone-like device worn by a tester/user, who is prompted via device 22 to move to the next measurement location in a space, so that equalization calculation processing unit 18 can modify the settings in audio signal processing circuitry 12 in order to optimize the sound produced by loudspeakers 14.—NAS

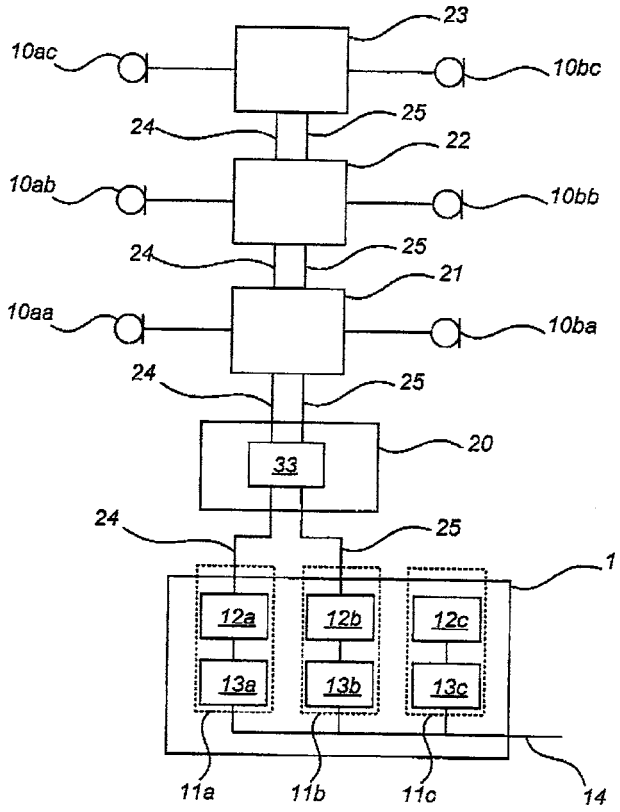


7,492,914

43.38.Tj AUDIO RECEIVING SYSTEM

Olof Arvidsson, assignor to Arva Trade
17 February 2009 (Class 381/92); filed 8 March 2002

Satellite units 21 22 23 accept microphone and/or line level signals, and combine and add them as one moves toward master unit 20, which in turn can be connected to traditional mixer unit 14, to reduce the number and length of cables needed to connect sources, in this case 10ax 10bx, to the control device 14. How this is done is described in what can be called very general terms. What is not stated is how one can adjust gain trim, equalization, and other parameters for each microphone/line source via the satellite and/or master unit.—NAS

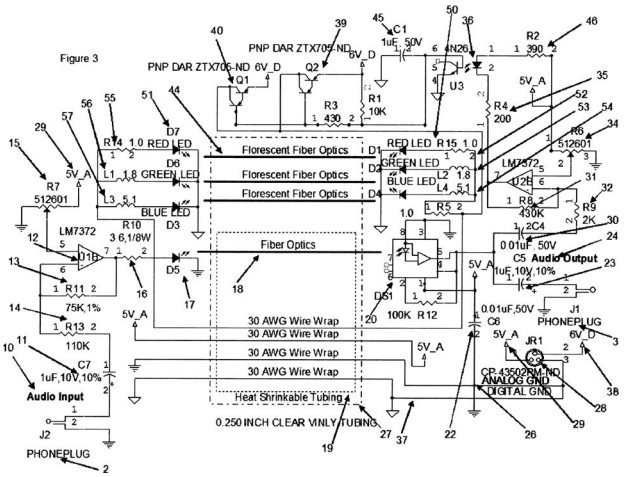


7,493,049

43.38.Tj AUDIO OPTIC CABLE

Donald Conemac, Santa Clarita, California
17 February 2009 (Class 398/141); filed 23 November 2005

An improved guitar cord, in which an arrow on the cord is actually necessary, is described that converts the electrical signal to an optical one at the guitar end in plug body 2. The amplitude modulated light signal is conveyed via fiber optic cable 18, which is then converted back to an electrical signal using circuitry in plug body 3. The cord can be implemented for



signal only, and by use of fluorescent fiber optics 44 the cable innards [within the "vinly" (sic) outer jacket] can emit either monochromatic or trichromatic light by selection of LEDs 50-51. The cable is said to be more robust and have better frequency response and dynamic range than conventional copper based cords as well as being electrically safe. This may be, but you need to have another connector 28 at the amplifier end, or a special connector, in order to power the assembly.—NAS

7,506,735

43.40.Tm VIBRATION DAMPING AND HEAT TRANSFER USING MATERIAL PHASE CHANGES

Petr Kloucek and Daniel R. Reynolds, assignors to William Marsh Rice University
24 March 2009 (Class 188/378); filed 14 June 2004

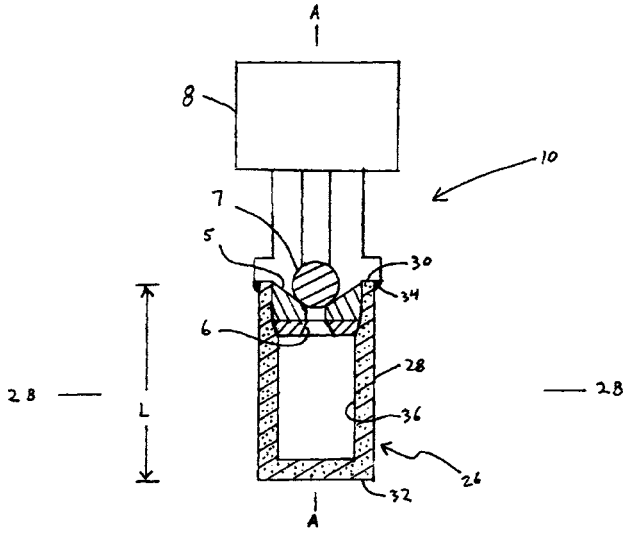
A layer of a shape memory alloy, which undergoes phase changes as its temperature changes, is adhered to a substrate or to an item whose vibrations are to be damped. Localized heating and cooling elements are attached to the alloy. These elements are actuated in response to signals generated by a vibration sensor so that vibrational energy is removed from the assembly due to hysteresis resulting from the induced phase changes. The patent indicates neither the control system that is to be used nor the fact that this use of shape memory alloys is limited to low frequencies in view of the inevitable slowness of the heat transfer process.—EEU

