CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 126, No. 4, Pt. 1 of 2 October 2009

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

On the sign of the adaptive passive fathometer impulse response (L)

James Traer,[a] Peter Gerstoft, H. C. Song, and William S. Hodgkiss
Scripps Institution of Oceanography, La Jolla, California 92093-0238

(Received 15 June 2009; revised 27 July 2009; accepted 28 July 2009)

Harrison [J. Acoust. Soc. Am. **125**, 3511–3513 (2009)] presented a mathematical explanation for a sign-inversion induced to the passive fathometer response by minimum variance distortionless response (MVDR) beamforming. Here a concise mathematical formulation is offered, which decomposes the cross-spectral density matrix into coherent and incoherent components and allows the matrix inversion to be obtained exactly by eigendecomposition. This shows that, in the region containing the bottom reflection, the MVDR fathometer response is identical to that obtained with conventional processing multiplied by a negative factor.
 [DOI: 10.1121/1.3206696]

PACS number(s): 43.30.Pc, 43.30.Re, 43.30.Wi, 43.30.Nb [AIT] Pages: 1657–1658

I. INTRODUCTION

A drifting vertical array can be used as a passive fathometer by cross-correlating up- and down-ward beams.[1] A more detailed analysis and an introduction to the application of adaptive beamforming to the passive fathometer are given in Ref. 2. Recently it was found that adaptive processing induced a sign change in the fathometer cross-correlation. An explanation was given in Ref. 3 using the Woodbury matrix identity to invert the cross-spectral density matrix (CSDM). Here we offer a simple explanation using eigendecomposition to perform this matrix inversion. This approach is exact given the idealized noise model described below.

A simple physical model is used consisting of a vertically downward propagating signal and its vertically upward propagating reflection in the presence of incoherent background noise. It should be noted that this is an idealization and a realistic ocean environment would likely contain a more complicated spatial structure. The passive fathometer requires the reflected signal to be coherent with the downward propagating signal to extract the depths of the reflecting layer boundaries. Consequently the two coherent signals are not separable and act as a single coherent component in constructing the CSDM. We show that, in the region of interest for the fathometer attributed to this coherent component, the adaptive response obtained using the minimum variance distortionless response (MVDR) beamformer is identical to that obtained using the conventional beamformer multiplied by a *negative* factor.

[a]Author to whom correspondence should be addressed. Electronic mail: jtraer@ucsd.edu

II. THEORY

Consider the response of the fathometer to a single frequency plane wave. The conventional response is given by[2]

$$C(\omega) = \mathbf{w}^T \mathbf{R} \mathbf{w}, \tag{1}$$

where $\mathbf{w}$ is the downward directed steering vector, the superscript T denotes the matrix transpose, and $\mathbf{R}$ is the CSDM. The MVDR adaptive response is given by[3]

$$C_{\mathrm{MVDR}}(\omega) = \Lambda(\omega) \mathbf{w}^T \mathbf{R}^{-1} \mathbf{w}, \tag{2}$$

where $\Lambda = |\mathbf{w}^H \mathbf{R}^{-1} \mathbf{w}|^{-2}$ is a positive normalization factor and the superscript H denotes the complex conjugate transpose. Thus, excluding the normalization factor, MVDR processing is of the same functional form as conventional processing with the CSDM inverse in place of the CSDM.

Decomposing the CSDM into coherent and incoherent components,

$$\mathbf{R} = \mathbf{d}\mathbf{d}^H + \sigma^2 \mathbf{I}, \tag{3}$$

where $\mathbf{d} = \mathbf{d}_{\mathrm{down}} + \mathbf{d}_{\mathrm{up}}$ is the sum of the downward propagating signal and the upward propagating reflection. $\mathbf{d}_{\mathrm{down}}$ is proportional to the steering vector $\mathbf{w}$. $\sigma^2 \mathbf{I}$ is the component due to incoherent background noise, with $\mathbf{I}$ designating the identity matrix. This allows the CSDM to be expanded by eigendecomposition to yield

$$\mathbf{R} = (\sigma^2 + a)\mathbf{u}_1 \mathbf{u}_1^H + \sigma^2 \sum_{j=2}^{N} \mathbf{u}_j \mathbf{u}_j^H, \tag{4}$$

where N is the number of array elements, a is the eigenvalue component corresponding to the coherent signal, and $\mathbf{u}_j$ are the normalized eigenvectors. The CSDM inverse is

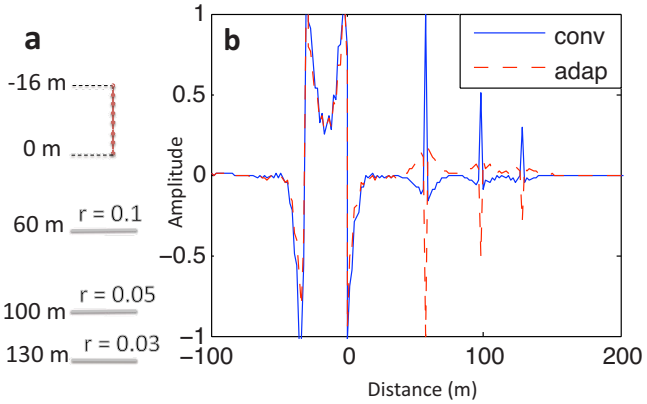


FIG. 1. (Color online) (a) A schematic of the model environment. The array is shown between −16 and 0 m (only every fourth hydrophone is shown) and the reflection layers with the associated reflection coefficients, r. (b) The conventional (solid) and MVDR (dashed) fathometer responses from simulated data.

$$\mathbf{R}^{-1} = \frac{1}{(\sigma^2 + a)}\mathbf{u}_1\mathbf{u}_1^H + \frac{1}{\sigma^2}\sum_{j=2}^{N}\mathbf{u}_j\mathbf{u}_j^H$$
$$= \frac{1}{\sigma^2}\left[\frac{\sigma^2}{(\sigma^2 + a)}\mathbf{u}_1\mathbf{u}_1^H + \mathbf{I} - \mathbf{u}_1\mathbf{u}_1^H\right]$$
$$= \frac{1}{\sigma^2}\left[\mathbf{I} - \frac{a}{\sigma^2 + a}\mathbf{u}_1\mathbf{u}_1^H\right]$$
$$= -\frac{1}{(\sigma^2 + a)^2}\mathbf{d}\mathbf{d}^H + \frac{1}{\sigma^2}\mathbf{I}, \tag{5}$$

where we have utilized $\Sigma\mathbf{u}_j\mathbf{u}_j^H = \mathbf{I}$. Thus the CSDM inverse contains the same terms as the CSDM with different coefficients. The component due to the coherent signal has been multiplied by a negative coefficient. The term due to background noise remains positive and the effect of this term is examined separately in both the frequency and time domains.

In the conventional case,

$$C_{\text{noise}} = \sigma^2\mathbf{w}^T\mathbf{I}\mathbf{w}. \tag{6}$$

The effect of this term depends on the choice of reference element. For convenience, the lowermost element of the array is defined as the reference [see Fig. 1(a)]. Thus the kth term of the steering vector is $e^{i\omega\tau_k}$, where τ_k is the vertical propagation time to the kth element. Substituting this yields

$$C_{\text{noise}} = \sigma^2\sum_{k=1}^{N} e^{2i\omega\tau_k}. \tag{7}$$

In the time domain, assuming infinite bandwidth, Eq. (7) corresponds to a series of delta functions at $t = -2[\tau_{N-1}, \tau_{N-2}, \ldots, \tau_2, 0]$. For an equispaced array with inter-element vertical travel time τ, this becomes $t = -2\tau[N-1, N-2, \ldots, 1, 0]$. The adaptive case is identical except the scalar coefficient Λ/σ^2 replaces σ^2. Note that defining the lowermost element as the reference ensures that these delta functions due to the background noise appear at negative times. Equations (1) and (2) are cross-correlations of down- and up-ward propagating beams. In this model, the upward beam is a sum of time-delayed reflections of the downward beam. As the reflections lag the downward beam, the correlation peaks will appear at positive times. Thus the seabed response will not be obscured by the contribution from background noise.

Neglecting the background noise component and substituting Eqs. (3) and (5) into Eqs. (1) and (2), respectively, reduces the fathometer response to

$$C(\omega) = \mathbf{w}^T\mathbf{d}\mathbf{d}^H\mathbf{w} \tag{8}$$

$$C_{\text{MVDR}}(\omega) = -\frac{\Lambda}{(\sigma^2 + a)^2}\mathbf{w}^T\mathbf{d}\mathbf{d}^H\mathbf{w}. \tag{9}$$

where $\Lambda/(\sigma^2+a)^2$ is a positive factor.

III. NUMERICAL SIMULATION

A simulation was constructed with a 32-element array with 0.5 m spacing (design frequency of 1500 Hz) over three reflection layers 60, 100, and 130 m below the bottom of the array with reflection coefficients of 0.1, 0.05, and 0.03, respectively [see Fig. 1(a)]. A boxcar function with a signal-to-noise ratio of −10 dB and a width of 1.3 ms was used as the down-going signal. The processing was done in the frequency domain with 1024 frequency bins from 0 to 750 Hz. A 40 Hz high pass filter was applied and the two responses were normalized such that the first reflection peak has an absolute magnitude of one.

The conventional and MVDR responses are shown in Fig. 1(b) against the depth associated with a two-way travel time in a medium with sound speed of 1500 m/s. At depths greater than 0, the conventional and MVDR traces are mirror images. In the region between −32 and 0 m, the conventional and MVDR responses are similar. This is the region dominated by the incoherent noise term. The delta functions predicted in Eq. (7) are twice convolved with the box car signal which obscures the individual peaks.

IV. SUMMARY

This analysis shows that adaptive processing will induce a negative sign to the seabed response given by conventional passive fathometer processing. In addition, it has been shown that the component from incoherent noise which obscures both conventional and adaptive processing can be confined to negative times by referencing the array elements relative to the bottom hydrophone.

[1]M. Siderius, C. H. Harrison, and M. B. Porter, "A passive fathometer technique for imaging seabed layering using ambient noise," J. Acoust. Soc. Am. **120**, 1315–1323 (2006).
[2]P. Gerstoft, W. S. Hodgkiss, M. Siderius, C. F. Huang, and C. H. Harrison, "Passive fathometer processing," J. Acoust. Soc. Am. **123**, 1297–1305 (2008).
[3]C. H. Harrison, "Anomalous signed passive fathometer impulse response when using adaptive beam forming," J. Acoust. Soc. Am. **125**, 3511–3513 (2009).

Time domain visualization using acoustic holography implemented by temporal and spatial complex envelope (L)

Choon-Su Park[a] and Yang-Hann Kim
Department of Mechanical Engineering, Center for Noise and Vibration Control (NOVIC), Korea Advanced Institute of Science and Technology (KAIST), 373-1 Guseong-dong, Yuseong-gu, Daejeon 305–701, Korea

(Received 2 April 2009; revised 15 July 2009; accepted 16 July 2009)

Spatial envelope was proposed to show the location of acoustic sources and overall radiation pattern by the authors [C.-S. Park and Y.-H. Kim, J. Acoust. Soc. Am. **125**, 206–211 (2009)]. The envelope can provide sufficient information on where the sources are and how the energy propagates into space. This concept is certainly useful for time domain acoustic holography since one can utilize not only spatial envelope to envisage what one needs to know but also temporal envelope to reduce the number of data. A holographic process to obtain spatial envelope is therefore introduced and verified, and how much one can reduce the processing time by implementing envelopes is compared with the conventional holography. © *2009 Acoustical Society of America.*
[DOI: 10.1121/1.3203919]

PACS number(s): 43.60.Sx, 43.60.Ac, 43.60.Jn, 43.60.Gk [EJS] Pages: 1659–1662

I. INTRODUCTION

Acoustic holography is a method to reconstruct sound fields in prediction plane using acoustic pressure measured in hologram plane (for instance, see Ref. 1). Time domain sound visualization using acoustic holography has been used particularly to show time-variant sound fields, for example, transient sound field radiated from automobile tire[2] or non-stationary sound field from engine during run-up test.[3] Various papers also have been published on the sound visualization in time domain.[4–7] The time domain sound visualization can exhibit the time-variant sound fields by convolving time data or taking Fourier transform on time and space, which is well summarized by Rochefoucauld *et al.*[7]

The time domain sound visualization is certainly useful in displaying sound fields as fast as possible and also in analyzing the results when it depicts only what we need to know (e.g., the locations of sources, etc.). The time domain visualization, however, usually takes a great deal of data to reconstruct the sound fields. In addition, the sound fields in the prediction plane often contain much more information than we want to know. For example, if we want to observe information from the visualized sound field for dealing with noise, where noise sources are and how the sound radiate would be the very information we want to know. In such a case, the whole variation in the sound field is often sufficient rather than the reconstructed sound field itself. The authors, therefore, raised these problems associated with processing time issue and analysis problem[8] in the previous paper,[9] and spatial envelope was proposed as a solution method for the analysis problem of sound visualization.

We propose a means to implement the complex envelope in time and space for visualizing the information where sound sources are and how the sound energy propagates into space with shorter calculation time than the conventional holography. The temporal envelope is used as a way for reducing calculation time by using less number of data and the spatial envelope[9] for showing the information we want to see. To obtain temporal and spatial envelopes, modulations in time and space are needed. The modulation of time data should be processed before sampling by analog modulation device to obtain reduced number of data, but the spatial modulation can be completely processed by data processing in the wavenumber domain as the propagation process of acoustic holography. Two-dimensional Fourier transform of the sound field gives wavenumber spectra for the sound field, and the spatial modulation is applied for each wavenumber (spatial frequency) in the band of interest. Then, spatial envelope for the given sound field is obtained by inverse Fourier transform of both space and time.

The proposed method can be applied to sound visualization devices overlaid with camera image for recognizing sound field easily, for example, devices for visualizing sound field which was induced by malfunctions of machinery for diagnosis or visual assist devices for the deaf. The users want to see not spatial fluctuation of sound field in detail, but simple sound field that can clearly give where and how loud the sound sources are for preparing possible dangers.

II. COMPLEX ENVELOPE IN TIME DOMAIN AND SPACE DOMAIN

A. Time domain complex envelope

The envelope of a narrow band signal shows overall change of the signal; for instance, see the fluctuating thick solid line along t-axis as depicted in Fig. 1. The envelope does not depend on where it is placed in frequency domain, as shown in Fig. 1, because an envelope is only associated with the bandwidth and its components within the band (for instance, see Ref. 10). This means that the envelope does not change by shifting (modulating in time domain) the narrow

[a]Author to whom correspondence should be addressed. Electronic mail: cs_park@kaist.ac.kr

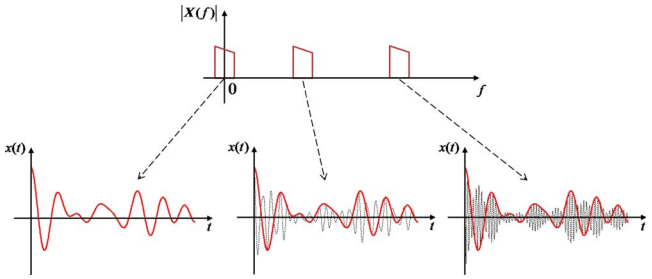


FIG. 1. (Color online) A narrow band signal (thin dotted line) and its complex envelope (thick solid line) in time with respect to its position in frequency domain. t and f represent time and frequency, respectively. $x(t)$ is the narrow band signal, and $|X(f)|$ is the magnitude of the frequency spectrum of $x(t)$.

band spectrum. This enables us to obtain the signal of interest with lower sampling rate. Low sampling rate allows us to acquire less data when we measure the same record length.

Complex modulation,[11] shifting the band signal into low frequency region and then low pass filtering the signal, provides us with a complex envelope of the signal. The modulation has to be realized with analog signal, and this can be realized by the quadrature modulation method.[12] In short, by introducing the concept of temporal envelope and using analog modulation circuit, we can make use of less data for time domain sound visualization.

B. Space domain complex envelope

Although the spatial envelope is an extension of the temporal envelope into the space domain, the spatial envelope has its own distinct feature distinguished from the temporal one: the geometric information.[9] The information inherently exists in the radiated sound field. The geometric information, of course, smeared out as the sound propagates owing to evanescent wave components which constitute the geometric information. Therefore, it is assumed that sound fields are measured in near field with respect to the wavelength of interest. Then, although the sound field varies complicatedly, the spatial envelope can show not only overall distribution of the sound field but also sources' geometric information. When we use the sound visualization as a way of providing information of noise sources, it is of importance to find out where the noise sources and how the energy transmits to space. The spatial envelope can envisage energy propagation of the acoustic pressure distribution and the geometric information of sources as well.

In addition, it is assured that the spatial envelope is acoustic pressure distribution. Therefore, other acoustic variables (particle velocity, intensity distribution, etc.) are obtainable from the spatial envelope, which is discriminated from other sound source localization methods that only estimate the location of sources (e.g., beamforming method).

III. MODIFIED HOLOGRAPHIC PROCESS TO OBTAIN A SPATIAL ENVELOPE

The spatial envelope of a sound field is obtained by the spatial modulation.[9] Therefore, to obtain the envelope in the prediction plane, the spatial modulation process should be added to the conventional holographic procedure, as shown

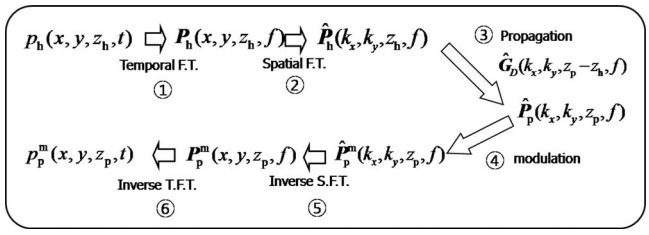


FIG. 2. The modified holographic process to get spatial envelope. $p_h(x,y,z_h,t)$ is a pressure distribution on the hologram plane at specified z_h, $p_p(x,y,z_p,t)$ is a pressure distribution on the prediction plane at specified z_p, $\hat{\boldsymbol{G}}_D(k_x,k_y,z_p-z_h,f)$ is a propagator satisfying Dirichlet boundary condition, the superscript m indicates that the variables are modulated, the bold-faced font for complex value and capital letters are for the variables in frequency domain, and a hat ($\hat{}$) expresses the value at wavenumber domain.

in Fig. 2. To develop and verify this process mathematically, let us consider the simplest case, an acoustic monopole source in the free field. A sound field radiated from the source radiating a center frequency f_c is placed at $(x_s,y_s,0)$, and the source has a slowly varying amplitude, which is composed of much lower frequencies $[\Delta f_n, (n=1,2,\ldots,N)]$ compared to the center frequency $(\Delta f_n \ll f_c)$,

$$\boldsymbol{a}(r,t)=\sum_{n=1}^{N} c_n e^{-j(2\pi\Delta f_n t-\Delta k_n R)}, \tag{1}$$

where c_n is the coefficient of each frequency component and $R=\sqrt{(x-x_s)^2+(y-y_s)^2+z^2}$. $\Delta k_n(=2\pi\Delta f_n/c)$ is corresponding wavenumber deviation. That is, the wavenumber (k_c) of center frequency plus the deviation (Δk_n) indicates the wavenumbers (k_n) of the narrow band signal. Then, the acoustic pressure at an arbitrary z plane can be formulated as

$$\boldsymbol{p}_h(x,y,z,t)=-\frac{j\rho_0 cq}{4\pi}\sum_{n=1}^{N} c_n k_n \frac{e^{-j(2\pi f_n t+k_n R)}}{R}, \tag{2}$$

where $\rho_0 c$ is the characteristic impedance of the media and q is the volume velocity. The complex envelope of Eq. (2) is given by

$$\boldsymbol{p}_{\mathrm{CE}}(x,y,z,t)=-\frac{j\rho_0 cq}{4\pi}\sum_{n=1}^{N} c_n k_n \frac{e^{-j(2\pi\Delta f_n t+\Delta k_n R)}}{R}, \tag{3}$$

without the terms associated with f_c.[9]

Then, let us follow the holographic procedure. The pressure in a hologram plane z_h is

$$\boldsymbol{p}_h(x,y,z_h,t)=-\frac{j\rho_0 cq}{4\pi}\sum_{n=1}^{N} c_n k_n \frac{e^{-j(2\pi f_n t+k_n R_h)}}{R_h}, \tag{4}$$

where $R_h=\sqrt{(x-x_s)^2+(y-y_s)^2+z_h^2}$. The wavenumber spectra for each frequency are obtained by spatial Fourier transform as (see Ref. 9 in detail),

$$\hat{\boldsymbol{P}}_h(k_x,k_y,z_h)=\rho_0 cqc_n k_n \frac{e^{jz_h\sqrt{k_n^2-k_x^2-k_y^2}}}{\sqrt{k_n^2-k_x^2-k_y^2}}, \tag{5}$$

which are obtained with respect to N numbers of frequency, and hat ($\hat{}$) means the variables are in the wavenumber domain. To obtain the pressure in the prediction plane, z_p, the propagator which satisfy the Dirichlet boundary condition[1] is

used. Then, the wavenumber spectra in the prediction plane are obtained. After obtaining the wavenumber spectra, the spatial modulation is applied to each wavenumber spectrum. Then, the modulated wavenumber spectra are obtained as

$$\hat{\boldsymbol{P}}_p^m(k_x,k_y,z_p)=\rho_0 cqc_n k_n \frac{e^{jz_p\sqrt{\Delta k_n^2-k_x^2-k_y^2}}}{\sqrt{\Delta k_n^2-k_x^2-k_y^2}} \tag{6}$$

by the spatial modulation method.[9] Eventually, the spatial envelope in the prediction plane is obtained by inverse spatial and temporal Fourier transform. That is,

$$\boldsymbol{p}_p^m(x,y,z_h,t)=-\frac{j\rho_0 cu_0 q}{4\pi}\sum_{n=1}^{N} c_n k_n \frac{e^{-j(2\pi\Delta f_n t-\Delta k_n R_p)}}{R_p}, \tag{7}$$

where $R_p=\sqrt{(x-x_s)^2+(y-y_s)^2+z_p^2}$. Equation (7) shows that the envelope in the prediction plane is obtained correctly, which is the same with the complex envelope of Eq. (3).

As a more general case, sound fields governed by Rayleigh's first integral formula,

$$\boldsymbol{P}(x,y,z)=-\frac{j\rho_0 ck}{2\pi}\int_{y_s}\int_{x_s}\frac{e^{jk\sqrt{(x-x_s)^2+(y-y_s)^2+z^2}}}{\sqrt{(x-x_s)^2+(y-y_s)^2+z^2}} \times u(x_s,y_s,0)dx_s dy_s, \tag{8}$$

are also available for the above process because the integral formula tells us that the sound from the velocity source at each source point propagates like monopole (e^{jkR}/R). The other Rayleigh formula (for instance, see Ref. 13) describing sound field radiated from acoustic pressure sources has different type propagator, namely, dipole. This propagator requires different modulator proposed in Ref. 9 because the wavenumber spectra are dissimilar from the modulator for monopole type. The wavenumber spectrum and corresponding modulator are given in the Appendix.

IV. THE ROLE OF COMPLEX ENVELOPES IN HOLOGRAPHIC PROCESS

A. Processing time reduction by using temporal envelope

In spite of the additional process (step 4 in Fig. 2), we can reduce calculation time for the whole process because of less data by temporal modulation. Although we stated processing time issue, the processing time is not adequate as a parameter to estimate how the computation is fast because the processing time can be varied according to hardware configurations (for instance, CPU or data transfer rate, etc.). The number of multiplication therefore can be an appropriate parameter to estimate how fast the process is because all the processes of holography are made up of multiplication, and the number of multiplication is fixed with respect to the fast Fourier transform algorithm used. For example, for "radix-2" algorithm, the number of multiplication for N numbers of data is $(N/2)\log_2 N$(e.g., see Ref. 14).

Then, let us assess how many times the number of multiplication needed for the modified process compared with the conventional one in case of the radix-2 algorithm. To compare the modified procedure with the conventional one, let us have $N_x \times N_y$ measurement points in space and for simplicity, $N_x=N_y=N_s$ (i.e., square measurement). In addition, we assumed to sample N_t points in time domain, and the number of the band of interests is $N_B=N_t/B$. B represents the ratio between initial number of data and the reduced number of data. Then the number of multiplication is summarized in Table I. For steps 1 and 6 and 2 and 5, the number of multiplication should be the same since they are Fourier transform pair. In steps 2 and 3, the constant "2" is for two-dimensional Fourier transform, and therefore $2 \times (N_s/2\log_2 N_s)\times(N_s/2)$ number of multiplication is needed with respect to N_B for the modified process and $N_t/2$ for the conventional process. In addition, the propagation and modulation should be calculated whole spatial points with respect to each frequency of interest. Therefore, the number of multiplication is $N_s^2 \times N_B$ for the modified process and $N_s^2 \times (N_t/2)$ for the conventional process. Finally, the ratio between the total number of multiplication (TNM) of the modified and conventional procedure is obtained as

TABLE I. The number of multiplication of the conventional and modified holographic procedures.

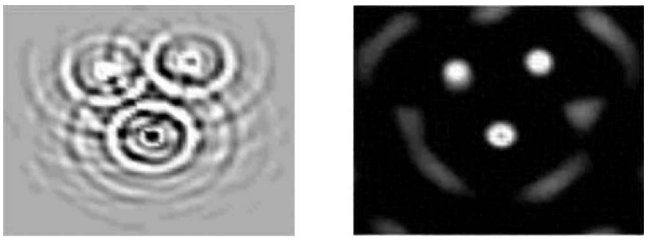
	Process	Modified process	Conventional process
1	Temporal Fourier transform	$\left(\frac{N_B}{2}\log_2 N_B\right)\times N_s^2$	$\left(\frac{N_t}{2}\log_2 N_t\right)\times N_s^2$
2	Spatial Fourier transform	$2\times\left(\frac{N_s}{2}\log_2 N_s\right)\times N_B$	$2\times\left(\frac{N_s}{2}\log_2 N_s\right)\times\frac{N_t}{2}$
3	Propagation	$N_s^2\times N_B$	$N_s^2\times\frac{N_t}{2}$
4	Modulation	$N_s^2\times N_B$	·
5	Inverse SFT	$2\times\left(\frac{N_s}{2}\log_2 N_s\right)\times N_B$	$2\times\left(\frac{N_s}{2}\log_2 N_s\right)\times\frac{N_t}{2}$
6	Inverse TFT	$\left(\frac{N_B}{2}\log_2 N_B\right)\times N_s^2$	$\left(\frac{N_t}{2}\log_2 N_t\right)\times N_s^2$

$$\frac{\mathrm{TNM}_m}{\mathrm{TNM}_c}=\frac{1}{B}\frac{2\log_2 N_t+\frac{4}{N_s}\log_2 N_s+2(2-\log_2 B)}{2\log_2 N_t+\frac{4}{N_s}\log_2 N_s+1}, \tag{9}$$

where the subscript m means modified procedure and c indicates the conventional one. As you can see in Eq. (9), the ratio is dominantly proportional to $1/B$ because another term composed of logarithmic values in the right-hand side has around 1. This result is reasonable since we handle the proposed process with less data with approximately $1/B$ during the procedure.

B. Spatial complex envelope for analysis problem

The spatial envelope is considered as a solution method for the analysis problem of acoustic holography. In other words, it is needed for the reconstructed sound field to provide correct location or distribution of sources because sound visualization for engineering problem is usually used to display where sound sources are and how they radiate. The complex envelope of the reconstructed sound field can there-



(a) (b)

FIG. 3. An acoustic pressure distribution and its spatial complex envelope in prediction plane obtained through modified holographic procedure. (a) An acoustic pressure distribution generated by three monopoles. (b) Spatial envelope of the pressure distribution in prediction plane.

fore be a possible candidate to show not only the geometric information of sources but also overall radiation pattern associated with energy propagation in space.

Figure 3 shows the reconstructed pressure field composed of three monopoles and its spatial complex envelope in prediction plane. Figure 3(a) shows the result of the conventional holography, and Fig. 3(b) is the spatial envelope of the reconstructed sound field. Comparing the results, spatial envelope shows the locations of the sources more clearly, and the radiated pattern is less complicated than the pressure field reconstructed by the conventional holography process.

V. CONCLUSIONS

The authors reviewed the temporal and spatial complex envelopes as a solution method to resolve the analysis problem and processing time issue for time domain sound visualization by acoustic holography. The temporal envelopes contribute to reduce the number of data. Therefore, the processing time to get results in prediction plane can be decreased by $1/B$ compared with the conventional process. In addition, by using the spatial envelope, the reconstructed pictures provide us the information we need: the geometric information and overall radiation pattern. We also examined the modified holographic procedure, and it should not be missed that the modulation must be applied after the propagation process to prevent elimination of the propagating waves.

ACKNOWLEDGMENTS

This study was partly supported by Agency for Defense Development (ADD), Korea Atomic Energy Research Institute (KAERI), and also BK21 (Brain Korea 21) project initiated by Ministry of Education, Science and Technology of Korea.

APPENDIX: A WAVENUMBER SPECTRUM OF A DIPOLE SOURCE AND ITS SPATIAL MODULATOR

A dipole source is composed of two adjacent monopole sources. The wavenumber spectrum of a dipole is obtained differently according to its direction of polarization, which is the inclined direction between the two monopoles. The wavenumber spectrum of a dipole source at an origin is derived in Ref. 15. That is,

$$\hat{\boldsymbol{P}}_{\mathrm{di}}(k_x,k_y,z) = 2\pi(n_x k_x + n_y k_y + n_z k_z)\frac{e^{-jz\sqrt{k^2-k_x^2-k_y^2}}}{\sqrt{k^2-k_x^2-k_y^2}}, \quad \text{(A1)}$$

where n_x, n_y, and n_z denote the directional cosine of the inclined direction $(n_x^2+n_y^2+n_z^2=1)$. For example, if the direction of polarization is on x-axis, then the spectrum is

$$\hat{\boldsymbol{P}}_{\mathrm{di},x}(k_x,k_y,z) = 2\pi k_x \frac{e^{-jz\sqrt{k^2-k_x^2-k_y^2}}}{\sqrt{k^2-k_x^2-k_y^2}}. \quad \text{(A2)}$$

However, if the direction is on z-axis, the spectrum is

$$\hat{\boldsymbol{P}}_{\mathrm{di},z}(k_x,k_y,z) = 2\pi k_z \frac{e^{-jz\sqrt{k^2-k_x^2-k_y^2}}}{\sqrt{k^2-k_x^2-k_y^2}} = 2\pi e^{-jz\sqrt{k^2-k_x^2-k_y^2}}. \quad \text{(A3)}$$

This concludes that the authors can modulate the wavenumber spectra of dipole sources when the direction of polarization is without regard to the z-direction, but they have to use another modulator for z-directional polarization. That is,

$$\boldsymbol{M}_{k_1\to k_2}(k_x,k_y,z) = e^{jz\{\sqrt{k_2^2-k_x^2-k_y^2}-\sqrt{k_1^2-k_x^2-k_y^2}\}} \quad \text{(A4)}$$

when the authors want to change the wavenumber from k_1 to k_2. The Rayleigh's second integral formula uses this type of propagator. Therefore, the modulator of (A4) should be used for the case.

[1] 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).
[2] E.-U. Saemann and J. Hald, "Transient tyre noise measurements using time domain holography," in Proceedings of SAE (1997), pp. 1423–1429.
[3] J. Hald, "Time domain acoustical holography and its applications," Sound Vib. **35**, 16–25 (2001).
[4] F. Deblauwe, J. Leuridan, J. L. Chauray, and B. Beguet, "Acoustic holography in transient conditions," in Proceedings of the Sixth International Congress on Sound and Vibration (1999), pp. 899–906.
[5] V. Grulier, J.-H. Thomas, J.-C. Pascal, and J.-C. Le Roux, "Time varying forward projection using wavenumber formulation," in Proceedings of Inter-Noise 2004, Prague, Czech Republic (2004), Abstract 406.
[6] M. Forbes, S. Letcher, and P. Stepanishen, "A wave vector, time-domain method of forward projecting time-dependent pressure fields," J. Acoust. Soc. Am. **90**, 2782–2793 (1991).
[7] O. de La Rochefoucauld, M. Melon, and A. Garcia, "Time domain holography: Forward projection of simulated and measure sound pressure fields," J. Acoust. Soc. Am. **116**, 142–153 (2004).
[8] Y.-H. Kim, "Can we hear the shape of a noise source?" in Plenary Lecture, The 18th International Congress on Acoustic, Kyoto International Conference Hall, Japan, (2004), pp. 3357–3370.
[9] C.-S. Park and Y.-H. Kim, "Space domain complex envelopes: Definition and a spatial modulation method," J. Acoust. Soc. Am. **125**, 206–211 (2009).
[10] S. J. Roome, "Analysis of quadrature detectors using complex envelope notation," IEE Proc. F, Radar Signal Process **136**, 95–100 (1989).
[11] D. G. Childers and M. Pao, "Complex demodulation for transient wavelet detection and extraction," IEEE Trans. Audio Electroacoust. **20**, 295–308 (1972).
[12] O. D. Grace and S. P. Pitt, "Sampling and interpolation of band-limited signals by quadrature methods," J. Acoust. Soc. Am. **48**, 1311–1318 (1970).
[13] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, San Diego, 1999), pp. 31–38.
[14] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. (Prentice-Hall, Englewood Cliffs, NJ, 1989), Chap. 9.
[15] H.-S. Kwon, "Sound visualization by using enhanced planar acoustic holographic reconstruction," Ph.D. thesis, KAIST (1997), Chap. 2 and Appendix.

Octave-shifted pitch matching in nonword imitations: The effects of lexical stress and speech sound disorder (L)[a)]

Beate Peter,[b)] Tara Larkin, and Carol Stoel-Gammon
Department of Speech and Hearing Sciences, University of Washington, P.O. Box 354875, Seattle, Washington 98105

(Received 2 April 2009; revised 10 July 2009; accepted 21 July 2009)

Perceptual similarities of musical tones separated by octave intervals are known as octave equivalence (OE). Peter *et al.* [(2008). Proceedings of the Fourth Conference on Speech Prosody, edited S. Maduerira, C. Reis, and P. Barbosa, Luso-Brazilian Association of Speech Sciences, Campinas, pp. 731–734] found evidence of octave-shifted pitch matching (OSPM) in children during verbal imitation tasks, implying OE in speech tokens. This study evaluated the role of lexical stress and speech sound disorder (SSD) in OSPM. Eleven children with SSD and 11 controls imitated low-pitched nonwords. Stimulus/response f_0 ratios were computed. OSPM was expressed preferentially in stressed vowels. SSD was associated with reduced expression of OSPM in unstressed vowels only. Results are consistent with the psycholinguistic prominence of lexical stress and prosodic deficits in SSD. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203993]

PACS number(s): 43.66.Hg, 43.70.Fq, 43.70.Dn, 43.66.Lj [DD] Pages: 1663–1666

I. INTRODUCTION

The perceptual similarity between two musical tones separated by one or more octave intervals is referred to as octave equivalence (OE). Musical tuning systems in different cultures around the world vary widely in the way scales are arranged, yet all utilize octave-based perceptual similarities (Burns and Ward, 1978). Psychophysical evidence for OE has been found in a small number of studies in adults and infants, and even animals, although results have been somewhat controversial (Burns, 1999).

The biological basis for perceptual OE in musical tones is unclear. Several theoretical approaches (reviewed in Burns, 1999) have attempted to account for OE, including the place code theory, which posits that octave-separated tones activate the same place on the basilar membrane in the cochlea because they share harmonic frequencies. Langner (2007) pointed out that neurons in the inferior colliculus respond not only to characteristic frequencies but also to their integer multiples.

While OE has been documented in the music literature, it is unknown whether it exists in speech signals as well. Evidence of pitch matching at an octave interval in a spoken imitation task would imply the presence of perceptual similarity in the speech tokens. To evaluate pitch matching in imitation tasks, the vocal pitch levels in the target speaker and the respondent must be considered because these vary by age and gender. In 4- and 5-year-old children, mean fundamental frequency (f_0) levels were estimated at 286 and 289 Hz, respectively (Eguchi and Hirsh, 1969). Typical f_0 mean values for male and female adults between 25 and 50 years range between 100 and 120 Hz and between 240 and 245 Hz, respectively; with advancing age, f_0 tends to rise in men and to drop slightly in women (Hollien and Shipp, 1972; Mysak and Hanley, 1959; Saxman and Burk, 1967).

In conversational speech, f_0 adjustments have been described as a function of discourse mechanics, for instance, in the form of pitch matching between speakers to convey agreement (Brazil *et al.*, 1985; Wennerstrom, 2001). Pitch is usually described relative to a speaker's vocal range (e.g., high, mid, or low key), not physical scales. In the context of an imitation task, Peter *et al.* (2008) unexpectedly found evidence of octave-shifted pitch matching (OSPM). They analyzed pitch in a study designed originally to evaluate the role of temporal aspects of speech and hand tasks in children (Peter and Stoel-Gammon, 2008). In two imitation tasks, nonwords and sentences, where the target voice was an adult male's, the participants tended to imitate the tokens one octave above the target level.

The purpose of the present study was to evaluate the role of lexical stress and speech sound disorder (SSD) in the expression of OSPM in a nonword imitation task. Stressed syllables appear to be associated with psycholinguistic prominence. Newborn infants demonstrate an awareness of language-specific stress patterns (reviewed in Echols *et al.*, 1997) and the ability to use stressed syllables as cues to segment words (Jusczyk *et al.*, 1993). Young children tend to omit unstressed syllables more frequently than stressed syllables (Ingram, 1974; Kehoe and Stoel-Gammon, 1997). In individuals who stutter, stressed syllables are affected predominantly (Hahn, 1942). Some children with SSD have difficulty speaking with appropriate prosody, which includes not only temporal aspects such as vowel durations and intensity levels but also aspects related to vocal pitch (Shriberg *et al.*, 2003). Lower expression of OSPM is therefore predicted in unstressed syllables and in participants with SSD.

[a)] Aspects of this study were presented at the Fourth Conference on Speech Prosody, Campinas, Brazil, May 6–9, 2008 and at the Child Phonology Conference, Purdue University, June 2–3, 2008.

[b)] Author to whom correspondence should be addressed. Also at Department of Medicine, Division of Medical Genetics, University of Washington, P.O. Box 357720, Seattle, WA 98195. Electronic mail: bvpeter@u.washington.edu

TABLE I. Schematics of stimuli in the nonword imitation task. Note: Capital letters indicate stressed syllables, lower-case letters indicate unstressed syllables, and periods mark syllable boundaries. Items 13 and 14 contain pauses and were not administered.

Item no.	Rhythm pattern
1	MA.ma
2	ma.MA
3	MA.ma.MA
4	ma.MA.ma
5	MA.ma.ma
6	MA.ma.MA.ma
7	ma.MA.MA.ma
8	ma.ma.MA.ma
9	ma.MA.ma.ma
10	ma.MA.ma.MA
11	MA.ma.ma.MA.ma.ma
12	ma.MA.ma.ma.MA.ma
13	MA.[pause]MA.ma
14	MA.ma.[pause]MA

II. PARTICIPANTS, PROCEDURES, AND RELIABILITY

This study was conducted with the approval of the University of Washington Human Subjects Division and in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Eleven children, ages 4 years and 7 months to 6 years and 6 months, with a diagnosis of a moderate to severe SSD of unknown origin and 11 age- and gender-matched controls with typical development (TD) participated in this study. All participants were required to fulfill the following criteria: age 4 years and 6 months to 7 years; monolingual English home environment; hearing screen passed at 25 dB sound pressure level (SPL); no history of any developmental or acquired disorder interfering with hearing, speech fluency, or neuropsychologic functioning. The participants with SSD additionally were required to have a diagnosis of moderate to severe speech disorder of unknown origin, documented with standardized speech test (articulation and/or phonology) scores of 1.5 standard deviations below the mean or lower.

The imitation task was based on the "Tennessee test of rhythm and intonation patterns" (T-TRIP; Koike and Asp, 1981), a nonstandardized test of prosody. During the Rhythm subtest, the participant hears 14 prerecorded sequences of the syllable "ma" and imitates them. Items 13 and 14 were excluded from analysis because they contain pauses. Items 1–12 contain a total of 45 syllables, of which 43, 18 stressed and 25 unstressed, were available for analysis; a reliable pitch measurement could not be obtained for two of the modeled syllables. The voice on the CD provided by the manufacturer was a male adult, with f_0 in the stressed vowels averaging 133 Hz and, in the unstressed vowels, 105 Hz (overall average f_0=116 Hz). Table I shows the metric structures of the test items.

Customary f_0 measures were measured in 5–12 utterances obtained from conversational speech samples collected during the study sessions. Acoustic analyses were performed by a team consisting of the first author and seven undergraduate or postbaccalaureate students in the Department of Speech and Hearing Sciences at the University of Washington using PRAAT, Version 4.2.09 (Boersma and Weenink, 2004) or more recent versions, measuring f_0 in Hz.

The octave equivalence range (OER) was defined as any pitch that fell within one semitone above or below twice the stimulus frequency. As the adult targets were, on average 116 Hz, the average OER was 220–247 Hz. For individual vowel imitations, a child/adult f_0 ratio was calculated. Ratios between 1.887:1 and 2.119:1 represented the range within one semitone of twice the adult target's f_0 level, following the formula

$$\mathrm{OER} = 2(f_0 2^{-1/12}), 2(f_0 2^{1/12}). \quad (1)$$

Statistical analyses included paired t tests, t tests with adjustments for unequal variances, and the nonparametric Wilcoxon signed-rank test. The acoustic measurements were checked for reliability by remeasuring approximately 15% of the tokens. The average discrepancy was 5.36 Hz for vowels and 5.05 Hz for the conversational samples. This was considered acceptable for the purposes of this study.

III. RESULTS

As a group, the participants with SSD produced an average of 35.6 syllables with correctly placed stress out of the 43 available model syllables (N=368). The participants with TD produced an average of 40.7 syllables with correctly placed stress (N=422). The participants with SSD had an average conversational f_0 of 285 Hz [standard deviation (SD)=42 Hz], and the participants with TD averaged 292 Hz (SD=43 Hz), which closely approximates conversational values expected for the participants' ages. Table II summarizes participant characteristics and f_0 measures.

107 of 421 (25%) unstressed and 160 of 368 (43%) stressed imitated vowels fell within the OER. Differences were statistically significant (paired t test using per-participant averages: t=9.11, $p<0.0001$; t test with adjustments for unequal variances using all vowel measurements: t=14.16, $p<0.0001$).

The SSD group included two participants whose average conversational f_0 levels fell within the defined OER. Of the remaining nine, whose conversational f_0 levels all fell above that range, one participant downshifted the conversational f_0 levels to the OER in the imitations; four participants lowered their conversational f_0 levels by at least 32 Hz without reaching the OER, and four did not adjust their conversational f_0 levels appreciably during the imitation task (difference $\leq$8 Hz). The TD group included one participant with both conversational and imitated f_0 levels within the OER, five participants who downshifted their conversational f_0 levels to fall within the OER during the imitations, four who lowered their conversational f_0 levels substantially during the imitations without reaching the OER, and one participant who did not adjust the conversational f_0 levels appreciably. Figure 1 shows conversational and imitated f_0 values for each participant, in relationship to the adult target and the OER.

124 of 368 (34%) vowels in the SSD group and 143 of 422 (34%) in the TD group fell into the OER. Nonparametric

TABLE II. Summary of participant gender, age, mean conversational and imitated f_0 levels, and child/adult f_0 ratios. Child/adult f_0 ratios within one semitone of twice the stimulus f_0 levels (range: 1.888, 2.119) are indicated with a *. [Table adapted with permission from Peter and Stoel-Gammon (2008).]

Code	Gender	Age in months	Conversational f_0	Imitated f_0	Mean C/A f_0 ratio	C/A f_0 ratio (unstressed syllables)	C/A f_0 ratio (stressed syllables)
D01	M	73	222	224	*1.933	*2.103	1.744
D02	M	60	288	248	*2.095	2.246	*1.961
D03	M	72	256	259	2.260	2.531	*1.930
D04	M	67	269	268	2.323	2.681	*1.988
D05	M	78	238	254	2.133	2.266	*1.990
D06	M	61	318	286	2.453	2.712	*2.116
D07	F	67	283	223	*1.929	*2.000	1.842
D08	M	55	313	268	2.224	2.421	*2.062
D09	F	59	355	262	2.251	2.398	*2.081
D10	F	59	331	328	2.806	2.993	2.596
D11	M	76	258	251	2.157	2.340	*1.934
T01	M	74	283	265	2.230	2.244	2.216
T02	M	61	312	236	*2.040	2.171	1.867
T03	M	72	325	249	2.164	2.369	*1.895
T04	M	63	279	223	*1.915	*1.956	1.869
T05	M	81	265	235	*2.054	2.216	1.839
T06	M	61	303	238	*2.018	2.108	*1.918
T07	F	73	229	232	*1.961	*2.064	1.857
T08	M	58	330	282	2.375	2.480	2.245
T09	F	58	353	283	2.391	2.650	2.170
T10	F	59	316	233	*2.002	2.131	*1.877
T11	M	76	214	220	*1.925	*1.988	1.838

testing for paired data, inputting the per-participant mean scores, did not show statistically significant SSD-TD group differences ($z=1.16$, $p=0.2477$). Parametric testing based on t testing for matched data, inputting the child/adult f_0 ratios from all analyzed vowels, showed a significant group difference ($t=6.24$, $p<0.0001$). Both participant groups deviated from the 2:1 ratio in both directions, but the distributions were skewed toward overshoots in the direction of the participants' conversational f_0 levels. The distribution in the SSD group showed a wider range and greater variability compared to the TD group.