7,412,972

43.50.Gf DISPERSION-TYPE SUPPRESSOR FOR ACOUSTIC NOISE REDUCTION OF A GASEOUS FUEL INJECTOR

Perry Robert Czimmek, assignor to Continental Automotive Systems US, Incorporated
19 August 2008 (Class 123/527); filed 3 March 2005

A compressed natural gas (CNG) internal combustion engine fuel injector 10 noise suppressor 28 is claimed comprising a porous sintered 316L stainless steel enclosure 32 through which gas and noise created by the release of CNG from manifold 8 by valve 5–7 through orifice 6 must flow. Noise reduction occurs via the flow resistance of 1.5 mm thick, 25 mm long cylindrical wall 32–36 having 40 μm pore size. Noise reduction of 3 or 5 dBA results from length L being 12.5 or 25 mm.—AJC



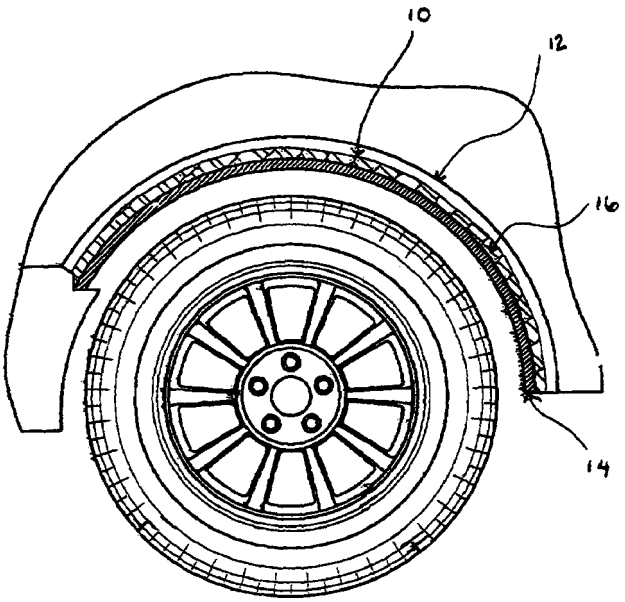
7,419,188

43.50.Gf VEHICLE UNDERSHIELD

Norbert Nicolai and Nelson Dias, assignors to Entwicklungsgesellschaft fuer Akustik (EFA) mit Beschraenkter Haftung
2 September 2008 (Class 280/847); filed in Germany 10 July 2001

A sound and sandblast absorbing wheel shield 14–16 is claimed where self-supporting and self-cleaning smooth-faced plastic sheet 14 faces the wheel. To absorb sound, needle-punched web material 16 is attached to the back side of impervious sheet 14, clear of wheel well 12. The elastic modu-

lus of sheet 14 is preferably between 600 and 1000 MPa with a loss factor of 0.1–0.3 and area density of 500–2000 g/m^2. The needle punched material should have a similar area density.—AJC



7,510,052

43.50.Gf ACOUSTIC SEPTUM CAP HONEYCOMB

Earl Ayle, assignor to Hexcel Corporation
31 March 2009 (Class 181/292); filed 22 March 2007

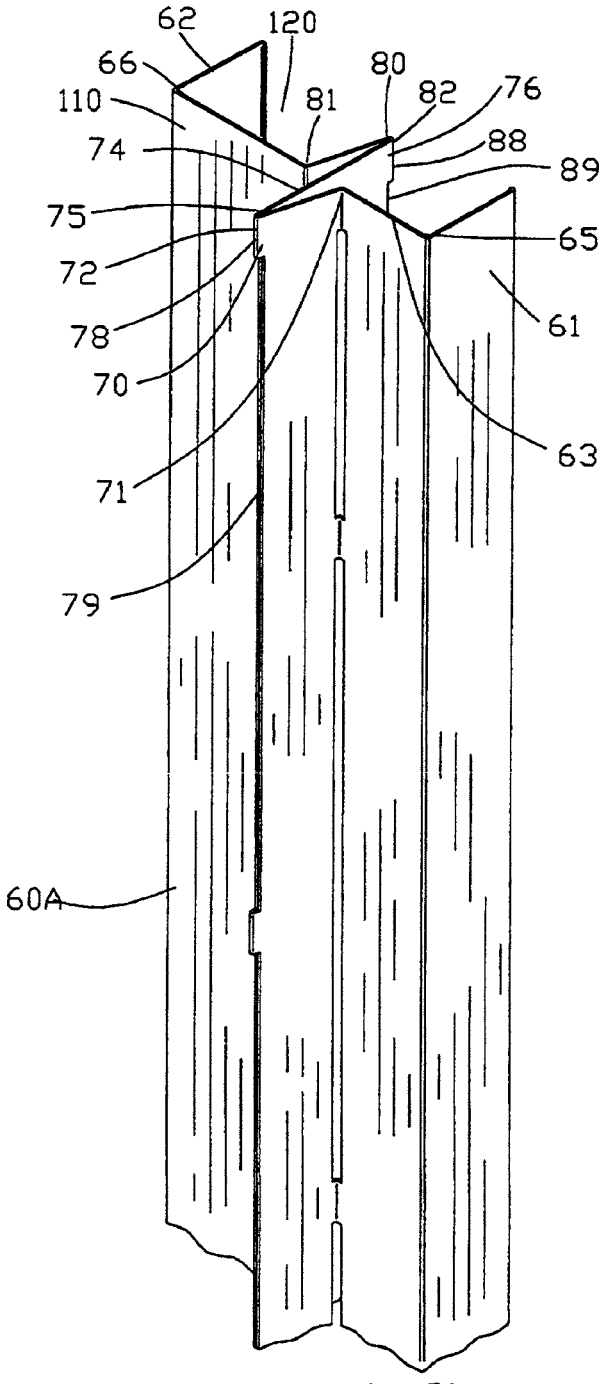
This is a mechanical support for a septum cap inserted into the cells of a honeycomb muffler of a jet engine cell. The method by which this cap is secured provides greater stability and acoustical control.—CJR

7,513,082

43.55.Rg SOUND REDUCING SYSTEM

Lahnie Johnson, Tampa, Florida
7 April 2009 (Class 52/144); filed 8 February 2005

The supports for a self-contained wall are attached through folded flanges that introduce resiliency (or separation) between the two sides of the wall that would otherwise be rigidly braced to each other.—CJR



7,510,771

43.55.Wk SOUND REDUCING POLYMER INTERLAYERS

Jun Lu, assignor to Solutia Incorporated
31 March 2009 (Class 428/437); filed 17 March 2005

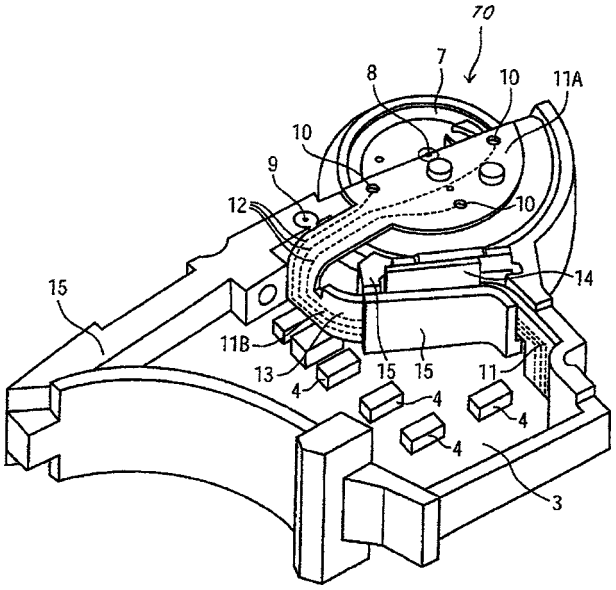
This patent claims to enhance the sound attenuation properties of lamina ted glass by using polymer interlayers composed of multiple thermoplastic sheets.—CJR

7,507,920

43.66.Ts HEARING AID WITH A CONTROL ELEMENT

Ehrenfried Erbe, assignor to Siemens Audiologische Technik GmbH
24 March 2009 (Class 200/14); filed in Germany 15 February 2005

To conserve space utilized in small hearing aids and to prevent damage to their subminiature components, the control element of a multi-function, user-activated switch is mounted directly to a flexible printed circuit board with other electrical components. The flex substrate has at least one bending zone to withstand mechanical stress. To provide multiple functions, such as a volume control and a switch to change memories, the control element may be rotated about a first axis and swiveled about a second axis.—DAP

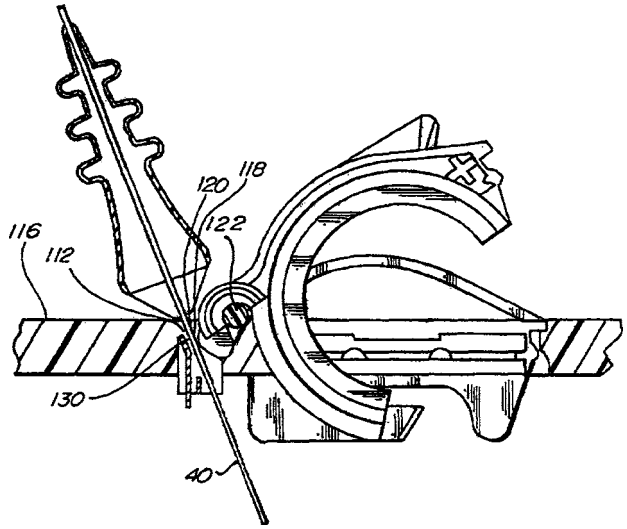


7,508,949

43.66.Ts FACE PLATE CONNECTOR FOR HEARING AID

Mark A. Neilson, assignor to In'Tech Industries, Incorporated
24 March 2009 (Class 381/314); filed 29 March 2005

In a variation on an old idea, a programmable custom hearing aid system comprises several programming terminals located in a very narrow slot behind a hinged battery drawer in the hearing aid faceplate and a flexible programming strip fits within the slot, mating with the programming terminals. When the battery drawer is closed, it presses the programming strip electrodes against the terminals.—DAP

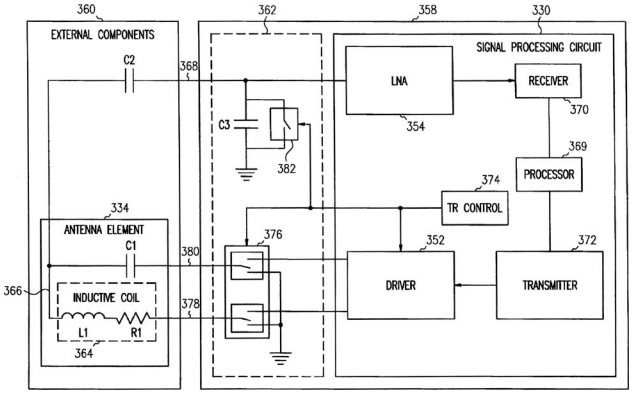


7,512,383

43.66.Ts TRANSMIT-RECEIVE SWITCHING IN WIRELESS HEARING AIDS

Mohamad Essabar and John David Terry, assignors to Starkey Laboratories, Incorporated
31 March 2009 (Class 455/78); filed 26 November 2003

A wireless hearing aid uses a single antenna coil that is switched to different connections in the radio for transmit and receive mode operation. Tuning is accomplished by transforming the coil, tuning capacitor, and dc blocking cap into an equivalent series resonant circuit to reduce the inductive load in transmit mode and into an equivalent parallel resonant circuit to increase the inductive load in receive mode.—DAP



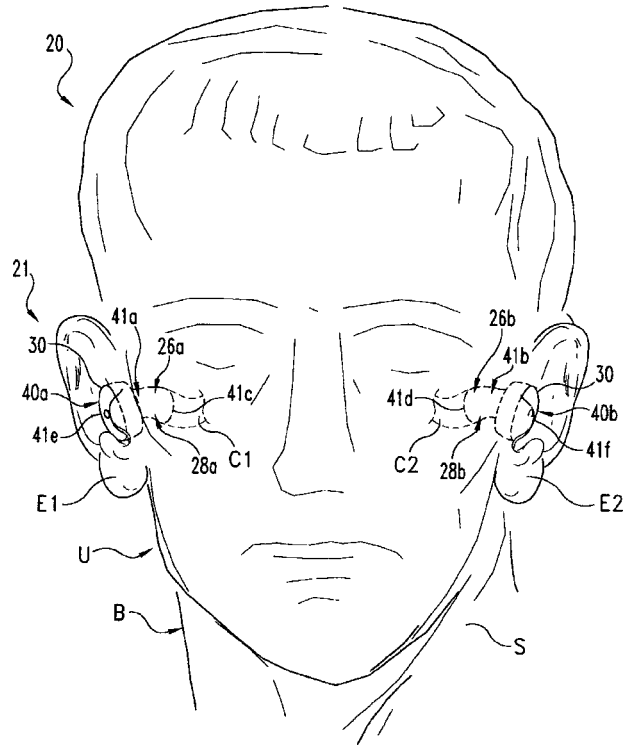
7,512,448

43.66.Ts ELECTRODE PLACEMENT FOR WIRELESS INTRABODY COMMUNICATION BETWEEN COMPONENTS OF A HEARING SYSTEM

Crista Malick *et al.*, assignors to Phonak AG
31 March 2009 (Class 607/136); filed 10 January 2003