Because OSPM was more robust in the stressed than in the unstressed vowels, the two groups were compared using data from the two stress types separately. For the unstressed vowels, the SSD group produced 38 of the 193 vowels (20%) within the OER, whereas the TD group produced 69 of the 228 vowels (30%) in the OER. For the stressed vowels, the SSD group produced 86 of 174 vowels (49%) within the OER, whereas the TD group produced 74 of the 194 vowels (38%) in the OER. Nonparametric testing failed to detect statistically significant differences between the SSD and TD groups in the unstressed vowels ($z=1.43$, $p=0.1549$) and in the stressed vowels ($z=0.98$, $p=0.3281$). Parametric testing showed a statistically significant difference between the SSD and TD groups for both the unstressed ($t=6.38$, $p<0.0001$)

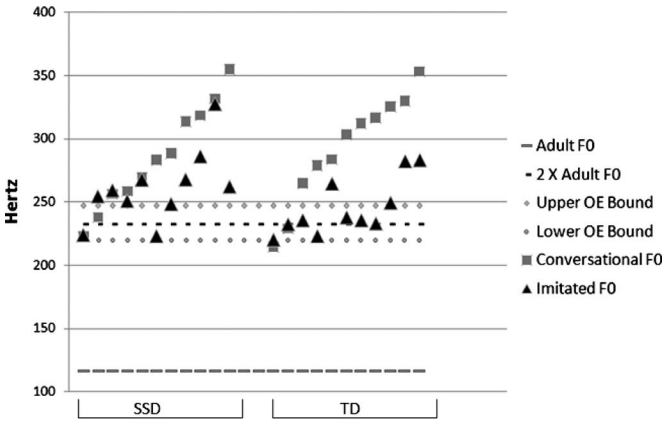


FIG. 1. Target adult f_0, double target adult f_0, upper and lower bounds of the OER, and mean conversational and imitated f_0 for each participant. Data points shown separately for each participant group, sorted by conversational f_0 level.

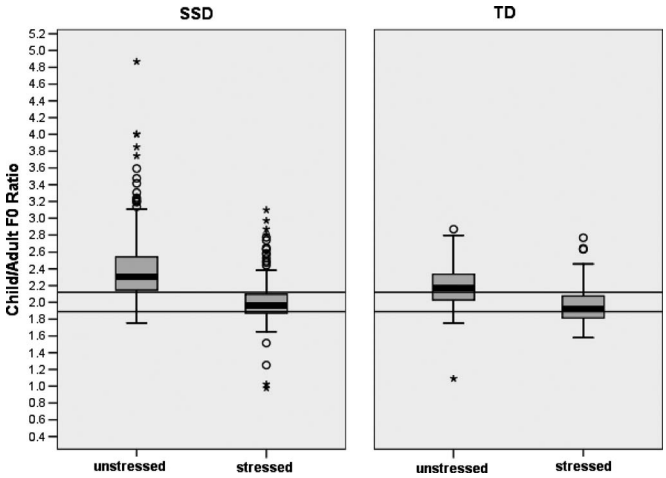


FIG. 2. Distributions of child/adult f_0 ratios by participant group and stress type. Reference lines bracket the OER.

and stressed ($t=2.01$, $p=0.0186$) vowels. Figure 2 shows the distributions of child/adult f_0 ratios for the two participant groups and stress types.

IV. SUMMARY AND DISCUSSION

This study addresses the role of lexical stress and presence of SSD in OSPM in imitated speech tokens, produced by 11 children with SSD and 11 controls. A shift from conversational f_0 levels to double the adult's f_0 level, as opposed to no pitch adjustment or direct pitch matching toward the actual adult f_0 level, is supported by several observations. First, none of the three participants whose conversational f_0 levels fell within one semitone of the doubled adult's f_0 levels dropped f_0 levels during the imitation task. This indicates that direct pitch matching at the adult's actual f_0 levels did not occur in these participants. Second, seven participants lowered their conversational f_0 levels to fall within one semitone of the doubled adult's f_0 levels. Third, the distributions of child/adult f_0 ratios in both participant groups show peaks near a ratio of 2:1, which would not necessarily be expected where direct pitch matching was attempted or no pitch matching occurred. Whether or not the eight participants who lowered their f_0 levels substantially but failed to reach the OER were aiming for this range or for the direct target pitch cannot be ascertained.

OSPM was preferentially expressed in stressed vowels. The distributions of the child/adult f_0 ratios showed that participants tended to lower their f_0 to a level equivalent to twice the targets' f_0 when the syllable was stressed while speaking closer to their conversational f_0 when the syllables were unstressed. This result is consistent with the literature on the psycholinguistic prominence of lexical stress and introduces a new manifestation of this phenomenon.

The participants with SSD did not adjust their conversational f_0 levels in the direction of twice the stimulus f_0 levels as frequently as their peers with TD. The distributions of the child/adult f_0 ratios show that this group difference resulted from a difference in spread, not average values, with the imitations in the SSD group being more variable. The observation that nonparametric testing did not show the group differences that were evident in parametric testing is consistent with this outcome. An important observation resulted from considering the joint effects of SSD and lexical stress. In the stressed vowels, the SSD group reached the OER slightly more frequently than the TD group, while showing much less evidence of OSPM in the unstressed vowels. This is consistent with the finding that children with SSD tend to omit unstressed syllables more frequently than peers with TD and allows the interpretation that unstressed syllables have even less psycholinguistic prominence in SSD than in TD.

An established literature on the perceptual aspects of OE in speech tokens and on direct pitch matching and OSPM in imitation tasks does not yet exist. Therefore, future studies should be conducted to corroborate these experimental results and to evaluate a number of hypotheses in speech-related tasks. They should address (1) the perceptual salience of octave-related speech stimuli, which are implied in this study but were not explicitly quantified; (2) direct pitch matching in contexts where the target f_0 is within the participants' conversational f_0 range but offset from their average f_0 values, for instance, in an imitation task where the target speaker is an adult female and the participants are children; (3) a variety of speech stimuli beyond nonwords, for instance, single words of varying phonologic complexity, pseudowords with varied word and syllable structures, and story retelling; and (4) other types of communication deficits seen in children with autism and other psychiatric disorders where prosody is affected, and the speech disfluencies in children who stutter.

ACKNOWLEDGMENTS

This project represents part of the first author's dissertation study, funded by an institutional training grant from NIDCD (Grant No. 05 T32 DC00033-10) and the Department of Speech and Hearing Sciences at the University of Washington.

Boersma, P., and Weenink, D. (**2004**). PRAAT, Version 4.2.09, Institute of Phonetic Sciences, Amsterdam, The Netherlands.

Brazil, D., Coulthard, M., and Johns, C. (**1985**). *Discourse Intonation and Language Teaching: Applied Linguistics and Language Study* (Longman, London).

Burns, E. (**1999**). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, San Diego), pp. 215–264.

Burns, E. M., and Ward, W. D. (**1978**). "Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals," J. Acoust. Soc. Am. **63**, 456–468.

Echols, C., Crowhurst, M., and Childers, J. (**1997**). "The perception of rhythmic units in speech discrimination in infants," J. Mem. Lang. **36**, 202–225.

Eguchi, I. S., and Hirsh, I. (**1969**). "Development of speech sounds in children," Acta Oto-Laryngol., Suppl. **2**, 57.

Hahn, E. (**1942**). "A study of the relationship between stuttering occurrence and phonetic factors in oral reading," J. Speech Hear Disord. **7**, 143–151.

Hollien, H., and Shipp, T. (**1972**). "Speaking fundamental frequency and chronologic age in males," J. Speech Hear. Res. **15**, 155–159.

Ingram, D. (**1974**). "Phonological rules in young children," J. Child Lang. **1**, 49–64.

Jusczyk, P., Cutler, A., and Redanz, N. (**1993**). "Infants' preference for the predominant stress patterns of English words," Child Dev. **64**, 675–687.

Kehoe, M. M., and Stoel-Gammon, C. (**1997**). "Truncation patterns in English-speaking children's word productions," J. Speech Lang. Hear. Res. **40**, 526–541.

Koike, K. J., and Asp, C. W. (**1981**). "Tennessee test of rhythm and intonation patterns," J. Speech Hear Disord. **46**, 81–86.

Langner, G. (**2007**). "Die zeitliche Verarbeitung periodischer Signale im Hörsystem: Neuronale Repräsentation von Tonhöhe, Klang und Harmonizität (Temporal processing of periodic signals in the auditory system: Neuronal representation of pitch, timbre, and harmonicity)," Z. f. Audiologie, Audiological Acoust. **46**, 8–21.

Mysak, E. D., and Hanley, T. D. (**1959**). "Vocal aging," Geriatrics **14**, 652–656.

Peter, B., and Stoel-Gammon, C. (**2008**). "Central timing deficits in children with primary speech disorders," Clin. Linguist. Phon. **22**, 171–198.

Peter, B., Stoel-Gammon, C., and Kim, D. (**2008**). "Octave equivalence as a measure of stimulus-response similarity during nonword and sentence imitations in young children," in Proceedings of the Fourth Conference on Speech Prosody, edited S. Maduerira, C. Reis, and P. Barbosa, Luso-Brazilian Association of Speech Sciences, Campinas, pp. 731–734.

Saxman, J. H., and Burk, K. W. (**1967**). "Speaking fundamental frequency characteristics of middle-aged females," Folia Phoniatr. (Basel) **19**, 167–172.

Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeny, J. L., and Nadler, C. J. (**2003**). "A diagnostic marker for childhood apraxia of speech: The lexical stress ratio," Clin. Linguist. Phonetics **17**, 549–574.

Wennerstrom, A. K. (**2001**). *The Music of Everyday Speech: Prosody and Discourse Analysis* (Oxford University Press, Oxford).

Acoustic standing wave suppression using randomized phase-shift-keying excitations (L)

Sai Chun Tang[a] and Gregory T. Clement
Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115

(Received 14 May 2009; revised 16 July 2009; accepted 17 July 2009)

Recent papers have demonstrated that acoustic standing waves can be inhibited by frequency-modulated spread-spectrum excitation. An alternative method is studied here that is designed to be more practical for implementation in phased arrays. The method operates using phase-shift-keying (PSK), which introduces phase shifts into the driving signal to break wave symmetry. Sequential and random binary-PSK (BPSK) and quadrature-PSK (QPSK) excitations are studied in water, using a carrier frequency of 250 kHz and a time segment of 10 cycles. The resulting acoustic field is measured with a transducer inside a plastic-walled chamber and compared with continuous wave excitation. Results indicate that both the random BPSK and QPSK methods can reduce time-averaged spatial intensity variation caused by standing waves by approximately six times. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203935]

PACS number(s): 43.80.Sh, 43.20.Ks [CCC] Pages: 1667–1670

I. INTRODUCTION

Standing waves clearly play an important role in many applied acoustic techniques,[1] however, in certain areas of acoustics, standing waves are undesirable. For example, in our laboratory's concentration on high intensity focused ultrasound (HIFU), standing waves have been identified as a possible source of undesirable energy concentrations in transcranial ultrasound therapy.[2–5] Motivated by this problem, we have investigated several techniques for eliminating standing waves using practical methods.[6,7] Previous works demonstrated that time-averaged acoustic standing wave effects can be reduced by sweeping the acoustic driving frequency.[8–12] We recently verified that this technique successfully reduces standing wave effects during transcranial ultrasound sonication and further demonstrated that random frequency modulation has superior performance in standing wave suppression than its swept-frequency counterpart.[6,7]

However, the use of frequency variation introduces a number of potential problems. In therapy, narrowband transducers are generally used in order to maximize power at a given frequency. The electrical impedance of the transducer and its matching network components are also frequency-dependent, and maximum power transfer can only be achieved at a single frequency. Furthermore, phased-array beam steering uses superposition at a given frequency, so that steering a signal with continuously changing frequency would be quite complicated.

In this study, we hypothesize that standing wave symmetry can be broken by randomly changing the phase over discrete time segments while keeping the signal frequency constant. This phase-varying method, known as phase-shift-keying (PSK), was originally used in digital communication.[13] To test this method for standing wave reduction, experiments are performed with four different PSK signal excitations, including sequential and random binary-PSK (BPSK) and sequential and random quadrature-PSK (QPSK). To assess efficacy, partial standing waves are set up in a water tank, and measurements using the various PSK excitations are compared. While the sequential BPSK and sequential QPSK approaches show only marginal reduction in standing waves, both random BPSK and random QPSK signals show significant reduction.

II. MATERIALS AND METHODS

A. PSK signal generation

BPSK and QPSK signals were generated with a custom-made PSK generator (Fig. 1) to demonstrate and compare the efficacy of the standing wave suppression. The carrier signal frequency was set to the transducer resonance (250 kHz). The PSK signals are described by $s=A_o \sin[\omega t+\phi(t)]$, where A_o is the signal amplitude, ω is the angular frequency, and ϕ is the phase angle shifted from the carrier signal.

The BPSK signal phase was switched between 0° and 180°, and the QPSK signal was shifted among 0°, 90°, 180°, and 270° from the carrier signal. The PSK signals were time-segmented over a length of 10 carrier cycles. The randomized BPSK and QPSK were achieved using a random noise generator to randomly select the phase shift of each segment. Figure 2 shows representative waveforms of the random BPSK and QPSK signals. The randomized PSK signals were filtered by a 300 kHz low-pass filter and excited the 250 kHz transducer through a radio frequency linear amplifier (240L, E&I, NY). The efficacy of standing wave suppression using *random* PSK excitations was compared to *sequential* PSK excitations, in which the segment phase shift for the BPSK and QPSK were given regular patterns of 0°, 180°, 0°, 180°,… and 0°, 90°, 180°, 270°, 0°, 90°,…, respectively. The sequential PSK signals were generated by replacing the noise generator with a binary counter.

[a]Author to whom correspondence should be addressed. Electronic mail: sct@bwh.harvard.edu

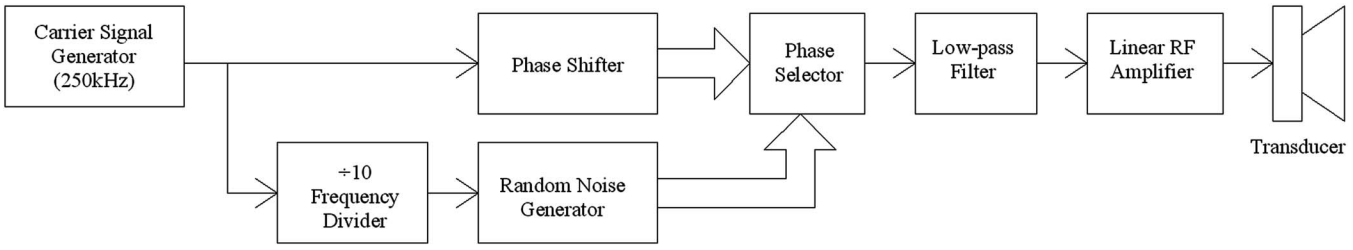


FIG. 1. A block diagram of the randomized PSK signal generator.

B. Experiments

Raster scans of the acoustic field were performed for each type of excitation. Measurements took place inside a tank filled with room temperature degassed, de-ionized water (Fig. 3). A PZT transducer (250 kHz, 50 mm diameter, 10 cm radius of curvature) was driven at a voltage set to 10 Vrms. Two acrylic plates ($150\times90\times5$ mm^3) were separated by 129 mm and aligned parallel to the transducer. The parallel plates are used to mimic the skull to setup the standing waves. The distance between the plate closest to the transducer and the transducer surface was 13 mm. A 2 mm diameter planar transducer (XMS-310-B, Olympus NDT, MA) was scanned by a personal computer controlled three-axis position system and its signal was recorded with an oscilloscope (TDS 3014B, Tektronix, OR). Acoustic line scans were first performed along the transducer axis of symmetry ($x=0$, $y=0$) with a step size of 0.5 mm from $z=20$ mm to 70 mm, where z represented the distance between the transducer faces. Two-dimensional (2D) raster scans were then performed on the x-z plane ($y=0$) from $(x,z)=(-30,20)$ to (30,70) at 1 mm steps.

III. RESULTS

The square of the acoustic pressure was used for analysis, as it is proportional to acoustic intensity. For the specific application of thermal ablation, acoustic intensity, in turn, is proportional to absorptive temperature rise. The z-axis and the x-z plane scans of the normalized square of the pressure field are shown in Figs. 4 and 5. Figure 4 shows the line scans of the square of the acoustic pressure from excitation by the sequential and random BPSK, sequential and random QPSK, and sinusoidal CW (no PSK). Corresponding plots on the x-z plane are shown in Fig. 5. The plots in Figs. 4 and 5 are normalized to their respective average values between $z=40$ mm and 60 mm for comparing the standing wave suppression efficacy. Results from the line scans and the 2D plots reveal partial standing waves when sinusoidal CW (no PSK) and sequential PSK excitations were used. These standing waves are effectively diminished with the random BPSK and QPSK excitations. We quantify the standing wave contribution to the signal by a comparison of the peak, $p^2_{\max}$, to minimum, $p^2_{\min}$, over a region with the average amplitude of p^2_{av},

$$R=\frac{p^2_{\max}-p^2_{\min}}{p^2_{\text{av}}}. \tag{1}$$

This ratio (in percentage) is provided along the z-axis from $z=40-60$ mm in Table I. Both the sequential modes were found to slightly reduce the standing wave effect by less than 10% in R over this distance. The random BPSK and QPSK modes suppressed the standing wave effectively, reducing the ratio from 44% to 7% and 8%, respectively.

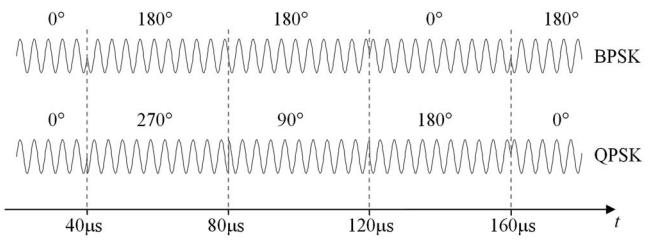


FIG. 2. Representative random BPSK and random QPSK signals with time segment of 10 signal cycles. The phase angles are relative to the carrier signal.

IV. DISCUSSION AND CONCLUSIONS

This study verified our hypothesis that randomly changing phase in time segments can break the symmetry of incident and multi-path reflected waves, thereby inhibiting standing waves. We examined and compared the performance of four potential methods for shifting phase, including sequential and random BPSK and QPSK methods. It was determined that the random PSK methods are more effective in standing wave suppression than their sequential counterparts. We also determined that the performance of BPSK is similar to that of QPSK. However, in practical system implementation, generation of BPSK is generally simpler and more cost-effective than that of QPSK, as the BPSK can be achieved by merely combining an inverter (for 180° phase-shift) and a 2-to-1 multiplexer.

Further investigation on the choice of PSK segment length still must be performed to optimize the efficacy of

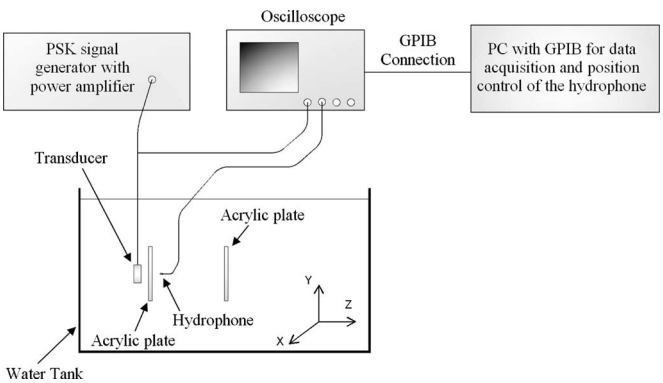


FIG. 3. An experimental setup for the acoustic field scan.

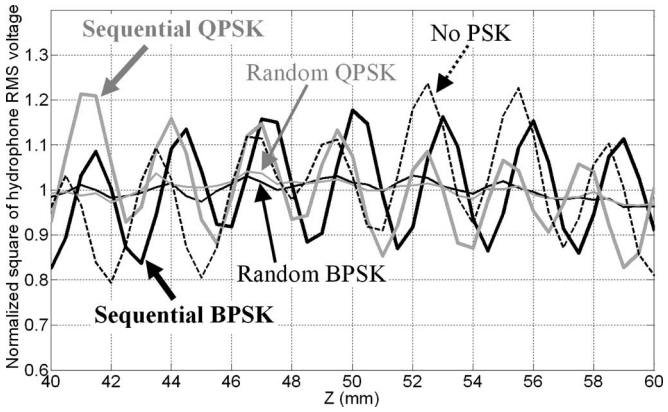


FIG. 4. Normalized square of the acoustic field pressure along the transmitting transducer axis of symmetry with sinusoidal excitation (no PSK) and excitations with random and sequential BPSK and QPSK.

TABLE I. Standing wave effect with different excitations quantified in terms of the R value which is defined as the comparison of the peak, p^2_{max}, over a region with the average amplitude of p^2_{av} as described in Eq. (1).

Excitation	R value
No PSK	44%
Sequential BPSK	35%
Random BPSK	7%
Sequential QPSK	38%
Random QPSK	8%

standing wave suppression while considering application specific operating conditions. The PSK signals used in this work are continuous waves (cws) as indicated in Fig. 2. Alternatively, phase shift in each time segment could be

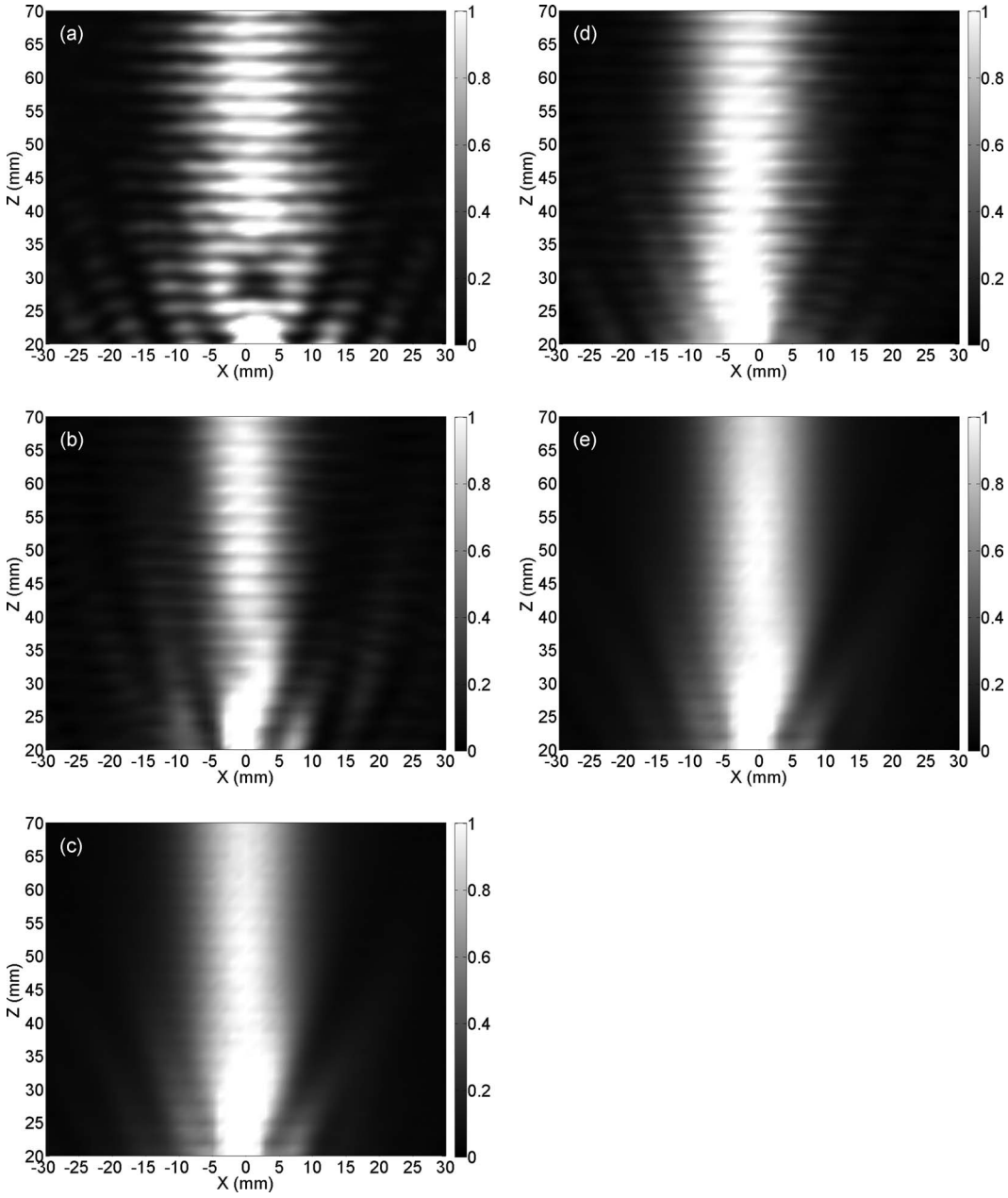


FIG. 5. Normalized square of the acoustic field pressure on the x-z plane (y=0): (a) sinusoidal CW (no PSK), (b) sequential BPSK, (c) random BPSK, (d) sequential QPSK, and (e) random QPSK.

achieved by time delay, relative to the carrier signal. While this delay time is equivalent to a phase shift in the CW PSK signal, during this delay the transducer is not actuated, and this time period can be lengthened for the purpose of time-averaged power control.