This patent involves bi-directional wireless communication between hearing device components mounted on or off the wearer's body. A pair of electrodes mounted on the hearing device is placed against the wearer's skin in one ear canal or behind the ear. A time-varying electrical potential between the electrodes carries electrical signals through the wearer's body to other components of the hearing system, such as a similar device in the

other ear or an implantable component. Multiple electrode pairs may operate as dipole antennae to communicate with off-the-body devices.—DAP



7,515,720

43.66.Ts METHOD AND DEVICE FOR ADAPTING HEARING AIDS

Christoph Schwob, assignor to Audiocare AG
7 April 2009 (Class 381/60); filed in Switzerland 1 December 2003

To assist the hearing aid wearer in identifying and reacting to different acoustic listening environments during the fitting process, typical visual life situations that correspond to the sound examples are presented simultaneously via video recordings.—DAP

7,515,725

43.66.Ts ELECTRONIC BATTERY-POWERED HEARING INSTRUMENT, AND METHOD FOR OPERATING SUCH A HEARING INSTRUMENT

Martin Fluit, assignor to Phonak AG
7 April 2009 (Class 381/312); filed 5 March 2004

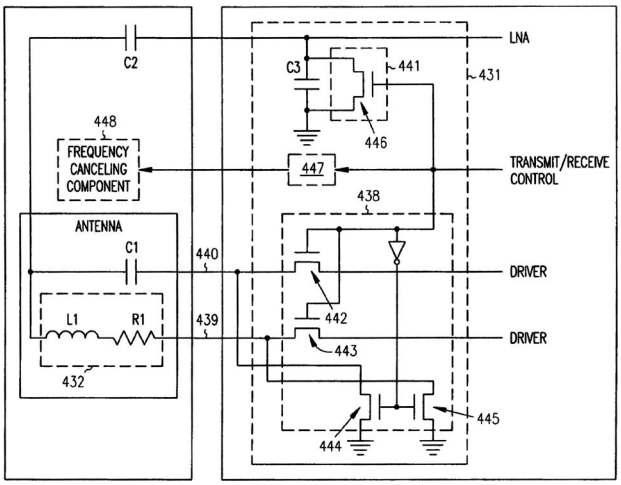
One problem with hearing aid use is that the battery may be inserted upside down inadvertently. A rectifier circuit in the hearing aid provides the correct positive and negative voltages to the first and second power supply terminals, regardless of how the battery is inserted. One terminal of the input device is connected to the first power supply terminal, and a second terminal is connected to a voltage of an interface line via a single electrical contact without need for a ground or reference voltage. To determine if the status of an input device, such as a memory select switch, has changed, the rectifier status may be determined.—DAP

7,515,881

43.66.Ts RESONANCE FREQUENCY SHIFT CANCELING IN WIRELESS HEARING AIDS

Mohamad Essabar, assignor to Starkey Laboratories, Incorporated
7 April 2009 (Class 455/78); filed 26 November 2003

The goal is to prevent detuning caused by a resonant frequency shift in the radio of a wireless hearing aid that uses a single antenna coil which is switched to different connections for transmit and receive modes of operation. The frequency shift occurs because a dc blocking capacitor contributes significantly to the resonant frequency in the transmit mode, but not in the receive mode. The frequency shift is prevented, resulting in substantially equal first and second resonant frequencies for transmit and receive modes, by excluding or including a predetermined capacitor in parallel or in series with a tuning capacitor in the resonant circuit comprising a coil and tuning capacitor.—DAP

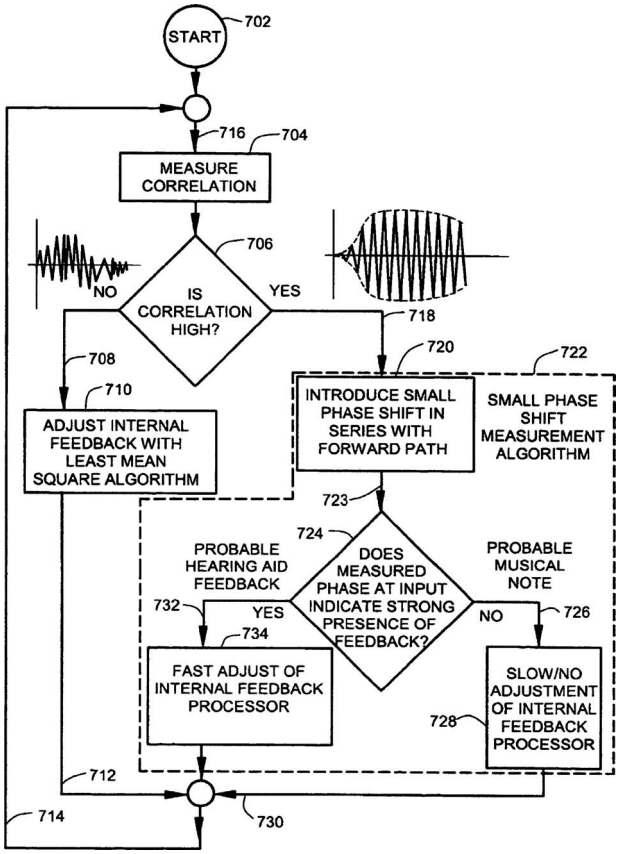


7,519,193

43.66.Ts HEARING AID CIRCUIT REDUCING FEEDBACK

Robert J. Fretz, assignor to Resistance Technology, Incorporated
14 April 2009 (Class 381/312); filed 1 September 2004

Undesirable audio artifacts may be created when a hearing aid acoustic feedback canceler erroneously attempts to cancel environmental signals, such as musical sounds, rather than acoustic feedback signals. This problem is said to be avoided by determining whether the signal source is environmental or feedback by using the output of a correlation detector at the input to inject a small, temporary phase shift in the forward signal path. If the phase of the input signal summer output does not change correspondingly, the source is assumed to be environmental, and the feedback canceler operation is slowed or suspended.—DAP