In a previous study,[7] acoustic standing waves within an *ex-vivo* human skull were significantly reduced using frequency randomization. With PSK excitations, only the phase of the segmented signal changes and the transducer driving frequency is kept constant. The PSK method avoids the potential problems caused by frequency-varying "spread-spectrum" techniques for standing wave suppression. Since the present study in a parallel-walled chamber represents an extreme case for standing waves, it is reasonable to suggest that the PSK method could have benefit to cw applications within the human skull. Future work will apply randomized PSK to the problem of transcranial propagation using phased arrays.

ACKNOWLEDGMENT

This work was supported by NIH Grant No. U41 RR019703.

[1]F. G. Mitri, U. Eberlein, and M. Fatemi, "Surface roughness imaging using the acoustic emission induced by the dynamic radiation force of ultrasound," Appl. Phys. Lett. **88**, 234105 (2006).
[2]T. Azuma, K. Kawabata, S. Umemura, M. Ogihara, J. Kubota, A. Sasaki, and H. Furuhata, "Schlieren observation of therapeutic field in water surrounded by cranium radiated from 500 kHz ultrasonic sector transducer," Proc.-IEEE Ultrason. Symp. **2**, 1001–1004 (2004).
[3]M. Daffertshofer, A. Gass, P. Ringleb, M. Sitzer, U. Sliwka, T. Els, O. Sedlaczek, W. J. Koroshetz, and M. G. Hennerici, "Transcranial low-frequency ultrasound-mediated thrombolysis in brain ischemia: Increased risk of hemorrhage with combined ultrasound and tissue plasminogen activator: Results of a phase II clinical trial," Stroke **36**, 1441–1446 (2005).
[4]W. C. Culp and T. C. McCowan, "Ultrasound augmented thrombolysis," Current Med. Imaging Reviews **1**, 5–12 (2005).
[5]J. F. Aubry and M. Fink, "Transcranial ultrasound-mediated thrombolysis: Safety issue," in Proceedings of the 7th International Symposium on Therapeutic Ultrasound (2007).
[6]S. C. Tang and G. T. Clement, "Approach for inhibiting standing waves in transcranial ultrasound therapy," Proceedings of the 8th International Symposium on Therapeutic Ultrasound (2008), pp. 3–7.
[7]S. C. Tang and G. T. Clement, "Standing wave suppression in transcranial ultrasound therapy using random-signal-modulation excitation," 2008 Proc. IEEE Int. Ultrason. Symposium (2009), pp. 2048–2051.
[8]F. G. Mitria and M. Fatemi, "Improved vibroacoustography imaging for nondestructive inspection of materials," J. Appl. Phys. **98**, 114901 (2005).
[9]B. Louis and G. B. Long, "Sonic washer," U.S. Patent No. 2985003 (1961).
[10]B. H. Candy, "Ultrasonic vibration generator and use of same for cleaning objects in a volume of liquid," U.S. Patent No. 5496411 (1996).
[11]F. G. Mitri, J. F. Greenleaf, and M. Fatemi, "Chirp imaging vibroacoustography for removing the ultrasound standing wave artifact," IEEE Trans. Med. Imaging **24**, 1249–1255 (2005).
[12]T. N. Erpelding, K. W. Hollman, and M. O'Donnell, "Bubble-based acoustic radiation force using chirp insonation to reduce standing wave effects," Ultrasound Med. Biol. **33**, 263–269 (2007).
[13]F. G. Stremler, *Introduction to Communication Systems*, 3rd ed. (Addison-Wesley, Reading, MA, 1990), Chap. 10, pp. 618–625.

Comments on "The Science and Applications of Acoustics" [Springer-Verlag (2000, 2006)] (L)

Kenneth A. Cunefare[a)]
The Woodruff School of Mechanical Engineering, The Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 4 November 2008; revised 12 May 2009; accepted 4 June 2009)

Raichel's "The Science and Applications of Acoustics" (Springer-Verlag, Berlin, 2000) has extensive material that appears to have been extracted from four other texts without proper attribution. The material so used extends to figures, detailed structure of entire sections and developments, and phrasing of individual passages. While in some instances like material between the texts has been paraphrased to some extent, in others it appears almost unaltered. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3177255]

PACS number(s): 43.10.Hj [ADP] Pages: 1671–1680

I. INTRODUCTION

This paper is intended to bring to the attention of the community a potential issue of integrity in scholarship with respect to Daniel R. Raichel's "The Science and Applications of Acoustics."[1] Raichel's textbook was reviewed in the Journal of the Acoustical Society of America in 2003 (Ref. 2) and in the Journal of Audio Engineering Society in 2000.[3] I have discovered that Raichel's book appears to use material from at least four other texts without adequate attribution and citation as is required in scholarly works. As will be explored in greater detail below, I have found that Raichel's text contains passages and developments traceable to *Fundamentals of Acoustics* by Kinsler *et al.*,[4] *Noise Control* by Wilson,[5] *Music, Physics, and Engineering* by Olson,[6] and *Principles of Underwater Sound* by Urick.[7] The degree of duplication extends to figures (including what appear to be scans of figures), detailed structure of entire sections and developments, and phrasing of individual passages. While in some instances like material between the texts has been paraphrased to some extent, in others it appears almost unaltered.

While the texts of Kinsler *et al.*, Olson, Wilson, and Urick are listed as references at the back of individual chapters of the Raichel text, the degree to which Raichel appears to have borrowed from these sources probably should have warranted explicit citation to the material, and even the use of quotation marks; this was not done. Further, there are instances of figures in the Raichel text that appear to be scans from other texts yet the Raichel figures lack attribution in their captions or in the associated written discussion.

Raichel's book is a textbook, and it would appear to be commonly accepted within the scholarly community that there is less an expectation of original scholarship for textbooks as compared to, for example, an archival journal publication. Discussion of this issue with respect to plagiarism may be found in Ref. 8 and in the American History Association's Standards of Professional Conduct.[9] Nonetheless, it is also commonly accepted that extensive paraphrasing and incorporation of material down to the structure of passage or section without attribution constitutes plagiarism. When considering a case of possible plagiarism, it is important to keep in mind that such is determined by the *similarities* between texts, and not the *differences*.

Sections II through V provide examples extracted from the texts Kinsler *et al.*, Wilson, Olson, and Urick and compared to like material extracted from Raichel. The examples range from brief sections of prose and end-of-chapter problems, to descriptive listings of material that is of such extensive nature that its replication here would be excessive; the reader is invited to compare these sections for themselves. The former are provided as examples that the problem exists, while the latter are provided as indications of the extent of the problem. Again, recall that in matters such as are of concern here, it is not the differences that are significant; rather, it is the similarities. In the following, **KFCS** will be used to denote the Kinsler, Frey, Coppens, and Sanders text (which appears to be the text that was used most extensively), while **Olson**, **Wilson**, **Urick**, and **Raichel** will be used to denote the texts by these respective authors. The examples below are relevant to the first edition of Raichel's text. While a second edition was published in 2006, most of the deficiencies remain in the second edition. Further, it must be understood that the following does not constitute a comprehensive listing of all instances of possible plagiarism within the Raichel text; rather, it is a listing derived from comparing a subset of the Raichel text against a subset of material from such other texts as I was readily able to recognize, and even then, only bringing forward a subset of what was found. The structure employed below to bring out the matter at hand would appear to be a common approach; the side-by-side comparison of like material from the subject texts.

II. COMPARISONS BETWEEN RAICHEL AND KFCS

Raichel appears to extensively use material from the **KFCS** text. Material from **KFCS** Chaps. 2, 3, 4, 9, and 10

[a)]Electronic mail: ken.cunefare@me.gatech.edu

may be found in Raichel's Chaps. 4–7. The similarities extend from order of presentation, through structure of sections, structure of developments, structure of paragraphs, examples, prose, figures, and end-of-chapter problems. The following represents a sampling of this material.

A. Brief examples of like prose

In each of the following paired comparisons, note the high degree of similarity between the **KFCS** prose and the **Raichel** prose, with the differences suggesting an effort to paraphrase **KFCS**.

KFCS Chap. 3.11, p. 74. "If a bar is struck in such a manner that the amplitudes of vibration of some of the overtones are appreciable, the sound produced has a metallic quality. However, these high-frequency overtones are rapidly damped out, so that the initial sound is soon mellowed into a nearly pure tone, whose frequency is that of the fundamental. A struck tuning fork exhibits the above characteristics of an initial metallic sound, which rapidly dies out leaving a nearly pure tone."

Raichel Chap. 5.8, p. 105. "When a metal bar is struck in such a manner that the amplitudes of the vibration of some of the overtones are fairly strong, the sound produced has a metallic cast. But these overtones die out rapidly, and the initial sound soon evolves into a mellower pure tone whose frequency is the fundamental. This is characteristic of the behavior of a tuning fork that emits a short metallic sound upon being struck before emitting a pure tone."

KFCS Chap. 4, p. 95. "The most important utilization of the vibrating thin plate is in the diaphragms of ordinary telephone microphones and receivers. Although the response of these devices is not uniform over a wide range of frequencies, they give adequate intelligibility and are simple and rugged in their construction. Another application is in sonar transducers used for producing sounds in water at frequencies below 1 kHz; sound is generated by the motion of relatively thin circular steel plates driven by alternations in the magnetic field of an adjoining electromagnet."

Raichel Chap. 6.8, p. 127. "The most apparent use of the vibrating thin plate is that of the telephone diaphragms (both receiver and microphone). While these diaphragms do not provide the flatter frequency response or frequency range of membranes in condenser microphones, they do provide adequate intelligibility, are generally far more rugged in their construction and cheaper to manufacture. Sonar transducers used to generate underwater sounds less than 1 kHz constitute another class of vibrating plates; the signals are produced by the variations of an electromagnetic field in an electromagnet positioned closely to a thin circular steel plate."

B. Specific example of parallel development

The following presents a portion of a development in **Raichel** and a corresponding development in **KFCS**. Note the identical sequence of equations and paraphrased prose. This is but one example of instances where it appears that large segments of **KFCS** have been used in **Raichel**.

KFCS 9.8 The waveguide of constant cross-section

Consider a waveguide with a rectangular cross-section, as shown in Figure 9.7. Assume the side walls to be rigid and the boundary $z=0$ to be a source of acoustic energy. The absence of another boundary on the z axis allows energy to propagate down the waveguide. This suggests a wave pattern consisting of standing waves in the transverse directions (x and y) and a traveling wave in the z direction.

Since the cross-section is rectangular and the boundaries are rigid, it can be seen intuitively and verified mathematically that acceptable eigenfunctions are

$$\mathbf{p}_{lm} = \mathbf{A}_{lm} \cos k_{xl} x \cos k_{ym} y e^{j(\omega t - k_z z)}, \qquad (9.53)$$

where substitution into the wave equation shows that

$$(\omega/c)^2 = k^2 = k_{xl}^2 + k_{ym}^2 + k_z^2 . \qquad (9.54a)$$

The allowed values of k_{xl} and k_{ym} are found from the rigid boundary conditions to be

Raichel 7.8 wave guide with constant cross-section

In Fig. 7.3 a waveguide of rectangular cross-section is assumed to have rigid side walls and source of acoustic energy located at its boundary $z=0$. There is no other boundary on the z-axis, which permits energy to propagate down the waveguide. This results in a situation where the wave pattern consists of standing waves in the transverse directions x and y and a traveling wave in the z-direction. The mathematical solution that contains applicable eigenfunctions is

$$p_{lm} = A_{lm} \cos k_{xl} x \cos k_{ym} y e^{j(\omega t - k_z z)} . \qquad (7.38)$$

Upon substitution equation (7.38) in the wave equation (7.31), we obtain the relationship

$$\left(\frac{\omega}{c}\right)^2 = k^2 = k_{xl}^2 + k_{ym}^2 + k_z^2 \qquad (7.39)$$

with permitted values of k_{xl} and k_{ym}, resulting from the boundary conditions of rigidity, these being

$$\begin{aligned} k_{xl} &= l\pi/L_x, \quad l = 0,1,2,\ldots, \\ k_{ym} &= m\pi/L_y, \quad m = 0,1,2,\ldots . \end{aligned} \qquad (7.40)$$

$$k_{xl} = l\pi/L_x, \quad l = 0,1,2,\ldots,$$
$$k_{ym} = m\pi/L_y, \quad m = 0,1,2,\ldots, \tag{9.54b}$$

and k_z is therefore by

$$k_z = \sqrt{(\omega/c)^2 - k_{xl}^2 - k_{ym}^2}. \tag{9.55}$$

Since ω can now have any value, $\mathbf{p}_{\mathrm{lm}}$ is a solution for *all* values of ω, in contrast with the cavity for which the allowed frequencies are quantized. Thus $\vec{k}$ is not fixed, and neither is k_z. It is convenient to define the *transverse component* k_{lm} of the propagation vector $\vec{k}$. For this rectangular cross-section,

$$k_{lm} = \sqrt{k_{xl}^2 + k_{ym}^2} \tag{9.65a}$$

so that equation (9.55) can be written more succinctly as

$$k_z = \sqrt{(\omega/c)^2 - k_{lm}^2}. \tag{9.56b}$$

When $\omega/c > k_{lm}$, then k_z is real. The wave advances in the $+z$ direction and the eigenfunction is called a *propagating mode*. The limiting value for ω/c for which k_{lm} remains real is given by $\omega/c = k_{lm}$, and defines the *cutoff frequency*

$$\omega_{lm} = ck_{lm} \tag{9.57}$$

for the (l, m) mode. If the input frequency is lowered below cutoff, the argument of the square root in (9.56b) becomes negative and k_z must be pure imaginary

$$k_z = \pm j\sqrt{k_{lm}^2 - (\omega/c)^2}. \tag{9.58}$$

The minus sign must be taken on physical grounds so that $\mathbf{p} \to 0$ as $z \to \infty$, and the eigenfunctions have the form

$$\mathbf{p}_{lm} = \mathbf{A}_{lm} \cos k_{xl}x \cos k_{ym}y \times \exp[- \sqrt{k_{lm}^2 - (\omega/c)^2}z]e^{j\omega t}, \tag{9.59}$$

which represents a standing wave that decays exponentially with z.

We can rearrange equation (7.39) to find k_z:

$$k_z = \sqrt{\left(\frac{\omega}{c}\right)^2 - k_{xl}^2 - k_{ym}^2}. \tag{7.41}$$

Because ω may have any value, equation (7.38) comprises a solution for all values of ω, in contrast to the totally enclosed cavity which allows for only quantized frequencies.

Setting

$$k_{lm} = \sqrt{k_{xl}^2 + k_{ym}^2}, \tag{7.42}$$

we can shorten equation (7.41) to

$$k_z = \sqrt{\left(\frac{\omega}{c}\right)^2 - k_{lm}^2}. \tag{7.43}$$

The value k_z is real when $\omega/c > k_{lm}$. We then obtain a propagating mode, as the wave moves in the $+z$ direction. The cutoff frequency, which occurs when $\omega/c = k_{lm}$ so defining the limit for which k_z remains real, is given by

$$\omega_{lm} = ck_{lm} \tag{7.44}$$

for the (l, m) mode. A frequency below the threshold value of ω_{lm} results in a purely imaginary value of k_z:

$$k_z \equiv \pm j\sqrt{k_{lm}^2 - (\omega/c)^2}. \tag{7.45}$$

We need to include the negative sign in equation (7.45) so that $\mathbf{p} \to 0$ as $z \to \infty$, and the eigenfunctions assume the form

$$\mathbf{p}_{lm} = A_{lm} \cos k_{xl}x \cos k_{ym}y \times e^{-\left(\sqrt{k_{lm}^2 - \left(\frac{\omega}{c}\right)^2}\right)z}e^{j\omega t}. \tag{7.46}$$

Equation (7.46) represents a standing wave that decays exponentially with z.

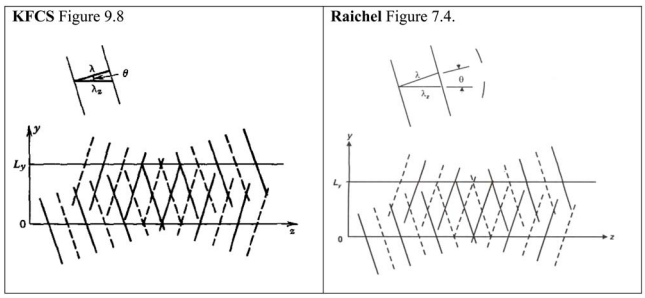
C. Sections of parallel/identical development

KFCS Chaps. 3.11 and 3.12, pp. 72–75. Also, compare **KCFS** Figs. 3.7–3.9 to **Raichel** Figs. 5.8–5.10.	**Raichel** Chap. 5.8, "Case 1" and "Case 2," pp. 103–107.
KFCS Chaps. 4.6 and 4.7, pp. 87–95.	**Raichel** Chap. 6.6, pp. 119–128.
KFCS Chap. 9.8, p. 216, through discussion of Eq. (9.64b), p. 220 and material on circular waveguide starting at bottom of p. 221. Also, compare **KFCS** Figs. 9.8 and 9.9 to **Raichel** Figs. 7.4 and 7.5.	**Raichel** Chaps. 7.8, 7.9, and 7.10, pp. 140–143.
KFCS Chaps. 10.1, 10.2, pp. 225–228.	**Raichel** Chap. 711, pp. 143–147

D. Example of matched figures

The following depicts an example of a figure in **Raichel** that appears to have been derived from **KFCS**. The figures address constant wavefronts in a rigid-walled waveguide (the figures may be found in Sec. 9.8 of **KFCS**, and Sec. 7.8 of **Raichel**; these are the sections a portion of which were presented above in "Sections of parallel/identical development"). While not a scan of **KFCS** Fig. 9.8, one finds in **Raichel** Fig. 7.4 the identical angles for the wavefronts, and the identical number of wavefronts and the style of how they

are depicted would suggest that the figure was generated by some manner of overlay or redrawing of the **KFCS** figure.



E. Examples of parallel homework problems

The following exemplar segments of end-of-chapter problems from **KFCS** and **Raichel** demonstrates that certain of **Raichel's** problem sections are of the same structure, sequence, and content as **KFCS**. For these parallels in structure and sequence to have occurred serendipitously would be unlikely. The numbering of the problems from the **KFCS** and **Raichel** texts is as in the original; problems from **KFCS** that do not appear in **Raichel** have been omitted from the following. As with other material considered in this document, the following represents only a subset of the subject material.

KFCS Chap. 2, problems 2.2–2.10, formatted as in **KFCS**.

2.2. By direct substitution show that each of the following are solutions of the wave equation: (a) $f_1(x-ct)$, (b) $\ln[a(ct-x)]$, (c) $a(ct-x)^2$, and (d) $\cos[a(ct-x)]$. Similarly, show that each of the following are *not* solutions of the wave equation: (e) $a(ct-x^2)$ and (f) $at(ct-x)$.

2.3. Sketch $y=A\exp(-a|ct-xl)$ for $t=0$, $t=1$, and $t=2$ s. Let $c=5$ cm/s, $a=3$ cm^{-1}, and $A=1$ cm. What is the significance of the displacement of these curves?

2.4. Consider the waveform $y=4\cos(3t\mathrm{t}-2x)$ propagating on a string of density 0.1 g/cm, where y and x are in centimeters and t is in seconds. (*a*) What is the amplitude, phase speed, frequency, wavelength, and wave number? (*b*) What is the particle speed of the element at $x=0$ at $t=0$?

2.5. Evaluate the mechanical impedance seen by the applied force driving a semi-infinite string at a distance L from a rigid end. Interpret the individual terms in the mechanical impedance.

2.7. A string is stretched between rigid supports a distance L apart. It is driven by a force $F\cos\omega t$ located at its midpoint. (a) What is the mechanical impedance at the midpoint? (b) Show that the amplitude of the midpoint is $(F/2kT)\tan(kL/2)$. (c) What is the amplitude of the displacement of the point $x=L/4$?

2.8. A string of density 0.01 kg/m is stretched with a tension of 5 N from a rigid support at one end to a device producing transverse periodic vibrations at the other end. The length of the string is 0.44 m, and it is observed that, when the driving frequency has a given value, the nodes are spaced 0.1 m apart and the maximum amplitude is 0.02 m. What are the (a) frequency and (b) amplitude of the driving force?

2.9. (a) Assume that a forced, fixed string is driven by a source that has constant speed amplitude $u(0,t)=U_0\exp(j\omega t)$, where U_0 is independent of frequency. Find the frequencies of maximum amplitude of the standing wave. (*b*) Repeat for a source that has a constant displacement amplitude $y(0,t)=Y_0=Y_0\exp(j\omega t)$. (c) Contrast the results of (a) and (b) with the frequencies of mechanical resonance for the forced, fixed string. (d) Does mechanical resonance always coincide with maximum amplitude of the motion?

2.10. Given a string, fixed at both ends, with ρ_L, c, L, f, and T specified so that c and f are known numbers, obtain the phase speed c' in terms of c and the fundamental resonance f' in terms of f if another string of the same material is used but (a) the length is doubled, (b) the density/length is quadrupled, (c) the cross-sectional area is doubled, (d) the tension is reduced to half, and (e) the diameter of the string is doubled.

Raichel Chap. 4, problems 1–9, formatted as in **Raichel**. Note that the content of each question matches to questions in **KFCS**, and in almost the same order. Minor changes have been made between the problems. Of this sample, **KFCS** 2.6 does not appear in **Raichel**. Notes in **bold** link the **Raichel** problems to the corresponding **KFCS** problems.

1. Show by direct substitution that each of the following expressions constitutes solutions of the wave equation: (**KFCS No. 2.2**)
   (a) $f(x-ct)$
   (b) $\ln[f(x-ct)]$
   (c) $A(ct-x)^3$,
   (d) $\sin[A(ct-x)]$
2. Show which of the following are solutions and not solutions to the wave equation: (**KFCS No. 2.2**)
   (a) $B(ct-x^2)$
   (b) $C(ct-c)t$

(c) $A+B\sin(ct+x)$

(d) $A\cos^2(ct-x)+B\sin(ct+x)$

3. Plot (by computer if possible) the expression $y=Ae^{-B(ct-x)}$ for times $t=0$ and $t=1.0$, with $A=6$ cm, $B=4$ cm^{-1}, and $c=3$ cm/s. Discuss the physical significance of these curves. (**KFCS No. 2.3**)

4. Consider a string of density 0.05 g/cm, in which a wave form $y=4\cos(5t-3x)$ is propagating. Both x and y are expressed in centimeters, and time t in seconds. (**KFCS No. 2.4**)

(a) Determine the amplitude, phase speed, frequency, wavelength, and the wave number.

(b) Find the particle speed of the string element at $x=0$ at time $t=0$.

5. A string is stretched with tension T between two rigid supports located at $x=0$ and $x=L$. It is driven at its midpoint by a force $F\cos\omega t$. (**KFCS No. 2.7**)

(a) Determine the mechanical impedance at the midpoint.

(b) Establish that the amplitude of the midpoint is given by $F\tan(kL/2)/(2kT)$.

(c) Find the amplitude of displacement at the quarter point $x=L/4$.

6. Determine the mechanical impedance with respect to the applied force driving a semi-infinite string at a distance L from the rigid end. What is the significance of the individual terms in the expression for mechanical impedance? (**KFCS No. 2.5**)

7. Consider a string of density 0.02 kg/m that is stretched with a tension of 8 N from a rigid support to a device producing transverse periodic vibrations at the other end. The length of the string is 0.52 m. It is noted that for a specific driving frequency, the nodes are spaced 0.1 m apart and the maximum amplitude is 0.022 m. What are the frequency and the amplitude of the driving force? (**KFCS No. 2.8**)

8. A device that has a constant speed amplitude $u(O,t)=U_oe^{i\omega t}$, where U_o is a constant, drives a forced, fixed string. (**KFCS No. 2.9**)

(a) Find the frequencies of maximum amplitude of the standing wave.

(b) Repeat the problem for a constant displacement amplitude $y(0,t)=Y_0e^{i\omega t}$.

(c) Compare the results of (a) and (b) with the frequencies of mechanical resonances for the forced fixed string. Does the mechanical amplitude coincide with the maximum amplitude of the motion?

9. Consider a string fixed at both ends, with specified values of ρ_L, c, L, f, and T. Express the phase speed c' in terms of c and the fundamental resonance f' in terms of f if another string of the same materials is used, but (**KFCS No. 2.10**)

(a) the length of the string is doubled;

(b) the density per unit length is doubled;

(c) the cross-sectional area is doubled;

(d) the tension reduced by half;

(e) the diameter of the string is doubled.

As another example, compare **KFCS** problems 4.1, 4.2, 4.4, 4.5, 4.8, 4.9, 4.10, and 4.11 to **Raichel** Chap. 6, problems 1–8. As above, the problems in **Raichel** appear to be substantially the same as in **KFCS**, and in the same sequence.

III. COMPARISONS BETWEEN RAICHEL AND WILSON

As compared to the volume of material apparently from **KFCS**, the use of material from **Wilson** within **Raichel** is to a lesser extent. Nonetheless, material from **Wilson** Chaps. 3, 6, and 7 may be found in **Raichel** Chaps. 9, 12, and 14, respectively.

A. Brief examples of like prose

Wilson Chap. 3.10, "Data Windows," p 119. "Consider a steady pure tone signal of unknown frequency. If this signal was to be analyzed, it would ordinarily be sampled over a short time interval (called the *window duration*). A *rectangular* window passes a segment of the original signal without adjustment. It might be assumed that the short segment so obtained was representative of the original signal. The assumption would be true of the segment contained an integer number of periods of the original signal."

Raichel Chap. 9.14. "If a steady pure-tone of an unknown frequency is to be analyzed, it would usually be sampled over a short timer interval that is termed window duration. A rectangular window will pass a portion of the input signal without adjustment. That short segment obtained is presumably representative of the original signal, and this would hold true if that segment embodies an integer number of periods of the original signal."

Wilson Chap. 6.1, p. 239. "When airborne sound reaches a wall, some of the sound energy is reflected, some energy is absorbed, and some energy is transmitted through the wall. …Sound pressure incident on one side of a wall can cause the wall to vibrate and transmit sound energy to the other side. The fraction of incident sound energy transmitted to the

wall depends on the impedance of the wall relative to the air."

Raichel Chap. 12.2, p. 261. "When airborne sound impinges on a wall, some of the sound energy is reflected, some energy is absorbed within the wall structure, and some energy is transmitted through the wall. Sound pressure against one side of the wall may cause the wall to vibrate and transmit sound to the other side. The amount of incident energy transmitted to the wall depends on the impedance of the wall relative to the air."

Wilson Chap. 7.2, p. 308. "Electric motors and internal combustion engines are ordinarily designed to operate at speeds of one-thousand to several thousand revolutions per minute. High-speed operation usually produces the greatest ratio of power to weight and the greatest ratio of the power to initial cost. Geared transmissions and other speed reducers are used when the driven machinery has high torque and low speed requirements."

Raichel Chap. 14.8, p. 349. "Internal combustion engines and electrical motors generally operate at speeds of one to several thousand revolutions per minute. These high speeds help maximize the power-to-weight and power-to-initial cost ratios. Gearing and other speed reducers are applied when the driven machinery requires high torque and low speeds."

B. Sections of parallel/identical development

Wilson Chaps. 6.1 and 6.2, pp. 239–249. Also, compare figures throughout.	**Raichel** Chaps. 12.2 and 12.3, pp. 261–271.
Wilson Chap. 6.2, pp. 252–253. Also, compare **Wilson** Figs. 6.2.8–6.2.9 to **Raichel** Figs. 12.8, 12.9.	**Raichel** Chap. 12.8, pp. 273–274.
Wilson Chap. 7.2, p. 317, "Contact Ratio" paragraph.	**Raichel** Chap. 14.8, p. 354, "Contact Ratio" paragraph.
Wilson Chap. 7.2, p. 321, "Other Considerations…" section.	**Raichel** Chap. 14.8, p. 355, "Other aspects…" section.

IV. COMPARISONS BETWEEN RAICHEL AND OLSON

The volume of material that appears to be from **Olson** is less than **KFCS**. In contrast to Raichel's apparent use of re-drawn figures from **KFCS**, Raichel appears to have used numerous figures scanned from **Olson**. The use is not restricted to figures, and includes prose.

A. Brief examples of like prose

Olson ChAP. 2.2, p. 25. "One of the principal reasons for the definite and unique frequencies of musical tones is that musical instruments are inherently resonant systems and therefore respond to only certain frequencies. Except for certain instruments, as for example, the trombone and the violin family, the resonant frequencies are fixed and cannot be altered at will."

Raichel Chap. 17.2, p. 474. "A principal reason that we can identify a musical instrument is that the musical instruments are essentially resonant instruments and therefore exhibit a response only certain frequencies. These resonant frequencies are fixed and cannot be altered, except for certain instruments such as members of the violin family and trombones.

Olson Chap. 5, p. 163. "A tuning slide or bit is provided so that the resonant frequencies of the trumpet can be made to coincide with other instruments. A mute in the form of a pear-shaped piece of metal or plastic which fits into the bell is used to change quality and attenuate the sound output."

Raichel Chap. 17.10, p. 511. "A tuning slide or bit is provided so that the resonant frequencies can be matched with other instruments…. A mute in the form of pear-shaped piece of metal or plastic can be inserted into the bell in order to change the quality of the tone and attenuate the output."

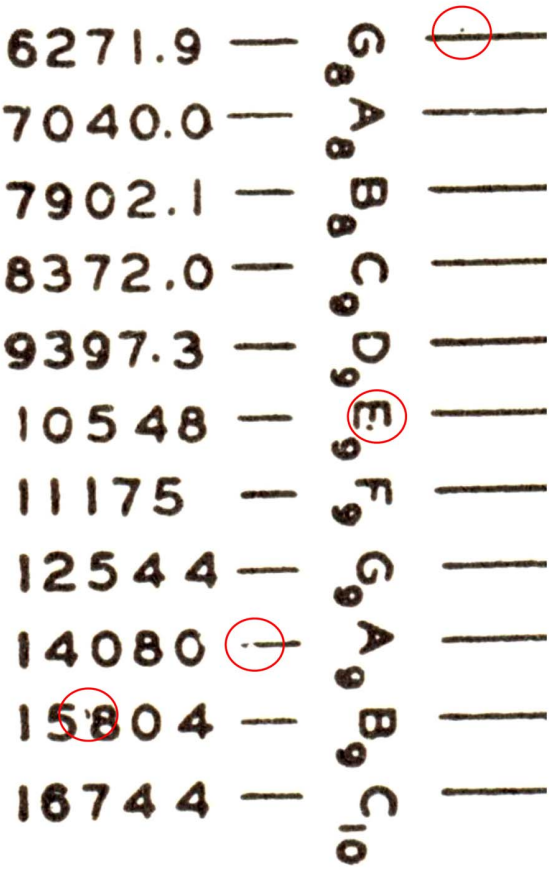
B. Sections of parallel/identical development

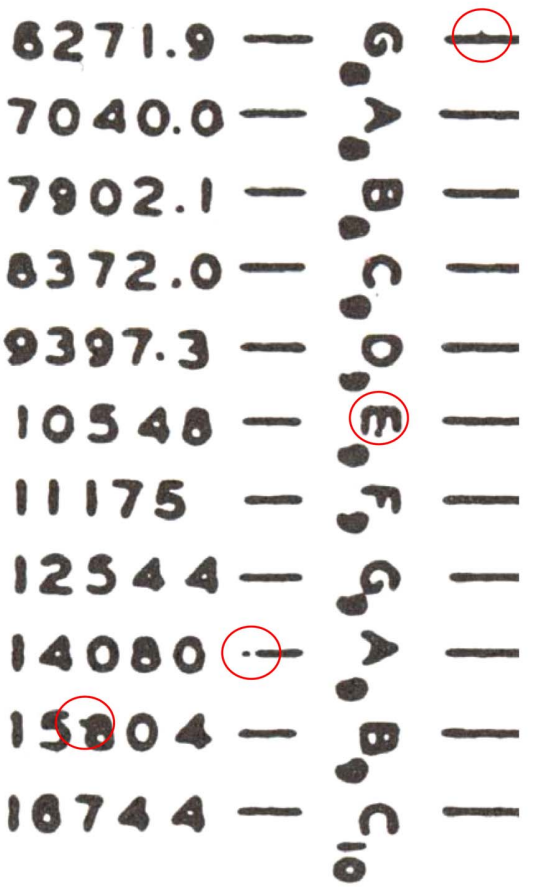
Olson Chap. 2 and its figures throughout, pp. 25–36.	**Raichel** Chaps. 17.2–17.6, pp. 473–484. Some of the figures are scans of the **Olson** figures, as may be seen by identical flaws in like figures, e.g., **Raichel** Fig. 17.7 is a scan of **Olson** 2.9 (this is demonstrated in the following).
Olson Chap. 5 through material on lip reed instruments, pp. 108–167.	**Raichel** Chaps. 17.8–17.10, pp. 485–512. Note that some figures from Olson were explicitly cited, while others were not. Some of the figures are scans of the **Olson** figures, e.g., **Raichel** Fig. 17.14 is a scan of **Olson** Fig. 5.5.
Olson Chaps. 5.7, 5.8, pp. 191–192.	**Raichel** Chap. 17.14, pp. 522–523.

C. Examples of seemingly duplicated figures

Olson Fig. 2.9, segment showing a portion with printing defects.

Raichel Fig. 17.7, segment with same printing defects, as indicated by the circled flaws common to both the **Olson** and **Raichel** figures. The **Olson** figure is not cited.





Olson Fig. 5.22

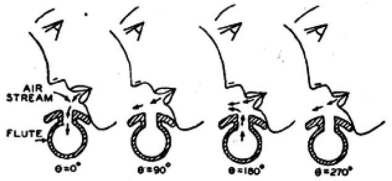


FIG. 5.22. The particle velocities at the embouchure of the flute for a complete cycle. The arrows indicate the direction of the air flow.

Raichel Fig. 17.25. This appears to be a touched-up scan of **OLSON** Fig. 5.22. The **Olson** figure is not cited.

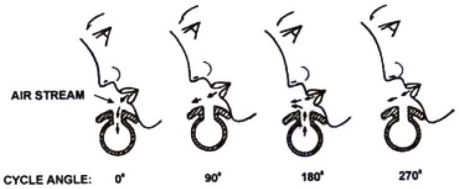


FIGURE 17.25. The action of the flute. The direction of the arrows indicate air flow direction.

V. COMPARISONS BETWEEN RAICHEL AND URICK

Material from **Urick** Chaps. 2, 5, and 6 may be found in **Raichel** Chap. 15. The use extends to prose and figures, with some figures in **Raichel** appearing to be scans of figures from **Urick**.

A. Brief examples of like prose

Urick Chap. 5.5, and figures, p. 116–120. "Just below the sea surface is the surface layer, in which the velocity of sound is susceptible to daily and local changes of heating, cooling, and wind action. The surface layer may contain a mixed layer of isothermal water that is formed by the action of the wind as it blows across the surface above."

Raichel Chap. 15.4, p. 390–293. "The surface layer lies just below the sea surface. The speed of sound in that layer is responsive to daily and local changes of heating, cooling, and action of the winds. The surface layer may consist of a mixed layer of isothermal water that is caused by action of the wind as it blows across the surface of the water."

Urick Chap. 5.3, p. 107. "The boric acid…ionization process is by no means a simple one, since the ionic interaction mechanism appears to depend in a complicated way on the other chemicals present in the seawater solution."

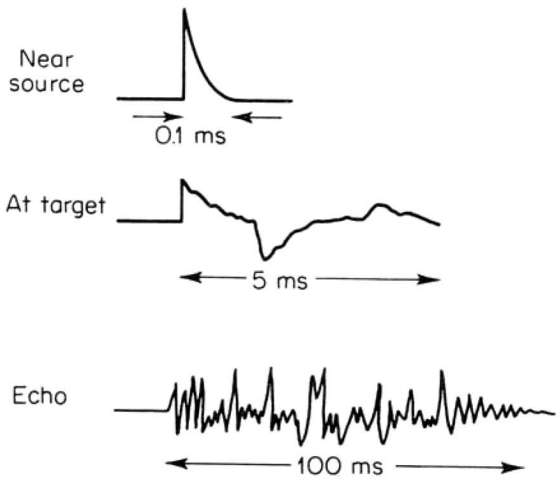
Raichel Chap. 15.5, p. 397. "The boric acid ionization process is not a simple one, because its mechanism seems to rely in a complicated way on the presence of other chemicals in the seawater solution,…"

B. Sections of parallel/identical development

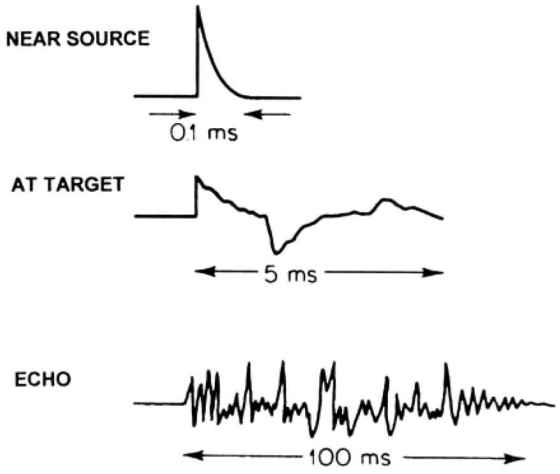
Urick Chaps. 5.2, 5.3, pp. 100–111.	**Raichel** Chaps. 15.5, 15.6. Some figures appear to be scans from **Urick**, e.g., **Raichel** Fig. 15.5 appears to be identical to **Urick** Fig. 5.7.
Urick Chap. 2.1–2.8, pp. 18–29.	**Raichel** Chaps. 15.12–15.16, pp. 404–414. Some figures appear to be scans from **Urick**, e.g., **Raichel** Fig. 15.13 is a relabeled scan of **Urick** Fig. 2.5.

C. Examples of scanned figures

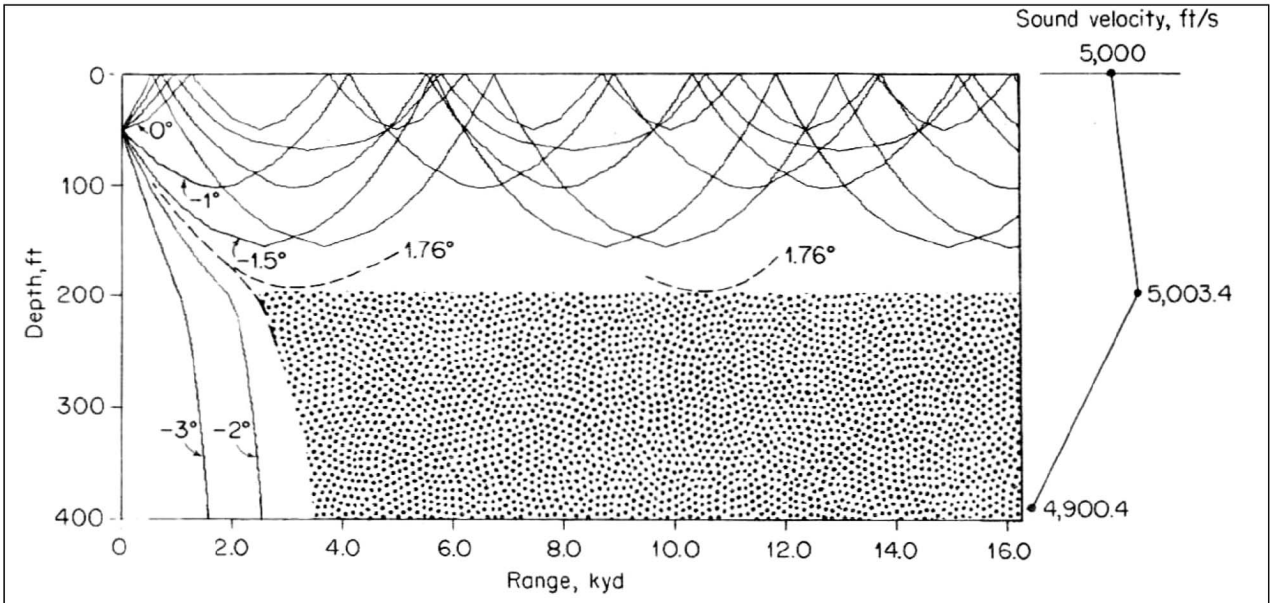
Urick Figure 2.5



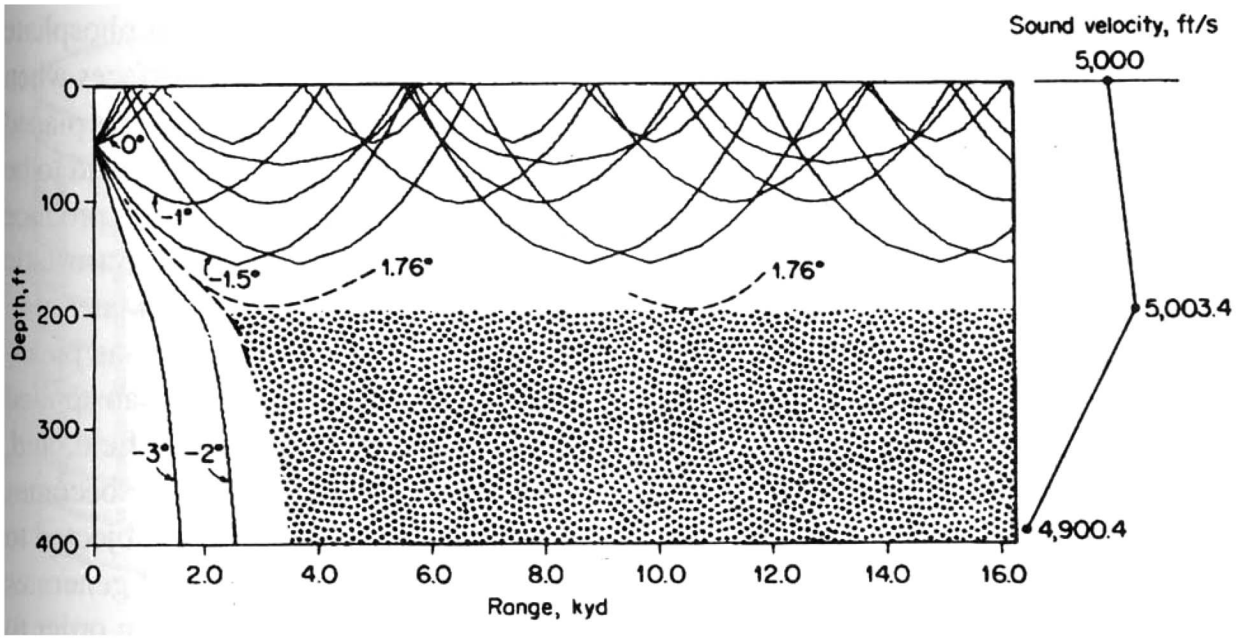
Raichel Fig. 15.13. Note the "NEAR SOURCE", "AT TARGET" and "ECHO" plots appear to be identical to the corresponding plots in **Urick** Fig. 2.5, possibly scans. **Urick** is not cited as the source of the figure.



Urick Figure 6.1



Raichel Figure 15.7. This appears to be a scan of **Urick** Fig. 6.1, as suggested by the dot pattern being identical, as well as the presence of printing flaws within the **Raichel** figure that are in the **Urick** figure. **Urick** is not cited as the source of the figure.



VI. CONCLUSION

As illustrated above, the degree to which the Raichel text appears to use material without attribution is, in my opinion, at worst, plagiarism, or at best, flawed scholarship. In light of what I have discovered with the Raichel text, the authorship of the Raichel text appears to go against the very practices we seek to inculcate within our students.

ACKNOWLEDGMENTS

The author would like to thank Professor Peter Rogers for assisting in identifying the material from the Urick text, and Joseph W. Dickey for bringing forward additional material from the KFCS text.

[1]D. R. Raichel, *The Science and Applications of Acoustics* (Springer, Ver-

lag, Berlin, 2000).
[2]D. R. Raichel, Author, U. J. Hansen, and Reviewer, "The science and applications of acoustics," J. Acoust. Soc. Am. **114**, 21–22 (2003).
[3]N. Shaw, "Review: Daniel R. Raichel: The science and applications of acoustics," J. Audio Eng. Soc. **48**, 990–989 (2000).
[4]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[5]C. E. Wilson, *Noise Control: Measurement, Analysis and Control of Sound and Vibration* (Krieger, Malabar, FL, 1994).
[6]H. F. Olson, *Music, Physics and Engineering* (Dover, New York, 1967).
[7]R. J. Urick, *Principles of Underwater Sound* (McGraw-Hill, New York, 1983).
[8]R. Clarke, "Plagiarism by academics: More complex than it seems," Journal of the Association for Information Systems **7**, 91–121 (2006).
[9]"Statement on Standards of Professional Conduct," (American History Association, http://www.historians.org/PUBS/Free/ProfessionalStandards.cfm (Last viewed 11/3/ 2008).