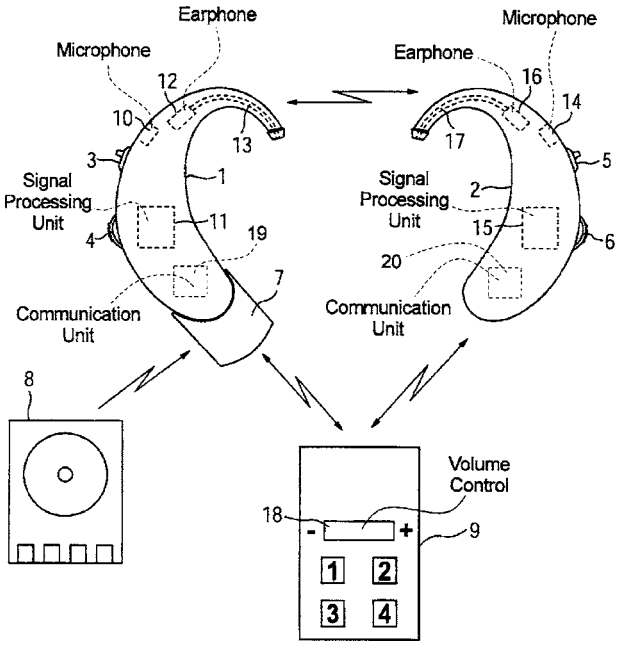


7,519,194

43.66.Ts HEARING AID SYSTEM AND OPERATING METHOD THEREFOR IN THE AUDIO RECEPTION MODE

Torsten Niederdränk and Gottfried Rückerl, assignor to Siemens Audiologische Technik GmbH
14 April 2009 (Class 381/330); filed in Germany 21 July 2004

In a wireless binaural hearing aid system, left and right hearing aids communicate automatically with each other to synchronize their operating parameters. However, if a direct audio input signal is supplied to one of the hearing aids from an external audio source, for example, via a plug-in audio shoe accessory, synchronization of the parameter settings of both hearing aids for varying acoustic environments is not desirable. Consequently, in this audio reception mode case, the parameter synchronization is limited to those relating to different audio sources.—DAP



7,269,560

43.72.−p SPEECH DETECTION AND ENHANCEMENT USING AUDIO/VIDEO FUSION

John R. Hershey *et al.*, assignors to Microsoft Corporation
11 September 2007 (Class 704/270); filed 27 June 2003

This patent is concerned with a method for harnessing information in a video of the talking face, in order to improve automatic speech recognition results. Like most Microsoft patents, it is quite detailed and worth studying. It specifically describes a "generative video model" that is able to learn to follow lip movements during speech. It is trained using expectation-maximization. "The system is adaptive to lip video from various angles." The speech enhancement method can infer correlations between the speech signal and pixel-based image data, by fusing the video model with an audio mixture model.—SAF

7,337,107

43.72.Ar PERCEPTUAL HARMONIC CEPSTRAL COEFFICIENTS AS THE FRONT-END FOR SPEECH RECOGNITION

Kenneth Rose and Liang Gu, assignors to The Regents of the University of California
26 February 2008 (Class 704/208); filed 2 October 2001

This patent first discusses a well-known problem with the usual acoustic speech features such as mel-frequency cepstral coefficients or perceptual linear predictive features—namely, that the transformed smoothed spectrum which these represent does not closely approximate the "spectral envelope" of voiced sounds, which is the desired information for speech recognition. Instead, the authors propose to use "perceptual harmonic cepstral coefficients," which are said to be derived directly from the spectral envelope by sampling it at the harmonic locations only. According to the patent, error rates in a Mandarin isolated digit recognition task decrease from 46% to 29% under a 0 dB white noise signal-to-noise ratio condition, when the perceptual harmonic cepstral coefficients are employed in place of mel-frequency cepstral coefficients. It is laudable for actually working on the recognition front end.—SAF

7,337,114

43.72.Ar SPEECH RECOGNITION USING DISCRIMINANT FEATURES

Ellen M. Eide, assignor to International Business Machines Corporation
26 February 2008 (Class 704/243); filed 29 March 2001

A method is documented for augmenting a speech recognition front-end by computing values of "distinctive features" of the sort employed in linguistic phonetics and phonology. The patent uses the terminology correctly, notwithstanding the incoherent "discriminant features" of the title. A reduction in word error rate of 33% is claimed for a system which includes such a component. The feature information is envisioned as being appended to a standard sort of cepstral representation. Gaussian mixture models are constructed for each feature; the speech waveform is then modeled under the presence and absence of each feature. The set of features modeled is then pruned by a mutual information criterion. The patent is valuable, if only for the relative rarity of actually redesigning the front end.—SAF

7,343,284

43.72.Dv METHOD AND SYSTEM FOR SPEECH PROCESSING FOR ENHANCEMENT AND DETECTION

Saeed Gazor and Mohamed El-Hennawey, assignors to Nortel Networks Limited
11 March 2008 (Class 704/226); filed 17 July 2003

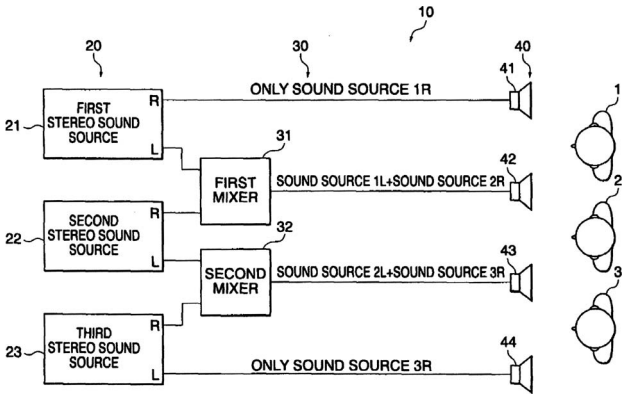
This patent exploits the rather recent idea (a good sign) that the feature vectors of speech are better modeled using some type of Laplace distribution rather than the usual Gaussians (although the incorrect term "Laplacian" is used throughout the patent). Meanwhile, it is also recognized that Gaussians make better models for noise. This leads to a method for separating speech from noise, which is one of those problems that is constantly worked on. A scheme is described in detail for obtaining and updating parameters of Laplace and Gaussian distributions from noise-contaminated speech. The feature vectors employed can be any of the decorrelated spectral variety.—SAF

7,512,246

43.72.Gy STEREO SOUNDING METHOD, STEREO SOUNDING SYSTEM, AND MUSICAL TONE GENERATION CONTROL SYSTEM

Yoshiki Nishitani *et al.*, assignors to Yamaha Corporation
31 March 2009 (Class 381/119); filed in Japan 28 April 2003

The goal is to produce realistic stereo sound for a large number of listeners with fewer loudspeakers than twice the number of stereo sound sources to be reproduced. The right-channel (R) output from at least one stereo sound source is mixed with the left-channel (L) output from another stereo sound source and the sum is output to at least one speaker. The result is that the R speaker for one stereo sound source also serves as the L speaker for the other stereo sound source, thereby reducing the number of required speakers.—DAP

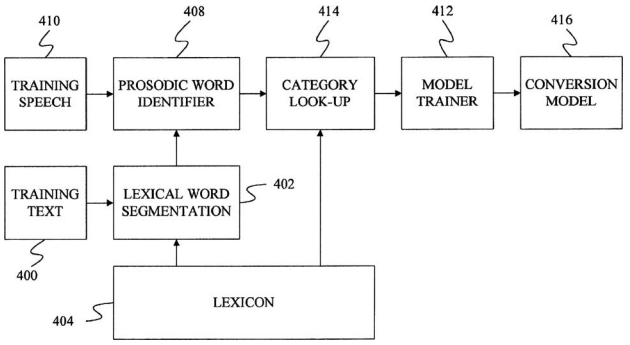


7,263,488

43.72.Ja METHOD AND APPARATUS FOR IDENTIFYING PROSODIC WORD BOUNDARIES

Min Chu and Yao Qian, assignors to Microsoft Corporation
28 August 2007 (Class 704/251); filed 7 May 2001

In Asian languages with non-alphabetic writing systems, a recalcitrant problem for text-to-speech synthesis stems from the ambiguous representation of so-called prosodic words in text. As stated, "the lexical word boundaries are not apparent from the text." Moreover, the "basic rhythm units" (prosodic words) often cross-cut the sequence of lexical words in ways which are not easily predicted, making it insufficient for a synthesis system to properly recognize the lexical words. This patent broadly lays claim to the general procedure of first segmenting a character-based text into lexical words, and then determining the prosodic words which may be used by a synthesis system. "In many embodiments, prosodic word identifier 408 is a panel of human listeners who listen to training speech signal 410 while reading the training text." Boundaries of prosodic words are then placed wherever a majority of the panel agrees.—SAF



7,269,557

43.72.Ja COARTICULATED CONCATENATED SPEECH

Scott J. Bailey and Nikko Strom, assignors to Tellme Networks, Incorporated
11 September 2007 (Class 704/251); filed 19 November 2004

Methods are described for "reducing the audible gap in concantenated recorded speech" used in speech synthesizers. The word "methods" is particularly appropriate in this case, since the patent describes in essence how to manually edit recorded speech in order to clip out desired words or phones so that their coarticulation portions are included right up to a point in the recording where the next sound could be heard. This "method" is called "coarticulation." I have little doubt that manually editing a speech synthesis database to include all possible coarticulated forms of speech sounds would certainly improve the quality of the synthesized speech. I do have consider-

able doubt that this is a practical procedure to carry out by hand for all the diphones or what not in a typical synthesizer. The envisioned application appears to be limited to prerecorded prompts or other longer speech fragments.—SAF

7,333,932

43.72.Ja METHOD FOR SPEECH SYNTHESIS

Horst-Udo Hain, assignor to Siemens Aktiengesellschaft
19 February 2008 (Class 704/258); filed in Germany 31 August 2000

A method is described for improving speech synthesis, which appears to be independent of the specific synthesis technology. The patent deals with the "out-of-vocabulary" (OOV) problem. In a typical text-to-speech synthesizer, unknown words are "sounded out" using grapheme-to-phoneme conversion rules, which we all know are pretty awful much of the time. The idea here is to leverage knowledge of "subwords" which may be present within OOV items, and to then assist the grapheme-to-phoneme conversion process by introducing pronunciation information about any known subwords. The success of the technique depends chiefly on the number of OOV items encountered which do contain known subwords. It is significant that the language for which the idea was conceived appears to be German.—SAF

7,334,183

43.72.Ja DOMAIN-SPECIFIC CONCATENATIVE AUDIO

Christopher Rusnak *et al.*, assignors to Oracle International Corporation
19 February 2008 (Class 715/500.1); filed 29 May 2003

This is an incredible patent. To be specific, it is incredible that it was granted. It somehow lays claim to the most simplistic variety of concatenative speech synthesis, as the described method involves finding recordings of individual words or "substrings" in a database according to the substrings parsed from an input text, and then splicing the recordings together to make a synthesized utterance. It hardly needs stating that not only is this idea very simple and probably quite ineffective but it is also very old.—SAF

7,266,497

43.72.Lc AUTOMATIC SEGMENTATION IN SPEECH SYNTHESIS

Alistair D. Conkie and Yeon-Jun Kim, assignors to AT&T Corporation
4 September 2007 (Class 704/258); filed 14 January 2003

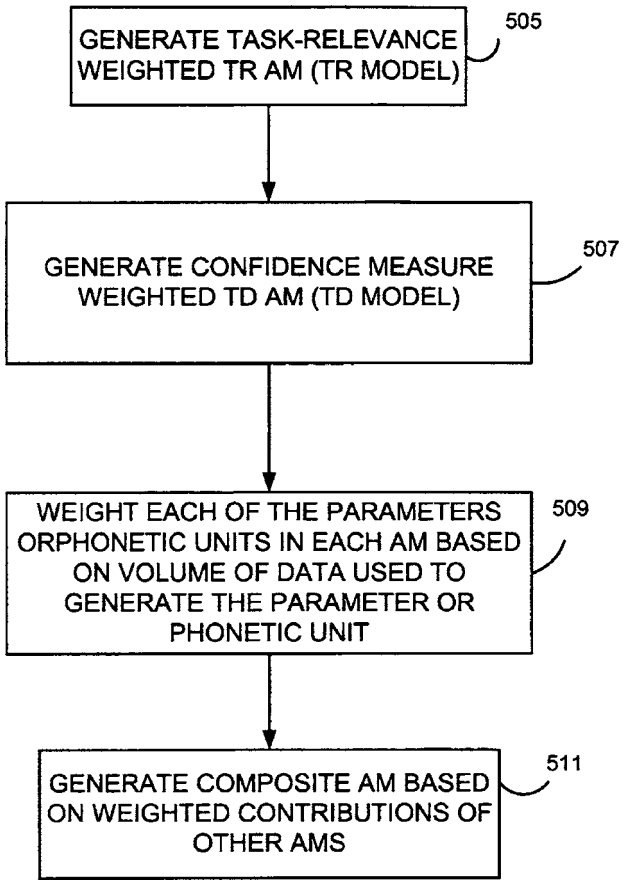
Typical concatenative speech synthesis is conducted using recorded speech units from a large corpus that must be managed somehow. A common procedure uses hidden Markov models (HMMs) to automatically segment the corpus into phones and provide phonetic labels. In a typical form, this procedure introduces badly segmented speech. This patent presents an iterative procedure involving HMM segmentation followed by a step which corrects the bad segmentation. The two steps may then be repeated at will. The crux of the patent lies in the second step. The general idea is that true phone boundaries often occur at a "bending point" of a spectral transition, which can be measured using a weighted slope metric over a sequence of cepstral coefficient vectors.—SAF