Propagation of nonlinear acoustic plane waves in an elastic gas-filled tube

Michal Bednarik and Milan Cervenka[a)]
Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague, Czech Republic

(Received 26 January 2009; revised 9 June 2009; accepted 17 July 2009)

This paper deals with modeling of nonlinear plane acoustic waves propagating through an elastic tube filled with thermoviscous gas. A description of the interactions between gas and an elastic tube wall is carried out by the continuity equation of a wall velocity. Simplification on the basis of the local reaction assumption enables to model an acoustic treatment on the tube wall by using a wall impedance. Because there are considerable losses due to wall friction, the influences of the acoustic boundary layer were also considered. Using certain assumptions a special form of the Burgers equation was derived which enables to describe the propagation of nonlinear waves in the elastic tube. This model equation takes into account nonlinear, dissipative, and dispersion effects which compete each other. Characteristic lengths of the supposed effects and numerical results with respect to the source frequency were used for a qualitative analysis of the model equation. Applicability of this model equation was demonstrated by series of measurements. By application of the long-wave approximation the Korteweg–de Vries–Burgers and Kuramoto–Sivashinsky equations were derived from the modified Burgers equation.
 [DOI: 10.1121/1.3203936]

PACS number(s): 43.25.Cb, 43.20.Mv [OAS] Pages: 1681–1689

I. INTRODUCTION

The problems concerning with the interactions between acoustic oscillations in a fluid-filled tube and the vibrations of its wall have received the attention of many investigators. Since time of Young who first found the pulse wave speed in human arteries, a number of scientists dealt with propagation of acoustic waves through elastic tubes filled with fluids. Their works differ each other by various assumptions, in particular, as far as media, wave modes, wave amplitudes, and tube walls are concerned.

Fay *et al.*[1] carried out an analytical and experimental investigation of a water-filled acoustic impedance tube. Their work was motivated by the fact that water-filled tube walls cannot be assumed to be rigid for acoustic waves. The paper by Jacobi[2] was also focused on the problems regarding the sound transmission through tubes filled with ideal liquid; however, in addition he considered the higher wave modes. In contrast to the above mentioned authors who considered only inviscid fluids, Morgan and Kiely[3] took into account also viscosity of the liquid and internal damping in the tube wall. Sondhi[4] for investigation of wave propagation in a lossy vocal tract used an approach which was based on a local wall admittance model. Guelke and Bunn[5] presented work which deals with application of the transmission line theory to linear acoustic wave propagation through tube with yielding walls. They limited themselves to vibrations only in the radial direction and considered only the breathing circumferential motion. In a similar spirit, Fredberg *et al.*[6] modeled mechanic oscillations of the respiratory system at high frequencies. Elvira-Segura[7] dealt with the study of speed and attenuation of an acoustic wave propagating inside a cylindrical elastic tube filled with a viscous liquid. This author extended influence of the liquid viscosity by the boundary layer effects. Because the flaws in the roundness of a tube induce coupling between the structural and acoustic modes which do not exist in the case where the cross-section is perfectly circular, Pico and Gautier[8] presented a model which allows us to take into account the small imperfections in the tube circularity. Gautier *et al.*[9] besides a well-arranged bibliographical review presented a study on cylindrical membranes submitted to a static tension. The authors of the above-cited papers supposed only the thin-walled tubes. Grosso[10] considered the exact longitudinal and shear wave equations for the multimode axial acoustic propagation in tubes of arbitrary wall thickness filled with inviscid liquid. Nonlinear effects were taken into account, e.g., by Yomosa[11] who described the propagation of weakly nonlinear waves in an infinitely long distensible thin-walled elastic tube filled with an ideal fluid. Erbay and Dost[12] investigated the propagation of weakly nonlinear waves in an infinitely long nonlinear viscoelastic thin tube filled with incompressible, inviscid fluid. By means of the long-wave approximation they derived a number of nonlinear evolution equations representing various regimes. Kamakura and Kumamoto[13] presented a work which concerned with investigation of nonlinear plane acoustic waves through an elastic tube. The authors assumed that an elastic tube wall reacts locally to the inner pressure. On the basis of their assumptions they derived a model equation. Validation of the model equation was justified experimentally. Unfortunately, the comparison of theoretical and experimental results was not presented in a way enabling to evaluate measure of the validation of their model equation

[a)]Author to whom correspondence should be addressed. Electronic mail: bednarik@fel.cvut.cz

rigorously. Though the problems concerning propagation of acoustic waves through tubes were widely studied by many authors, investigation seems to be missing which would simultaneously include most of effects that influence acoustic waves. A model equation, which takes into account more effects at the same time, can offer qualitatively new results. This fact determines the main object of this work.

Hence this work is focused on the description of nonlinear plane acoustic waves propagating through an elastic tube filled with thermoviscous gas. For this purpose, a special form of the Burgers equation was derived on the basis of the local wall reaction hypothesis. This model equation extends the standard Burgers equation by terms which represent dispersion and dissipative effects caused by the wall elasticity and acoustic (Stokes) boundary layer which plays an important role in the course of wave forming, in particular, for lower frequencies. Theoretical results were verified experimentally. The Korteweg–de Vries–Burgers (KdVB) and Kuramoto–Sivashinsky equations are derived from the modified Burgers equation by means of the long-wave approximation. These equations are supplemented by the term which takes into account boundary layer effects.

Section II is dedicated to derivation of the model equations. Analysis of numerical solutions of the modified Burgers equation is presented in Sec. III. Then in Sec. IV we compare theoretical and experimental data to justify applicability of the model equations.

II. MODEL EQUATIONS AND DISPERSION RELATIONS

A. Derivation of the modified Burgers equation for a gas-filled elastic tube

If we take into account the acoustic boundary layer effects we can write the following one-dimensional continuity equation (see Ref. 14):

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2}{r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\rho\int_0^\infty \frac{\partial v(x,t-\chi)}{\partial x}\frac{d\chi}{\sqrt{\chi}}. \tag{1}$$

Here x is space coordinate in the direction of the tube axis, t is time, $\rho=\rho'+\rho_0$ is density of fluid, where ρ' is the acoustic density and ρ_0 is density corresponding to the equilibrium fluid state, v is the acoustic velocity, ν is the kinematic viscosity, $\mathrm{Pr}=\rho_0\nu c_p/\varkappa$ is the Prandtl number, $\gamma=c_p/c_V$ is the ratio of specific heats (c_p and c_V are the specific heats under constant pressure and volume, respectively), $\varkappa$ is the coefficient of heat conductivity, and r_0 is an equilibrium tube inner radius.

Using the continuity equation (1) is conditioned by the following relations:[15,16]

$$\delta \ll \lambda, \quad \delta \ll r_0,$$

where δ is a boundary layer thickness and λ is a wavelength.

Further, when it is satisfied

$$\frac{2}{r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\sim \mu,$$

where $\mu \ll 1$ is a small dimensionless parameter (the peak Mach number of the source), then within the scope of the second order nonlinear theory we can replace the density $\rho=\rho'+\rho_0$ by an ambient fluid density ρ_0 in Eq. (1). After the replacement and using the relation between the acoustic velocity and the velocity potential $v=\partial\varphi/\partial x$ Eq. (1) can be written in the form

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2\rho_0}{r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\int_0^\infty \frac{\partial^2\varphi(x,t-\chi)}{\partial x^2}\frac{d\chi}{\sqrt{\chi}}$$
$$=\frac{2\rho_0}{r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\frac{\partial^2}{\partial x^2}\int_0^\infty \varphi(x,t-\chi)\frac{d\chi}{\sqrt{\chi}}. \tag{2}$$

With help of the linear one-dimensional wave equation

$$\frac{\partial^2\varphi}{\partial x^2}=\frac{1}{c_0^2}\frac{\partial^2\varphi}{\partial t^2},$$

it is possible to rewrite Eq. (2)

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2\rho_0}{c_0^2 r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\frac{\partial^2}{\partial t^2}\int_0^\infty \varphi(x,t-\chi)\frac{d\chi}{\sqrt{\chi}}$$
$$=\frac{2\rho_0}{c_0^2 r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\int_0^\infty \frac{\partial^2\varphi(x,t-\chi)}{\partial t^2}\frac{d\chi}{\sqrt{\chi}}. \tag{3}$$

The equality $\partial^2\varphi/\partial t^2=\partial^2\varphi/\partial\chi^2$ enables us to express Eq. (3) as

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2\rho_0}{c_0^2 r_0}\sqrt{\frac{\nu}{\pi}}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right)\int_0^\infty \frac{\partial^2\varphi(x,t-\chi)}{\partial\chi^2}\frac{d\chi}{\sqrt{\chi}}. \tag{4}$$

The integral on the right hand side (rhs) of Eq. (4) can be rewritten as

$$\frac{1}{\sqrt{\pi}}\int_0^\infty \frac{\partial^2\varphi(x,t-\chi)}{\partial\chi^2}\frac{d\chi}{\sqrt{\chi}}=\frac{1}{\sqrt{\pi}}\int_{-\infty}^t \frac{\partial^2\varphi(x,\chi)}{\partial\chi^2}\frac{d\chi}{\sqrt{t-\chi}}. \tag{5}$$

The rhs of Eq. (5) represents the fractional derivative (A3) (see, e.g., Refs. 17, 15, and 16). We can express Eq. (3) as follows:

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2\rho_0 B}{c_0^2}\frac{\partial^{3/2}\varphi}{\partial t^{3/2}}, \tag{6}$$

where

$$B=\frac{\sqrt{\nu}}{r_0}\left(1+\frac{\gamma-1}{\sqrt{\mathrm{Pr}}}\right) \tag{7}$$

is the boundary layer parameter.

If we suppose that acoustic waves propagate through a tube with an elastic wall (a variable cross-section) and we do not take into account the boundary layer effects then we can express the continuity equation as

$$\frac{\partial(\rho S)}{\partial t}+\frac{\partial(\rho S v)}{\partial x}=0, \tag{8}$$

where $S=S(x,t)$ is an inner tube cross-section. Further, Eq. (8) can be modified into the form

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}+\frac{\rho}{S}\frac{dS}{dt}=0, \tag{9}$$

where the operator d/dt is given as

$$\frac{d}{dt}=\frac{\partial}{\partial t}+v\frac{\partial}{\partial x}. \tag{10}$$

If the elastic tube wall is assumed to be locally reacting (see, e.g., Refs. 18 and 5) then we can consider the inner cross-section S only as a function of time, i.e., $S=S(t)=\pi r^2(t)$, where $r(t)=r_0+r'(t)$ is a total inner tube radius and r' represents a change in r.

If we use the fact that $r' \ll r_0$, then it is possible to use the following simplification:

$$\frac{1}{S}\frac{dS}{dt}\simeq\frac{1}{S_0}\frac{dS}{dt}=\frac{1}{\pi r_0^2}\frac{d(\pi r^2)}{dt}=\frac{2}{r_0}\frac{dr}{dt}=\frac{2}{r_0}v_w, \tag{11}$$

where v_w is a radial wall velocity.

Because within the scope of the second order nonlinear theory we neglect terms which are of the third order or higher, it is possible to simplify Eq. (9) by using relation (11) and adopting the supposition $v_w \sim \mu^2$,

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=-\frac{2\rho_0}{r_0}v_w. \tag{12}$$

Using the linear relation

$$\frac{\partial p}{\partial t}=-\rho_0\frac{\partial^2\varphi}{\partial t^2}, \tag{13}$$

we can write the following convolution integral for the radial wall velocity v_w:

$$v_w=\int_{-\infty}^{t}k_w(t-t')\frac{\partial p(x,t')}{\partial t'}dt'$$

$$=-\rho_0\int_{-\infty}^{t}k_w(t-t')\frac{\partial^2\varphi(x,t')}{\partial t'^2}dt', \tag{14}$$

where $k_w(t)$ is a kernel function representing behavior of the considered tube wall.

If we take into account both the acoustic boundary layer [Eq. (6)] and the tube elasticity [Eq. (12)] together with relation (14) we obtain the resulting continuity equation

$$\frac{\partial \rho}{\partial t}+\frac{\partial(\rho v)}{\partial x}=\frac{2\rho_0^2}{r_0}\int_{-\infty}^{t}k_w(t-t')\frac{\partial^2\varphi}{\partial t'^2}dt'+\frac{2\rho_0 B}{c_0^2}\frac{\partial^{3/2}\varphi}{\partial t^{3/2}}. \tag{15}$$

Using the basic equations of hydromechanics, namely, the Navier–Stokes equation of motion, the heat transfer equation, the state equation (see, e.g., Refs. 19–21), and the modified continuity equation (15) we can derive in the second approximation the following one-dimensional modified Kuznetsov's equation[22] for the velocity potential φ:

$$\frac{\partial^2\varphi}{\partial t^2}-c_0^2\frac{\partial^2\varphi}{\partial x^2}=\frac{b}{\rho_0 c_0^2}\frac{\partial^3\varphi}{\partial t^3}-\frac{\partial}{\partial t}\left[\frac{\beta-1}{c_0^2}\left(\frac{\partial\varphi}{\partial t}\right)^2+\left(\frac{\partial\varphi}{\partial x}\right)^2\right]$$

$$-\frac{2\rho_0 c_0^2}{r_0}\int_{-\infty}^{t}k_w(t-t')\frac{\partial^2\varphi}{\partial t'^2}dt'-2B\frac{\partial^{3/2}\varphi}{\partial t^{3/2}}, \tag{16}$$

where c_0 is the small-signal sound speed, $b=\zeta+4\eta/3+\varkappa(1/c_V-1/c_p)$ is the coefficient of sound diffusivity, ζ and η are the coefficients of bulk and shear viscosity, and β is the coefficient of nonlinearity.

As the profile of simple waves varies slowly in the space we can search for the solution of Eq. (16) in the moving reference frame in the form

$$\varphi=\varphi\left(x_1=\mu x,\tau=t-\frac{x}{c_0}\right). \tag{17}$$

Then taking into consideration that $b\sim B\sim\mu$ we can derive the following modified Burgers equation from Eq. (16) within frame of the second order nonlinear theory:

$$\frac{\partial v}{\partial x}-\frac{\beta}{c_0^2}v\frac{\partial v}{\partial\tau}+\frac{\rho_0 c_0}{r_0}\int_{-\infty}^{\tau}k_w(\tau-\tau')\frac{\partial v}{\partial\tau'}d\tau'+\frac{B}{c_0}\frac{\partial^{1/2}v}{\partial\tau^{1/2}}$$

$$-\frac{b}{2\rho_0 c_0^3}\frac{\partial^2 v}{\partial\tau^2}=0. \tag{18}$$

It is obvious that it is possible to use the model [Eq. (18)] also for the nonlinear plane waves which propagate through a hard-walled tube. In this case we can consider the term representing the elastic properties of the tube wall as equal to zero.[15]

B. Derivation of dispersion relations

Under the above mentioned restrictions, the tube can be regarded as consisting simply of a series of ring-shaped elements whose radial motion is caused only by elastic circumferential stresses and radial inertia. As the tube wall yields locally to the inner pressure we can write the following expression for the force acting on the ring:

$$F(t)=p(t)A(t)=p(t)[A_0+A'(t)]=p(t)2\pi r(t)\Delta l$$

$$=p(t)[2\pi r_0\Delta l+2\pi r'(t)\Delta l], \tag{19}$$

where p is an inner pressure and Δl is the width of the ring element. If we suppose that r' is sufficiently small compared to r_0 then we can simplify relation (19),

$$F(t)\simeq p(t)A_0. \tag{20}$$

With using the complex representation we can rewrite Eq. (20) as

$$\hat{F}\simeq\hat{p}A_0. \tag{21}$$

For the ring element we can use the equivalent electromechanic circuit which is sketched in Fig. 1. We suppose that air surrounds the ring and consequently the radiation load on the outside of the ring can be neglected.[23]

On the basis of the equivalent circuit we obtain

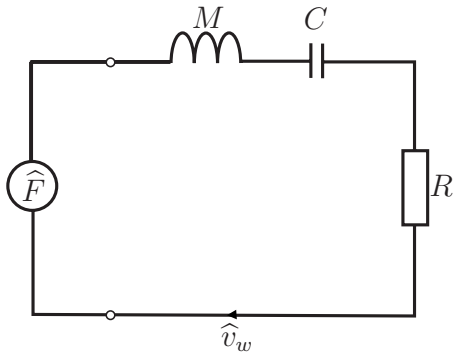


FIG. 1. The equivalent electromechanic circuit, M is the inertance, C is the compliance, R is mechanical resistance, and $\hat{v}_w$ is radial wall velocity.

$$\hat{F} = \left(R + j\omega M + \frac{1}{j\omega C}\right)\hat{v}_w = \hat{Z}_w \hat{v}_w, \tag{22}$$

where M is the inertance, C is the compliance, R is mechanical resistance, and $\hat{Z}_w$ is the resulting wall mechanical impedance.

Using relations (21) and (22) we can write

$$\hat{v}_w \simeq \frac{A_0}{\hat{Z}_w}\hat{p} = \frac{2\pi r_0 \Delta l}{\hat{Z}_w}\hat{p}. \tag{23}$$

With the help of the convolution integral (14) we can also write that

$$\hat{v}_w = j\omega \hat{k}_w \hat{p}. \tag{24}$$

By comparing Eqs. (23) and (24) we obtain

$$\hat{k}_w = \frac{2\pi r_0}{j\omega \hat{Z}_m}, \tag{25}$$

where

$$\hat{Z}_m = \frac{\hat{Z}_w}{\Delta l} = R_m + j\omega M_m + \frac{1}{j\omega C_m}, \tag{26}$$

where $R_m = R/\Delta l$ is the specific mechanical resistance (kg m^{-1} s^{-1}), $M_m = M/\Delta l$ is the specific inertance (kg m^{-1}), and $C_m = C\Delta l$ (kg^{-1} m s^2) is the specific compliance.

With help of expressions (25) and (26) it is possible to write (see Ref. 13)

$$\frac{2\hat{k}_w}{r_0} = \frac{4\pi}{j\omega \hat{Z}_m} = \frac{4\pi C_m}{1 - \left(\frac{\omega}{\omega_m}\right)^2 + j\omega R_m C_m}, \tag{27}$$

where ω_m represents the mechanical resonance angular frequency which is given by the following relation:

$$\omega_m = \frac{1}{\sqrt{C_m M_m}}. \tag{28}$$

After linearizing modified Kuznetsov's equation

$$\frac{\partial^2 \varphi}{\partial t^2} - c_0^2 \frac{\partial^2 \varphi}{\partial x^2} - \frac{b}{\rho_0 c_0^2}\frac{\partial^3 \varphi}{\partial t^3} + \frac{2\rho_0 c_0^2}{r_0}\int_{-\infty}^{t} k_w(t-t')\frac{\partial^2 \varphi}{\partial t'^2}dt' + 2B\frac{\partial^{3/2}\varphi}{\partial t^{3/2}} = 0, \tag{29}$$

we can find the following dispersion relation:

$$k = \frac{\omega}{c_0}\sqrt{1 - j\frac{\omega b}{\rho_0 c_0^2} + \frac{2\rho_0 c_0^2 \hat{k}_w}{r_0} - \frac{2B(j\omega)^{3/2}}{\omega^2}}, \tag{30}$$

where k is the complex wave number.

Expression (30) can be simplified on the basis of the binomial series

$$k \simeq \frac{\omega}{c_0}\left(1 - j\frac{\omega b}{2\rho_0 c_0^2} + \frac{\rho_0 c_0^2 \hat{k}_w}{r_0} + \frac{B}{\sqrt{j\omega}}\right). \tag{31}$$

Since $\hat{k}_w/r_0$ is complex, see relation (27), we can write

$$\frac{\hat{k}_w}{r_0} = \Re\left(\frac{\hat{k}_w}{r_0}\right) + j\Im\left(\frac{\hat{k}_w}{r_0}\right) = \frac{2\pi C_m\left[1 - \left(\frac{\omega}{\omega_m}\right)^2\right]}{\left[1 - \left(\frac{\omega}{\omega_m}\right)^2\right]^2 + R_m^2 C_m^2 \omega^2} - j\frac{2\pi R_m C_m^2 \omega}{\left[1 - \left(\frac{\omega}{\omega_m}\right)^2\right]^2 + R_m^2 C_m^2 \omega^2}. \tag{32}$$

It is possible to simplify formula (32) when the following relations are satisfied:[13]

$$\omega \ll \omega_m, \quad \omega R_m C_m \ll 1. \tag{33}$$

We can write that

$$\frac{\hat{k}_w}{r_0} \simeq 2\pi C_m\left[1 + \left(\frac{\omega}{\omega_m}\right)^2\right] - j2\pi R_m C_m^2 \omega\left[1 + 2\left(\frac{\omega}{\omega_m}\right)^2\right]. \tag{34}$$

Expression (34) is valid only for lower frequencies; it means that the dispersion and dissipation effects dominate above the nonlinear ones. In the opposite case the higher harmonic components, which arise in the course of the wave propagation, do not satisfy conditions (33). When we can take into account only small number of harmonics then it is further possible to simplify relation (34) as follows:

$$\frac{\hat{k}_w}{r_0} \simeq 2\pi C_m\left[1 + \left(\frac{\omega}{\omega_m}\right)^2\right] - j2\pi R_m C_m^2 \omega. \tag{35}$$

C. The KdVB and Kuramoto–Sivashinsky model equation for an elastic tube

We can rewrite Eq. (18) into the form[24]

$$\frac{\partial v}{\partial x} - \frac{\beta}{c_0^2} v\frac{\partial v}{\partial \tau} + \int_{-\infty}^{\infty} K(\tau - \tau')\frac{\partial v(x,\tau')}{\partial \tau'}d\tau' = 0, \tag{36}$$

where $K(\tau)$ represents a kernel function which describes assumed dispersion and dissipation properties. Suppose we know the dispersion relation

TABLE I. Real parameters for air and the used elastic tube (23 °C).

c_0 (m s^{-1})	ρ_0 (kg m^{-3})	b (kg m^{-1} s^{-1})	B (s$^{-1/2}$)	C_m (m s^2 kg^{-1})	R_m (kg m^{-1} s^{-1})	M_m (kg m^{-1})
345.22	1.193	4.578×10^{-5}	0.7227	4.45×10^{-8}	154.0	0.118

$$D(\omega^*,k')=0, \tag{37}$$

which corresponds to the kernel function $K(\tau)$, $\omega^*=\omega$ and

$$k'=k-\omega/c_0=\frac{\omega}{c_0}\left(\frac{\rho_0 c_0^2\hat{k}_w}{r_0}+\frac{B}{\sqrt{j\omega}}-j\frac{\omega b}{2\rho_0 c_0^2}\right) \tag{38}$$

is a complex wave number that is related to the independent variables τ, x in contrast to the complex wave number (31) that is connected with the independent variables t and x. Provided that certain conditions, which are discussed later, are fulfilled then we can simplify the dispersion relation (37) into the asymptotic form

$$D_a(\omega,k')=0. \tag{39}$$

After linearizing Eq. (36) we can write[25]

$$\frac{\partial v}{\partial x}+\int_{-\infty}^{\infty}K_a(\tau-\tau')\frac{\partial v(x,\tau')}{\partial\tau'}d\tau'=0, \tag{40}$$

where the kernel function $K_a(\tau)$ corresponds to asymptotic dispersion relation (39). We assume the solution of Eq. (40) in the form $v=v_m\exp[j(\omega\tau-k'x)]$ (v_m is the acoustic velocity amplitude). We can substitute this solution into Eq. (40) and we obtain

$$-jk'+\int_{-\infty}^{\infty}K_a(\tau-\tau')j\omega\exp[-j\omega(\tau-\tau')]d\tau'=0. \tag{41}$$

Equation (41) can be rewritten into the form

$$\frac{k'}{\omega}=\mathcal{F}[K_a(\tau)]=\int_{-\infty}^{\infty}K_a(\tau)\exp(-j\omega\tau)d\tau, \tag{42}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform. The rhs of Eq. (42) is the Fourier transform of the given kernel $K_a(\tau)$. On the basis of the inverse Fourier transform we have

$$K_a(\tau)=\mathcal{F}^{-1}\left[\frac{k'}{\omega}\right]=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{k'}{\omega}\exp(j\omega\tau)d\omega. \tag{43}$$

Substituting expression (34) into relation (38) we obtain

$$k'\simeq\omega\left\{2\pi\rho_0 c_0 C_m\left[1+\left(\frac{\omega}{\omega_m}\right)^2\right]-j2\pi\rho_0 c_0 R_m C_m^2\omega\left[1+2\left(\frac{\omega}{\omega_m}\right)^2\right]+\frac{B}{c_0\sqrt{j\omega}}-j\frac{\omega b}{2\rho_0 c_0^3}\right\}. \tag{44}$$

On the basis of relation (43) we can find the asymptotic kernel function $K_a(\tau)$ by using formula (44). After substitution of this asymptotic kernel function into Eq. (36) we get

$$\frac{\partial v}{\partial x}-\frac{\beta}{c_0^2}\left(v-\frac{2\pi\rho_0 c_0^3 C_m}{\beta}\right)\frac{\partial v}{\partial\tau}+\frac{B}{c_0}\frac{\partial^{1/2}v}{\partial\tau^{1/2}}-\frac{b+4\pi\rho_0^2 C_m^2 R_m c_0^4}{2\rho_0 c_0^3}\frac{\partial^2 v}{\partial\tau^2}-\frac{2\pi\rho_0 c_0 R_m C_m}{\omega_m^2}\frac{\partial^3 v}{\partial\tau^3}+\frac{4\pi\rho_0 c_0 R_m C_m^2}{\omega_m^2}\frac{\partial^4 v}{\partial\tau^4}=0. \tag{45}$$

Equation (45) represents the Kuramoto–Sivashinsky (Benney) equation[26,27] which is modified by the term representing the boundary layer effects. This equation can be simplified into the KdVB equation (see, e.g., Ref. 28) by setting the last term equal to zero when propagation wave distances are relatively short or the source frequency is sufficiently low and the dispersive and dissipative effects are insofar important that it is possible to take into account a reasonably small number of harmonics.