7,263,487

43.72.Ne GENERATING A TASK-ADAPTED ACOUSTIC MODEL FROM ONE OR MORE DIFFERENT CORPORA

Mei Yuh Hwang, assignor to Microsoft Corporation
28 August 2007 (Class 704/243); filed 29 September 2005

Methods are disclosed for developing a task-related acoustic model for a speech recognizer. First, each word in the training corpus can be weighted according to its task relevance. A 3% error rate reduction is a reported result of this procedure alone. For the next step, unsupervised speech data that are known to be task-dependent can be provided to a speech recognizer. An acoustic model is then generated based on the recognition result, by weighting the speech data according to a confidence measure "so that only data that have been recognized with a high degree of confidence will weigh heavily in generation of the acoustic model." Lastly, the Gaussian mean of the resulting task-dependent acoustic model is smoothed together with the mean of the previous task-related model. Equations and methods are provided in detail.—SAF



7,269,555

43.72.Ne UNSUPERVISED INCREMENTAL ADAPTATION USING MAXIMUM LIKELIHOOD SPECTRAL TRANSFORMATION

Dongsuk Yuk and David M. Lubensky, assignors to International Business Machines Corporation
11 September 2007 (Class 704/234); filed 30 August 2005

A method is described which purports to allow "rapid speaker and environment adaptation" in a speech recognition system. The method involves transforming feature vectors of received speech in order to match the expectations of the recognition system. The "maximum likelihood spectral transformation" models the differences between training speech and real

utterances by separating the two factors of convolutional noise and additive noise. The procedure estimates these parameters "such that the likelihood of testing utterances is increased" when the transformed features are used in a linear spectral space. A method for estimating the noise parameters so that this likelihood is maximized forms the core of the patent.—SAF

7,272,561

43.72.Ne SPEECH RECOGNITION DEVICE AND SPEECH RECOGNITION METHOD

Toshiyuki Miyazaki and Yoji Ishikawa, assignors to Asahi Kasei Kabushiki Kaisha
18 September 2007 (Class 704/256); filed 13 July 2001

This patent presents a somewhat different sort of algorithm for combining hidden Markov models in speaker-independent recognition. It is quite common to have a number of different models in these circumstances, pertaining to different speaker classes such as male vs female. The inventors claim a new method of combining these for a Viterbi search which saves on storage and computational steps over the most common approaches. Their successful results appear in the figure.—SAF

RECOGNITION METHOD	AVERAGE RECOGNITION RATE	NO. OF PERSONS WITH RECOGNITION RATE OF LESS THAN 60%	WORKING MEMORY (RAM)	VOCABULARY STORAGE MEMORY (ROM)
COMMON ACOUSTIC MODEL FOR BOTH MALES AND FEMALES	60.1(%)	8	19.5KBYTES	6.4KBYTES
SEPARATE MULTI-TEMPLATE MODELS FOR MALES AND FEMALES	71.8(%)	4	29.5KBYTES	12.8KBYTES
EMBODIMENT OF THIS INVENTION	74.0(%)	3	19.5KBYTES	6.4KBYTES

7,340,396

43.72.Ne METHOD AND APPARATUS FOR PROVIDING A SPEAKER ADAPTED SPEECH RECOGNITION MODEL SET

Mark Thomson *et al.*, assignors to Motorola, Incorporated
4 March 2008 (Class 704/234); filed 18 February 2003

This is one of many patents working on the problem of adapting a speaker-independent acoustic "feature space" model set to the speech of a given speaker. A speaker-dependent feature space model set has to be obtained from the speech of one person, even if poorly estimated such as by a single set of speech feature vectors. Then the methods concentrate on determining the alignment between the speaker-dependent model set and the speaker-independent model set. A "transformation estimation unit" receives the alignment information and determines a transformation from the speaker-independent model means to the speaker feature model classes. Then this information is passed to the "model transformation unit," which applies the indicated transformations to the appropriate hidden Markov model probability density functions.—SAF

7,509,258

43.72.Ne PHONETIC, SYNTACTIC AND CONCEPTUAL ANALYSIS DRIVEN SPEECH RECOGNITION SYSTEM AND METHOD

Philippe Roy, assignor to Conceptual Speech LLC
24 March 2009 (Class 704/256); filed 18 August 2007

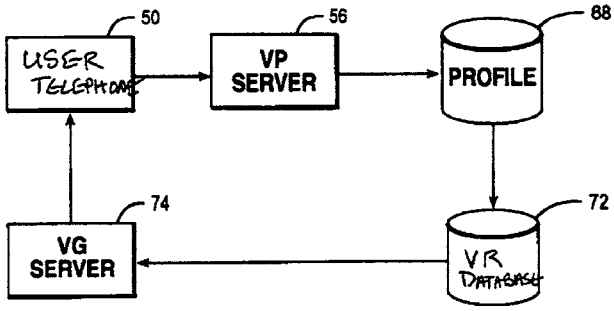
The goal is speaker-independent speech recognition that will recognize words uttered by persons having different accents without use of training, word spotting, context-free grammars, or comparison of incoming audio to stored phoneme profiles. Multiple phonemes are recognized and used to build words, and sequences of words are built via syntactic analysis. Conceptual analysis is then used to analyze for selected syntactically valid sequences of words applicable to context, optionally using artificial intelligence and recurrent neural networks. In the process, potential words and syntactic structures are assigned a score or probability.—DAP

7,515,694

43.72.Ne APPARATUS FOR PERSONALIZING COMPUTERIZED CUSTOMER SERVICE

Nathan Raymond Hughes *et al.*, assignors to International Business Machines Corporation
7 April 2009 (Class 379/88.06); filed 26 June 2003