III. ANALYSIS OF NUMERICAL SOLUTIONS

This section is devoted to qualitative analysis of numerical results of the model equation (18) for a silicon rubber tube (hose). This analysis enables to show different scenarios of possible wave evolution for various source frequencies and hence demonstrates how the individual terms in the model equation depend on the source frequency. The real physical parameters are presented in Table I and for these parameters the analysis of numerical solutions is carried out. In order to show frequency limits of the KdVB and Kuramoto–Sivashinsky equation we can depict the courses of function $\hat{k}_w/r_0$ in dependence on frequency according to relations (32), (34), and (35), see Figs. 2 and 3. It is evident from Figs. 2 and 3 that in this case the Kuramoto–

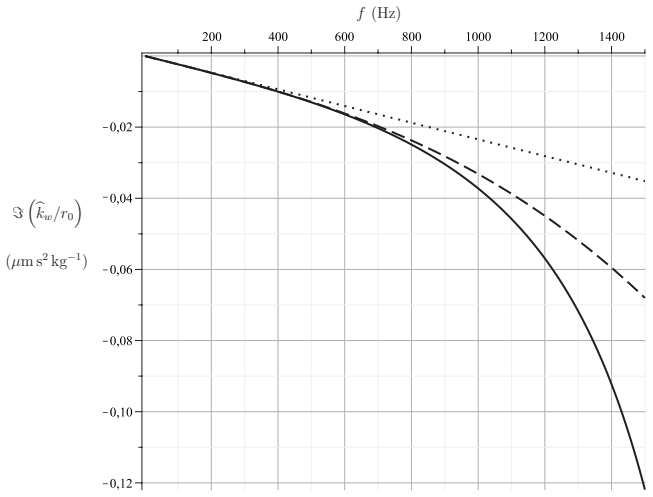


FIG. 2. Comparison of imaginary parts of relations (32) (solid line), (34) (dashed line), and (35) (dotted line).

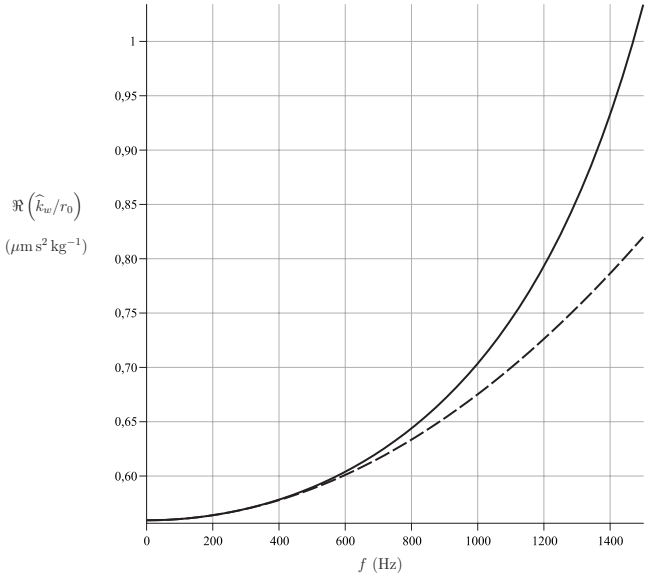


FIG. 3. Comparison of real parts of relations (32) (solid line) and (34) (dashed line).

Sivashinsky equation can be used in the frequency range approximately up to 1000 Hz, whereas the KdVB equation only up to about 800 Hz.

The model equation (18) takes into account three effects, i.e., the nonlinear, dissipative, and dispersion ones. There is a competition between the nonlinear effect, on the one hand, and the dispersion and dissipative effect on the other hand. In this case we can estimate a wave behavior on the basis of the characteristic lengths (see, e.g., Ref. 19). The nonlinear length is given as

$$L_N = \frac{c_0^2}{\beta v_m \omega}, \tag{46}$$

the dissipative length can be expressed as

$$L_D = \frac{1}{\alpha}, \tag{47}$$

and the dispersion or coherent length for the second harmonic component as

$$L_C = \frac{\pi}{\left|2\omega\left[\frac{1}{c_{\mathrm{ph}}(2\omega)} - \frac{1}{c_{\mathrm{ph}}(\omega)}\right]\right|}. \tag{48}$$

It holds that the shorter characteristic length means the greater influence of the considered effect (in our case the diffraction length tends to infinity because we suppose that plane waves propagate in the tube). On the basis of the characteristic lengths we can outline several scenarios of possible wave evolutions with respect to a chosen source frequency (amplitude of the source is supposed to be constant, 4000 Pa).

First, we suppose that the source frequency is chosen so that $L_N/L_C > 1$ and $L_N/L_D < 1$. It means that the dispersion effects dominate the nonlinear ones and at the same time the nonlinear effects dominate the dissipative ones. For instance, the source frequency of 1000 Hz satisfies (if we suppose the physical parameters in Table I) the above mentioned relations. For this case the wave distortion is negligible (the first harmonic component dominates) which affirms that dispersion effects dominate nonlinear ones and thus nonlinear acoustic interactions are ineffective. Thus, we could ignore the nonlinear effects and use only the linearized form of Eq. (18). When the supposed relations are satisfied also for lower source frequencies, in our case for a smaller source amplitude, it is possible to use the linearized Kuramoto–Sivashinsky or KdVB equation.

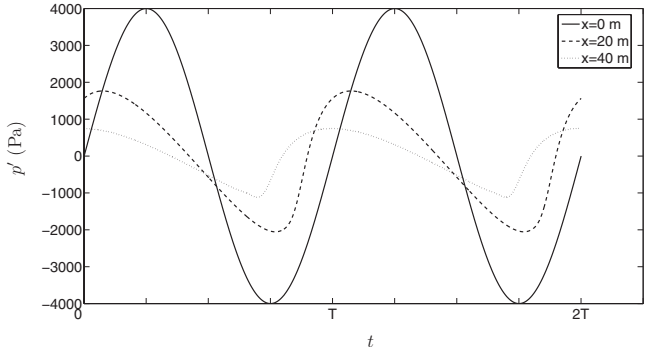


FIG. 4. Two periods of wave forms at different distances from the source; the source frequency is 80 Hz.

Further, we assume that the relations $L_N/L_C < 1$ and $L_N/L_D < 1$ are satisfied for at least the first ten harmonics. For this purpose we can choose the source frequency, e.g., 80 Hz. The satisfaction of these relations means that a sawtooth wave forms during propagation. The boundary layer dispersion causes the waveform asymmetry which occurs behind the shock formation. This result is demonstrated in Fig. 4. Furthermore, when the relations are satisfied then it is possible to replace Eq. (18) by the standard Burgers equation which is supplemented by the term representing the boundary layer effects.[15,16]

As the next case, we assume that the relations $L_N/L_C < 1$ and $L_N/L_D < 1$ are approximately satisfied only for the first two or three harmonics; hence, the generation of higher harmonics is suppressed. Figure 5 illustrates this situation.

When the boundary layer effects are small and again the relations $L_N/L_C < 1$ and $L_N/L_D < 1$ are satisfied, we can observe a gradual degeneration of the original waveform into individual solitons.[24,28,29] The solutions for this case are depicted in Figs. 6 and 7. In the figures we can see a train of the

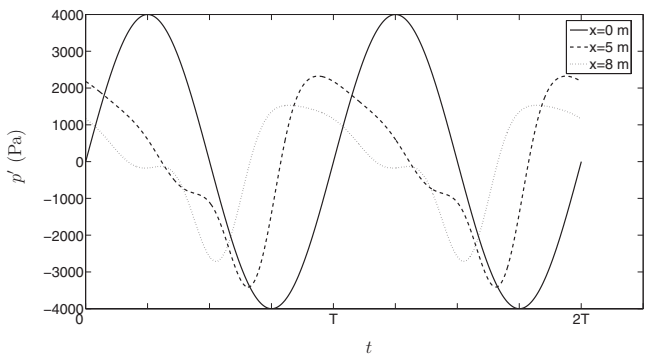


FIG. 5. Two periods of wave forms at different distances from the source; the source frequency is 400 Hz.

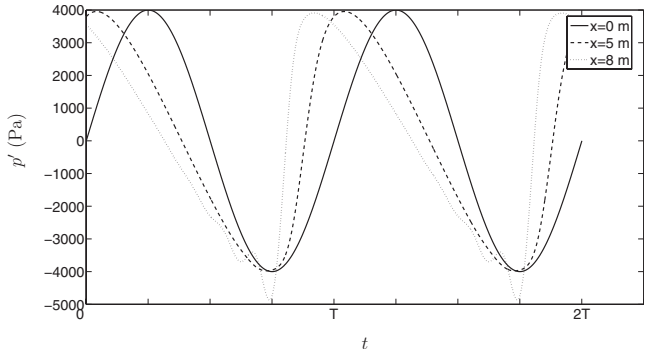


FIG. 6. Two periods of wave forms at different distances from the source; the source frequency is 200 Hz.

solitons which are gradually attenuated as the wave propagates through the considered tube. First, we can observe slight oscillations, see Fig. 6. This fact is related to the harmonics generation because dispersion effect starts to play a more important role for higher harmonics. Gradually the slight oscillations develop into a series of solitons which break up the previous waveform, see Fig. 7.

As the last case, we can consider the relations $L_N/L_C \ll 1$ and $L_N/L_D < 1$. It is obvious that $\hat{k}_w/r_0$, see formula (27), tends to zero for high source frequencies. It means that we can ignore the radial wall vibrations for high source frequencies. The considered relations for the characteristic lengths are satisfied for higher source frequencies.

IV. EXPERIMENT

To validate applicability of the model equations, in particular, the local reaction hypothesis, we compared the theoretical and experimental data.

A. Experimental setup

Traveling sound waves were driven in silicone-rubber hose (with inner diameter of 16 mm, wall thickness of 2 mm), length of the hose was 20 m, it was lengthened by 20 m of polyvinyl chloride hose of the same inner diameter in order to suppress standing waves due to reflections. All measurements were made at room temperature of 23 °C.

Harmonic driving signal was generated with direct-digital-synthesis function generator Motech FG 503 which was amplified by power amplifier Mackie M1400. High amplitude acoustic waves were generated by two speakers B&C 10MD26 (each of 350 W input) that were screwed one against another with a plastic ring for fastening of the hose. At the place where the hose was fastened to the plastic ring, reference 1/8 in. GRAS Type 40DP microphone was attached to measure the input acoustic pressure. In the hose, there were 13 small cuts distributed with 1 m distances for measurement of acoustic pressure with probe-microphone GRAS Type 40SA with stainless-steel 20 mm probe tube of 1.25 mm outer diameter (inner diameter of 1 mm).

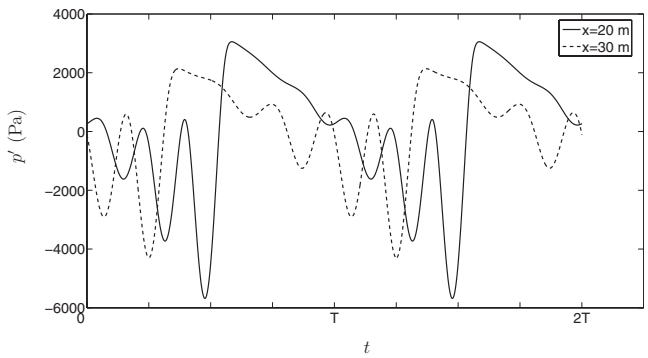


FIG. 7. Two periods of two wave forms at different distances from the source; the source frequency is 200 Hz.

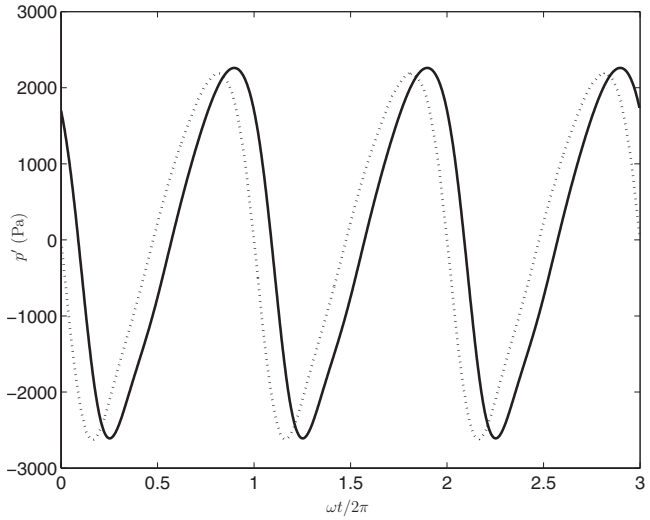


FIG. 8. Comparison of theoretical (solid line) and measured (dotted line) wave forms at the distance $x=3$ m, $p_m=3602$ Pa; the driving frequency is 400 Hz.

Measured signals were sampled by 16 bit National Instruments PCI-6251 plug-in computer data acquisition card at the rate of 100 kS/s per channel. Software for data acquisition and its processing was written in LABVIEW system.

On the basis of measurement of the silicone-rubber hose we found its material parameters which are shown in Table I.

In the course of measuring it was necessary to take into account the fact that the driving speakers heated up excessively. For this reason it was not possible to realize measuring with the same driving pressure at all points; hence the driving pressures were registered for individual measurements apart. The value of the driving pressure was around 3600 Pa ($v_m=8.74$ m s^{-1}). We confined ourselves to the relatively low source frequency in order that the dispersion and nonlinear effects could dominate above the dissipative ones.

B. Comparison of theoretical and experimental results

The measured and theoretical (computed numerically) data are depicted in Figs. 8–10. By comparison of measured and theoretical waveforms at different distances from the harmonic source it is clear that the model equation (18) enables to describe behavior of nonlinear traveling waves in the elastic tube and that the local reaction hypothesis is sufficient when the conditions mentioned in Sec. II are fulfilled. From these figures it is obvious that the shock waveform did not arise. This result is given by the fact that the relations $L_N/L_C < 1$ and $L_N/L_D < 1$ are fulfilled approximately only

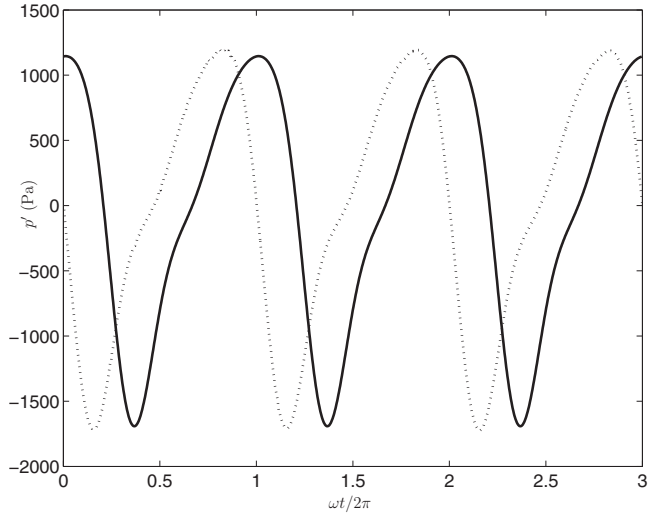


FIG. 9. Comparison of theoretical (solid line) and measured (dotted line) wave forms at the distance $x=7$ m, $p_m=3604$ Pa; the driving frequency is 400 Hz.

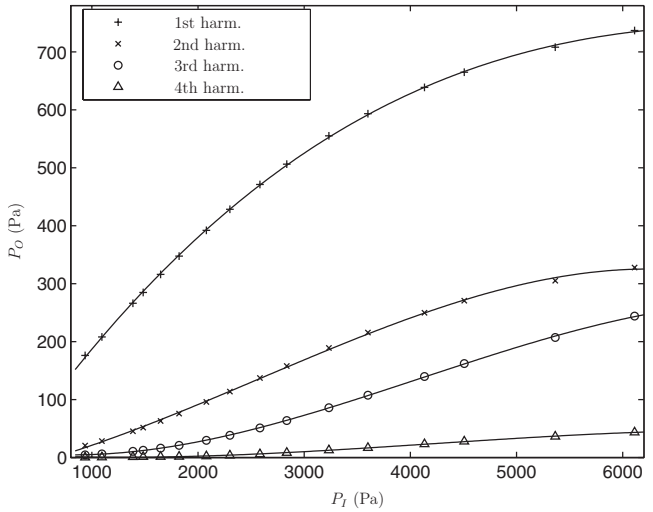


FIG. 11. Dependence of amplitude of the first four harmonics at distance of 13 m on input pressure amplitude.

for the first three harmonics; thus the generation of higher harmonics is suppressed. It means that the relation $L_N/L_C > 1$ holds for higher harmonics.

Figure 11 shows development of the first four harmonics of acoustic pressure at the distance of 13 m when input acoustic pressure ($f=333$ Hz) increases. It is apparent from the figure that even if dispersion is present, several higher harmonics appear and the amplitude of the first and second harmonics is not proportional to the pressure amplitude at the input. This effect is called *acoustic saturation* and it has connection with the nonlinear attenuation. It is obvious from Fig. 11 that the fourth harmonics and the third one depend on the pressure amplitude at the input almost linearly. It means that the cascade process of harmonics generation is interrupted or in other words, the characteristic coherent length is shorter than the nonlinear one ($L_N > L_C$) for higher harmonics in contrast to the lower ones.

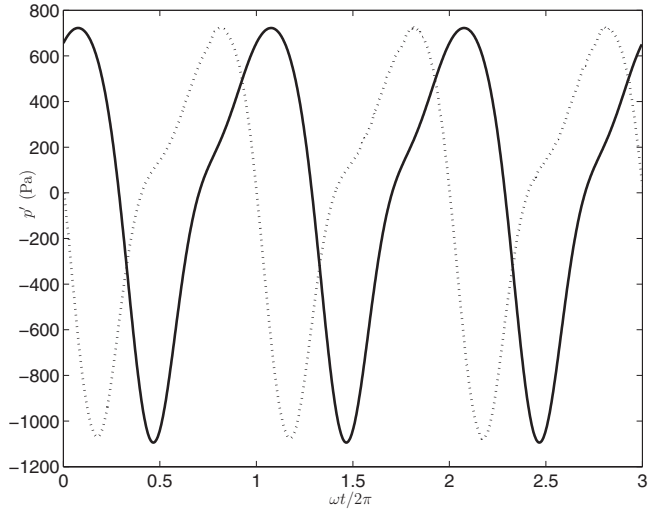


FIG. 10. Comparison of theoretical (solid line) and measured (dotted line) wave forms at the distance $x=10$ m, $p_m=3597$ Pa; the driving frequency is 400 Hz.

V. CONCLUSION

We derived the modified Burgers equation (18) which enables to describe nonlinear acoustic plane waves propagating through the elastic tube filled with a thermoviscous gas. The model equation takes into account only the breathing circumferential damped vibrations of the considered tube wall. Since the viscosity of gas is considered, it is necessary to suppose that the thin acoustic boundary layer appears near the tube wall. The acoustic boundary layer affects the acoustic field outside of this layer (the mainstream) in a considerable way. For this reason the derived model equation also contains the term which represents the boundary layer effects. The applicability of the modified Burgers equation was verified experimentally. The realized measurements demonstrate the fact that the dispersion induced by the elastic tube wall causes nonlinear acoustic interactions which are almost ineffective and further that it is not possible to ignore the influences of the acoustic boundary layer. It means that the shock formation does not arise and so we can observe only a small number of harmonics. Because we consider nonlinear, dispersion, and dissipative wave effects, that compete each other, it is possible to distinguish different regimes. These regimes are classified by means of the relations between characteristic lengths of the supposed wave effects. The supposed regimes lead to different wave evolutions. In the case when the nonlinear effects dominate the dispersion and dissipative one then the original waveform degenerates gradually into individual solitons. However, we can observe the solitons only when the acoustic boundary layer effects are weak. Contrary if it is not possible to suppose that these effects are weak (e.g., in the case of a small tube radius) then the generation of solitons is suppressed. This fact shows how the boundary layer plays important role.

Using the long-wave approximation, the modified Burgers equation was reduced to the KdVB and Kuramoto–Sivashinsky equations, which were investigated by many authors; however, they have not studied influence of the boundary layer effects. Considering that acoustic boundary

layer plays more important role for lower frequencies, we cannot neglect its influence in the case of long-wave approximation.

Considering that the KdVB equation can be used only for a relatively narrow frequency range we derived the Kuramoto–Sivashinsky equation which can be used for a wider frequency range and thus it can take into account more higher harmonics correctly.

We intend to extend further the presented results and methods for the case of nonlinear standing waves in elastic resonators.

ACKNOWLEDGMENT

This work was supported by GACR Grant No. 202/09/1509.

APPENDIX: FRACTIONAL DERIVATIVE OPERATOR

Let us introduce the following fractional integration operator of the order α:

$$I^{\alpha}[f(t)] = \int_{-\infty}^{t} \frac{(t-\chi)^{\alpha-1}}{\Gamma(\alpha)} f(\chi) d\chi, \tag{A1}$$

where Γ is the gamma function. The αth fractional derivative is given as

$$\frac{d^{\alpha} f(t)}{dt^{\alpha}} = I^{n-\alpha}\left[\frac{d^{n} f(t)}{dt^{n}}\right], \tag{A2}$$

where $n=1+[\alpha]$, where $[\alpha]$ represents the whole part of α.