To improve user satisfaction, the language usage pattern of computer-driven interactive responses are adapted automatically to previously stored language usage pattern preferences of the user, including personal pronoun preferences, pacing, and volume, independent of the language spoken or written by the user.—DAP



7,516,070

43.72.Ne METHOD FOR SIMULTANEOUSLY CREATING AUDIO-ALIGNED FINAL AND VERBATIM TEXT WITH THE ASSISTANCE OF A SPEECH RECOGNITION PROGRAM AS MAY BE USEFUL IN FORM COMPLETION USING A VERBAL ENTRY METHOD

Jonathan Kahn, assignor to Custom Speech USA, Inc.
7 April 2009 (Class 704/235); filed 19 February 2004

This patent addresses speaker-trained speech recognition software systems widely used on personal computers to convert spoken material into a text format for generating reports, filling in forms, or for further editing. Three problem areas with such software are discussed: text version mismatches, large file sizes, and form entry alignment problems. Because of spoken asides, corrections, verbal instructions to an assistant, etc., a text which matches the spoken material may differ from the desired final text. Existing software packages typically do not handle this situation well. The patented system provides a more general method of segmenting portions of the spoken and textual material and associating the portions with other objects, such as blanks of a form. This segmentation system is also used to isolate portions of the spoken material, indicating whether such portions are to be retained in a final text version. A file compression system uses the segmentation system to improve handling of the materials, such as across a network link. Further, an improved system is said to allow better recognition accuracy with a less cumbersome user enrollment process.—DLR

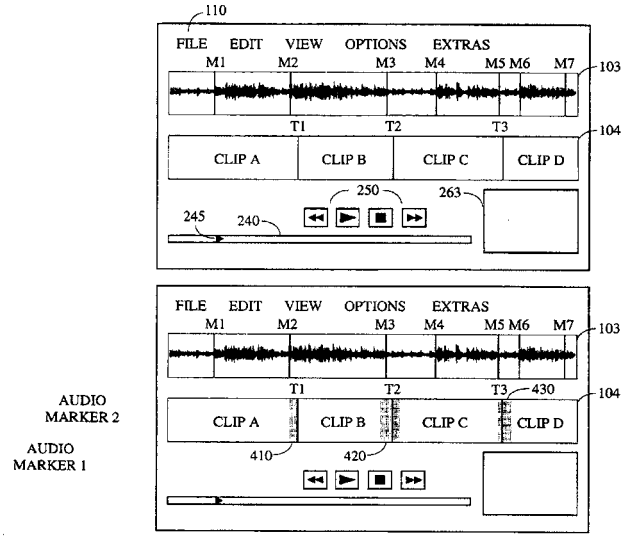
7,512,886

43.75.Zz SYSTEM AND METHOD OF AUTOMATICALLY ALIGNING VIDEO SCENES WITH AN AUDIO TRACK

Tilman Herberger and Titus Tost, assignors to Magix AG
31 March 2009 (Class 715/723); filed 15 April 2004

The patent describes a method of aligning audio soundtracks, which

may be music, narration, or sound effects, to segments of prerecorded stills or video material. A screen display shows what appears to be a fairly typical video editing facility. Using a mouse, for example, the user would place marks in both the audio and the video material indicating the breaks between the segments. Special marks may also be added to indicate certain time points, such as musical beats, video lighting changes, etc. The patented software would then adjust the durations, typically of the audio track, but possibly also the video material, so as to time-align the marked breaks. Special functions are available to adjust the timing of selected segues, to insert, to perform fades, etc. Various options allow manual overrides of the programmed transitions.—DLR



7,507,204

43.80.Vj APPARATUS AND METHOD FOR FORMING 3D ULTRASOUND IMAGE

Jaeyoon Shim *et al.*, assignors to Medison Company, Limited
24 March 2009 (Class 600/443); filed in Republic of Korea 1 April 2004

A region of interest is defined using a two-dimensional ultrasound image. The boundary of an object in the region of interest is detected. The detected boundary is used to render a three-dimensional ultrasound.—RCW

7,510,529

43.80.Vj ULTRASOUND RECONSTRUCTION UNIT

Ching-Hua Chou *et al.*, assignors to ZONARE Medical Systems, Incorporated
31 March 2009 (Class 600/443); filed 13 June 2006

Sets of echo signals in a receive aperture are adaptively selected by a controller, and the selected echo signals are passed to an image reconstruction processor.—RCW

7,513,870

43.80.Vj FINITE AMPLITUDE DISTORTION-BASED INHOMOGENEOUS PULSE ECHO ULTRASONIC IMAGING

Ted Christopher, assignor to Research Corporation Technologies, Incorporated
7 April 2009 (Class 600/443); filed 18 August 2004

An ultrasound signal is transmitted into a sample. First-order and higher-order signal components resulting from propagation and reflection of the signal are received. One of the higher-order components of the received signal is processed to form an image.—RCW

7,517,317

43.80.Vj SUBARRAY FORMING SYSTEM AND METHOD FOR ULTRASOUND

John C. Lazenby and David A. Petersen, assignors to Siemens Medical Solutions USA, Incorporated
14 April 2009 (Class 600/459); filed 26 February 2004

Different subarray sizes are used to form a building block. Larger arrays of elements are divided into rows that are associated with the subarrays. The rows in the larger arrays are combined into partial subarrays by using the subarray building blocks. This arrangement allows variation in subarray size with reduced switching requirements.—RCW

7,517,318

43.80.Vj REGISTRATION OF ELECTRO-ANATOMICAL MAP WITH PRE-ACQUIRED IMAGE USING ULTRASOUND

Andres Claudio Altmann and Assaf Govari, assignors to Biosense Webster, Incorporated
14 April 2009 (Class 600/459); filed 26 April 2005

A position sensor and an ultrasound imaging probe both in a catheter and at least one electrode are used with an already acquired ultrasound image to form an electrophysiologic map that is registered with the ultrasound image.—RCW

7,517,346

43.80.Vj RADIO FREQUENCY ABLATION SYSTEM WITH INTEGRATED ULTRASOUND IMAGING

Todd Sloan *et al.*, assignors to Boston Scientific Scimed, Incorporated
14 April 2009 (Class 606/41); filed 8 February 2005

Tissue in a region is ablated by using an ultrasound imaging probe aligned with the cannula of the ablation device to guide the ablation procedure.—RCW