On the basis of the mentioned definition of the fractional derivative we can write for the fractional derivative of the order $\alpha=3/2$ that

$$\frac{d^{3/2} f(t)}{dt^{3/2}} = I^{2-3/2}\left[\frac{d^{2} f(t)}{dt^{2}}\right] = I^{1/2}\left[\frac{d^{2} f(t)}{dt^{2}}\right]$$

$$= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{t} \frac{d^{2} f(\chi)}{d\chi^{2}} \frac{d\chi}{\sqrt{t-\chi}}, \tag{A3}$$

where we used that $\Gamma(1/2)=\sqrt{\pi}$.

[1]R. D. Fay, R. L. Brown, and O. V. Fortier, "Measurement of acoustic impedances of surfaces in water," J. Acoust. Soc. Am. **19**, 850–856 (1947).
[2]W. J. Jacobi, "Propagation of sound waves along liquid cylinders," J. Acoust. Soc. Am. **21**, 120–127 (1949).
[3]G. W. Morgan and J. P. Kiely, "Wave propagation in a viscous liquid contained in flexible tube," J. Acoust. Soc. Am. **26**, 323–328 (1954).
[4]M. M. Sondhi, "Model for wave propagation in a lossy vocal tract," J. Acoust. Soc. Am. **55**, 1070–1075 (1974).
[5]R. W. Guelke and A. E. Bunn, "Transmission line theory applied to sound wave propagation in tubes with compliant walls," Acustica **48**, 101–106 (1981).
[6]J. J. Fredberg, M. E. B. Wohl, G. M. Glass, and H. L. Dorkin, "Airway area by acoustic reflections measured at the mouth," J. Appl. Physiol. **48**, 749–758 (1980).
[7]L. Elvira-Segura, "Acoustic wave dispersion in a cylindrical elastic tube filled with a viscous liquid," Ultrasonics **37**, 537–547 (2000).
[8]R. Pico and F. Gautier, "The vibroacoustics of slightly distorted cylindrical shells: A model of the acoustic input impedance," J. Sound Vib. **302**, 18–38 (2007).
[9]F. Gautier, J. Gilbert, J.-P. Dalmont, and R. P. Vila, "Wave propagation in a fluid filled rubber tube: Theoretical and experimental results for Korteweg's wave," Acta. Acust. Acust. **93**, 333–344 (2007).
[10]V. A. D. Grosso, "Analysis of multimode acoustic propagation in liquid cylinders with realistic boundary conditions—Application to sound speed and absorption measurements," Acustica **24**, 299–311 (1971).
[11]S. Yomosa, "Solitary waves in large blood vessels," J. Phys. Soc. Jpn. **56**, 506–520 (1987).
[12]S. E. H. A. Erbay and S. Dost, "Wave propagation in fluid filled nonlinear viscoelastic tubes," Acta Mech. **95**, 87–102 (1992).
[13]T. Kamakura and Y. Kumamoto, "Waveform distortions of finite amplitude acoustic wave in an elastic tube," in Frontiers of Nonlinear Acoustics: Proceedings of the 12th ISNA, edited by M. F. Hamilton and D. T. Blackstock (1991), pp. 333–338.
[14]W. Chester, "Resonant oscillations in closed tubes," J. Fluid Mech. **18**, 44–64 (1964).
[15]S. N. Makarov and E. V. Vatrushina, "Effect of the acoustic boundary layer on a nonlinear quasiplane wave in a rigid-walled tube," J. Acoust. Soc. Am. **94**, 1076–1083 (1993).
[16]M. Bednarik and P. Konicek, "Propagation of quasiplane nonlinear waves in tubes and the approximate solutions of the generalized Burgers equation," J. Acoust. Soc. Am. **112**, 91–98 (2002).
[17]I. M. Sokolov, J. Klafter, and A. Blumen, "Fractional kinetics," Phys. Today **55**, 48–54 (2002).
[18]M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (MIT Press, Cambridge, 1986).
[19]O. V. Rudenko and S. I. Soluyan, *Theoretical Foundations of Nonlinear Acoustics* (Consultants Bureau, New York, 1977).
[20]K. Naugolnykh and L. Ostrovsky, *Nonlinear Wave Processes in Acoustics* (Cambridge University Press, Cambridge, 1998).
[21]M. F. Hamilton and D. T. Blackstock, *Nonlinear Acoustics* (Academic, New York, 1997).
[22]S. Makarov and M. Ochmann, "Nonlinear and thermoviscous phenomena in acoustics, Part II," Acta. Acust. Acust. **83**, 197–222 (1997).
[23]M. L. Munjal and P. T. Thawani, "Acoustic analysis and design of compliant cable-hose systems," Noise Control Eng. J. **45**, 235–242 (1997).
[24]J. Engelbrecht, *Nonlinear Wave Dynamics* (Kluwer Academic, Dordrecht, 1997).
[25]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1999).
[26]A. D. Polyanin and V. F. Zaitsev, *Handbook of Nonlinear Partial Differential Equations* (Chapman and Hall, London, 2004).
[27]L. Jiang, X. Chen, Z.-T. Fu, S.-K. Liu, and S.-D. Liu, "Periodic solutions to KdV-Burgers-Kuramoto equation," Commun. Theor. Phys. **45**, 815–818 (2006).
[28]V. R. Kodali, "Spectral analysis of numerical solutions to the Burgers-Korteweg-DeVries equation," J. Acoust. Soc. Am. **79**, 26–30 (1986).
[29]N. J. Zabusky and M. D. Kruskal, "Interactions of solitons in a collisionless plasma and the recurrence of initial states," Phys. Rev. Lett. **15**, 240–243 (1965).

Numerical investigation of nonlinear propagation distortion effects in helicopter rotor noise[a)]

Penelope Menounou and Panagiotis A. Vitsas
Department of Mechanical and Aeronautical Engineering, University of Patras, Patras 26504, Greece

(Received 7 August 2008; revised 13 July 2009; accepted 16 July 2009)

The effect of nonlinear propagation distortion on helicopter rotor noise is presented based on measured data for low-speed descent and numerical calculations that predict the noise level away from the helicopter with and without nonlinear effects. It is shown that for some frequency bands the difference between linear and nonlinear calculations can be as high as 7 dB. Blade vortex interaction (BVI) noise, the dominant noise contributor during descent, is mainly examined. It is shown that advancing side BVI noise is affected by nonlinear distortion, while retreating side BVI noise is not. Based on signal characteristics at source, two quantities are derived. The first quantity (termed polarity) is based on the pressure gradient of the source signal and can be used to determine whether a BVI signal will evolve as an advancing or a retreating side signal. The second quantity (termed weighted rise time) is a measure of the impulsiveness of the BVI signal and can be used to determine at which frequency nonlinear effects start to appear. Finally, polarity and weighted rise time are shown to be applicable in cases of BVI noise generated from different blade tips, as well as in cases of non-BVI noise. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203916]

PACS number(s): 43.25.Cb, 43.28.Bj [JWP] Pages: 1690–1699

I. INTRODUCTION

Nonlinear propagation distortion effects manifest themselves when the noise signal is of sufficiently high intensity. Each part of the signal travels with its own velocity, namely, the speed of sound plus the local velocity of the signal. Parts of the signal travel faster than the speed of sound, while others move slower. As a result, the original shape of the signal distorts, and the frequency distribution of the signal changes accordingly. The effect has been studied in jet aircraft noise propagation, where it has been demonstrated that nonlinear effects are responsible for the higher values measured at high frequencies compared to the values predicted using linear propagation theory.[1–4]

Literature on nonlinear propagation of helicopter noise is far less extensive. Nonlinear distortion has been considered insignificant for helicopter noise, as the noise field produced by a helicopter rotor is, in general, of relatively low intensity (compared to the noise field produced by a jet aircraft). Existing literature focuses in the special cases of transonic/supersonic flows around the blade tip, where shock waves of large pressure amplitudes are formed and, after dislocating from the blade surface, propagate in the surrounding medium.[5,6] The purpose of the present work is to demonstrate the effect of nonlinear propagation for subsonic cases, which to the best of the authors' knowledge, have not been considered before. The aim is to show which frequency bands of the noise spectrum are affected and by how much. In the case of jet aircraft noise, measurements first became available, and their differences with linear acoustics predictions prompted research on nonlinear propagation distortion. In the case of helicopters, however, there is almost a complete lack of experimental data appropriate for nonlinear propagation studies. The present investigation is thus numerical. Specifically, the Burgers equation is employed to numerically predict the evolution of noise signals, measured close to the helicopter rotor, at distances far away from the rotor. The calculations are performed twice, with and without including nonlinear effects. The difference between the calculations is a measure of the effect of nonlinear distortion. The numerical results from the present work might stimulate experimental research on the topic.

Helicopter rotors produce a very complicated directional noise field through several distinct noise generation mechanisms. The reader is referred to review papers, such as those of Brentner and Farassat[7,8] or Schmitz,[9] for a detailed description of helicopter noise. A very brief outline is given next. The noise field produced by a helicopter rotor consists of deterministic discrete-frequency noise components (thickness and loading noise, blade vortex interaction (BVI) noise, and high-speed impulsive (HSI) noise), as well as of nondeterministic broadband noise components. The various noise types are attributed to different generation mechanisms. Thickness noise is caused by the displacement of the air by the rotor blades while loading noise by the accelerating force on the air that is generated by the moving blade surface. BVI noise is an impulsive loading noise generated when a rotor blade passes within a close proximity of the shed tip vortices from the previous blades. HSI noise is an extreme case of thickness noise generated as the flow around the blade becomes transonic. Finally, broadband noise (a type of loading noise) has its origin in the interaction of the moving rotor blade with atmospheric turbulence and turbulence that is shed from the blade itself or from the previous blades.

[a)]Portions of this research were presented at the 155th Meeting of the Acoustical Society of America, Seventh EURONOISE, Paris, France, July 2008.

The present study does not investigate the generation of helicopter noise. The noise signals at source are considered known from available measurements, and they are considered characteristic of the flight conditions for which they were measured. For example, HSI noise[10] occurs during forward flight and is characterized by pulses of large negative pressure amplitude, while BVI noise[11] occurs mainly during descent and is characterized by pulses that contain both positive and negative pressure amplitudes. The nonlinear propagation of these signals, irrespective of the mechanism that generated them, is the focus of the present study. More specifically, a set of measured BVI noise source signals characteristic of low-speed descent noise is analyzed. BVI noise is particularly annoying when it occurs and very often sets the standards of acceptance for helicopter noise, and its nonlinear propagation has not been studied before. Analysis of the available database reveals the magnitude of nonlinear propagation effects on BVI noise. It also reveals certain characteristics of the noise source signals, which can be used to qualitatively predict the nonlinear behavior of other BVI noises (coming, for example, from different blades shapes), as well as the nonlinear behavior of non-BVI noise.

The present paper is structured as follows: In Sec. II, the numerical calculations are described. In Sec. III, results for nonlinear propagation of BVI noise are presented and commented on. In Sec. IV, characteristics of BVI noise source signals that affect their nonlinear evolution are presented. The effects of the blade tip shape and of broadband frequency noise on the nonlinear evolution of BVI noise are also investigated. Finally, in Sec. V, the nonlinear evolution of non-BVI noises is briefly considered.

II. NONLINEAR PROPAGATION DISTORTION: NUMERICAL PREDICTION

A. Augmented Burgers equation

The Burgers equation is the simplest equation that describes the combined effect of nonlinearity and thermoviscous attenuation in the propagation of a sound wave. In its augmented form, the Burgers equation includes the effect of geometrical spreading (spherical spreading in this paper's case), as well as the absorption and dispersion attributed to the molecular relaxation of O_2 and N_2 in the atmosphere. It can, therefore, be employed for the prediction of finite-amplitude sound emanating from a point source and propagating through the atmosphere. The atmosphere is considered homogeneous and at rest. It is described by its temperature and relative humidity, which in turn determine the absorption due to molecular relaxation. A form of the augmented Burgers equation[12,13] is the following:

$$\frac{\partial p}{\partial r} = -\frac{1}{r}p + \frac{\beta}{\rho_0 c_0^3} p \frac{\partial p}{\partial \tau} + \frac{\delta}{2c_0^3}\frac{\partial^2 p}{\partial \tau^2} + \sum_{\nu} \frac{c'_\nu}{c_0^2} \int_{-\infty}^{\tau} \frac{\partial^2 p}{\partial \tau'^2} e^{-(\tau-\tau')/t_\nu} d\tau', \quad (1)$$

where p is the sound pressure, r is the propagation distance, $\tau = t - (r - r_0)/c_0$ is the retarded time, with r_0 being the radius of the point source, ρ_0 is the ambient density, c_0 is the small signal sound speed, β is the coefficient of nonlinearity, δ is the diffusivity of sound for viscosity and heat conduction, and $\nu = 1, 2$ is 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. The first term in the right-hand-side of Eq. (1) represents the spherical spreading, the second represents the nonlinear distortion, the third describes the thermoviscous attenuation, and the fourth combines the two relaxation effects, corresponding to O_2 and N_2. The third and fourth terms together describe mathematically the atmospheric absorption.

The augmented Burgers equation is solved numerically in the time domain using the "Texas algorithm" described in Refs. 14 and 15. The algorithm marches in the propagation direction, and for each marching step the aforementioned mechanisms are applied sequentially. This procedure is referred to as "operator-splitting," and for sufficiently small steps the solution approaches that obtained by applying all mechanisms simultaneously. The algorithm allows the calculations to be performed by including only the effects considered at the time (spherical spreading, atmospheric absorption, and/or nonlinear propagation distortion). In the following, linear calculations shall mean the solution of Eq. (1) in the absence of the nonlinear term.

B. Accuracy of the augmented Burgers equation

The augmented Burgers equation in all its forms has been used and tested extensively for the prediction of nonlinear noise propagation.[14,16] It has also been employed specifically for the prediction of jet aircraft noise[3] and sonic boom[17] propagation in the atmosphere. In the following, results obtained by Eq. (1) are compared with, to the best of the authors' knowledge, the only available experimental data for helicopter noise. The data are from Schmitz *et al.*[10] and concern HSI noise. Measurements were taken at two microphones positioned in the rotor plane along an imaginary line from the rotor hub directly ahead of the model rotor, the second microphone being exactly twice the distance of the first [see Fig. 1(a)]. The ratio of the peaks (P_2/P_1) of the two measured signals was reported as a function of the advancing tip Mach number (M_{AT}) [see Fig. 1(b)]. It was shown that the ratio was approximately 2 for all M_{AT}, thus verifying the spherical spreading/far field hypothesis, for which the experiments were performed. However, deviations from the nominal value of 2 were observed. The deviations at low M_{AT} can be attributed to flow unsteadiness and/or experimental uncertainties, while for higher M_{AT}, and particularly above delocalization ($M_{AT} \simeq 0.9$), the difference can be explained by nonlinear propagation distortion. Table I and Fig. 1(b) show the comparison between experimental data, linear, and nonlinear calculations. The predictions have been performed for International Standard Atmosphere (ISA) standard day with 15 °C temperature and 0% relative humidity according to the wind tunnel's conditions. It can be observed that nonlinear predictions agree very well with experimental data, while linear predictions deviate from them. The energy

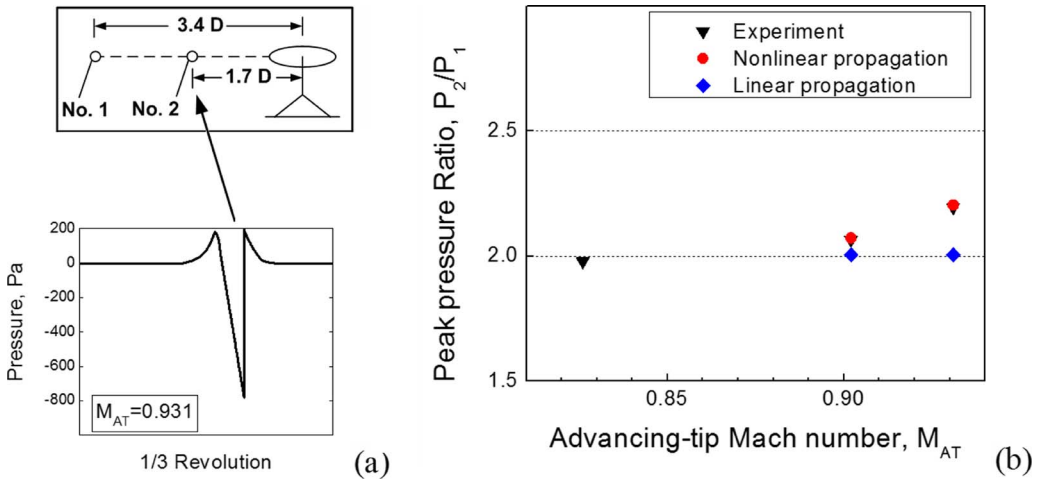


FIG. 1. (Color online) (a) Nonlinear propagation distortion can explain differences between measurements and linear predictions; experimental data from Ref. 10 for HSI noise; details of experimental setup (measurement positions and measured signal at Mic 2). (b) Comparisons between experimental data and numerical predictions employing Eq. (1).

absorbed at the shock is, therefore, responsible for the extra decay observed at high M_{AT}, which is accurately predicted by the Burgers equation.

C. Application to subsonic BVI noise and related numerical issues

The application of the augmented Burgers equation to subsonic BVI noise is discussed next. As noise source signals, the authors considered sound pressure time signals measured during the HELISHAPE project[18] for low-speed descent. This was one of the noisiest flight conditions tested, where BVI noise is dominant. Measurements were conducted in the German–Dutch wind tunnel (DNW) with a four bladed rotor having a 4.2 m diameter and a swept-back parabolic-anhedral blade tip shape. The flight speed was 35 m/s, and the descent flight had a 6° path angle. The acoustic pressure was measured with 0.5-in. pressure type condenser microphones at an array of 11 equally spaced (0.54 m apart) microphones. The array span was positioned normal to the flow and symmetrically arranged with respect to the rotor center [see Fig. 2(a)]. Measurements were taken at all 11 microphones and at several streamwise locations of the array span. The array's vertical position was 2.3 m below the rotor hub.

For the purpose of the present study, the following should be noted regarding the noise source signals: (i) The noise source signals stem from measurements conducted with a model rotor. Before being employed in the numerical study, they have been transformed into full scale rotor noise signals using the appropriate scaling procedure.[9] A scale factor of $\gamma=0.34^{-1}$ has been used in the transformation. Accordingly, lengths and times have been scaled by γ. It should be noted that the low advance ratio in the experiments, $\mu=0.165$, indicates an acceptable scalability[19,20] between model and full scale data for BVI noise, while the advancing tip Mach number, $M_{AT}=0.724$, does not imply transonic blade effects.[20] (ii) No time window has been applied to the measured noise signals. The ensemble averaged sound pressure time signal (with 30 averages) has been considered at each measurement location, not the instantaneous time signal. The signals considered do not contain high levels of broadband rotor noise, which is largely eliminated by the averaging procedure. However, propagation of instantaneous signals is also briefly examined in the present work. (iii) Throughout the study, both the measurement location and the set of calculations corresponding to that location are identified by the streamwise position of the array (x) and the microphone number. For example, Mic=6/x=0 indicates the measurement point right underneath the rotor head center.

The noise source signals were propagated numerically by employing the augmented form of the Burgers equation [Eq. (1)]. Predictions regard sound coming from a directive stationary point source located at the rotor head center and propagating distance R from the rotor head center through the microphone grid to receiver locations on the surface of a hemisphere, as shown in Fig. 2(b). The calculations have been performed twice, once including only linear propagation effects (geometrical spreading and atmospheric absorption) and a second time adding nonlinear propagation distortion to the linear calculations. Nonlinear effects are manifested as the difference in the predicted sound pressure level (SPL) values at each one-third-octave frequency band:

TABLE I. Comparison of results obtained by linear and nonlinear calculations [Eq. (1)] with experimental data (see also Fig. 1).

M_{AT}	P_2 (Pa)	$P_1^{(linear)}$ (Pa)	$P_1^{(nonlinear)}$ (Pa)	$(P_2/P_1)^{(linear)}$	$(P_2/P_1)^{(nonlinear)}$	$(P_2/P_1)^{(measured)}$
0.902	−412.34	−205.6	−199.0	2.005	2.072	2.064
0.931	−800.04	−398.8	−363.0	2.006	2.204	2.194

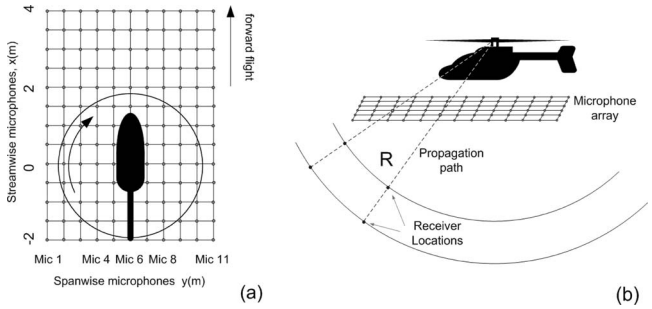


FIG. 2. (a) Location of measurement points on a plane 2.3 m below the rotor head center [HELISHAPE project (Ref. 18)]. (b) Propagation paths in the present numerical investigation.

$$\mathrm{DSPL}(f_{1/3};R) = \mathrm{SPL}_{\mathrm{nonlinear}}(f_{1/3};R) - \mathrm{SPL}_{\mathrm{linear}}(f_{1/3};R), \tag{2}$$

where R is the propagation distance and $f_{1/3}$ is the center frequency of the one-third-octave frequency bands. Positive values of DSPL indicate that nonlinear propagation distortion enhances the noise spectrum at the given frequency band.

The measured time signals had duration of one rotor revolution and were digitized with 2048 points. In the computations, however, time signals having duration of three rotor revolutions and 10240 points per rotor revolution were used instead, while results were obtained by analyzing the middle of the three rotor revolutions. This was done to ensure that numerical inaccuracies at the ends of the elongated time signal do not contaminate the middle part, which was actually analyzed. The time sampling increase was done by linearly interpolating additional points between the initial points of the measured signal in order to avoid excess numerical dissipation around the peak values of the signal.

III. DEMONSTRATION OF NONLINEAR PROPAGATION DISTORTION EFFECTS IN BVI NOISE

In this section, representative results obtained by the above described method for BVI noise are presented and commented. Consider the noise spectrum at source as measured at Mic $6/x=-2.5$ [see Fig. 3(a)]. The spectrum can be subdivided into three frequency regions:[21] (i) the low frequency region up to approximately 100 Hz (or, equivalently, up to the fifth blade passage frequency harmonic), which corresponds to thickness noise; (ii) the mid frequency region approximately from 100 to 1000 Hz (or, equivalently, from the 5th to the 40th–50th blade passage frequency harmonic), which corresponds to BVI noise; and (iii) the high frequency region from 1000 Hz to the end of the spectrum, corresponding to broadband noise. The Gol'dberg number[12] for the low and mid frequency region of the spectrum is $\Gamma \simeq 0.4 < 1$, while for the high frequency region $\Gamma \simeq 0.01 \ll 1$. It is recalled that the Gol'dberg number is a measure of the relative importance of nonlinear and absorption effects and is defined as the ratio $\Gamma = l_\alpha / \bar{x}$ between the absorption length l_α ($l_\alpha = a^{-1}$, where α is the attenuation coefficient) and the shock formation distance $\bar{x}$ of a plane sinusoidal wave in the absence of dissipation ($\bar{x} = 1/\beta\varepsilon k$, with $\varepsilon = p_0/\rho_0 c_0^2$ being the acoustic Mach number at source and $k = \omega/c_0$ the corresponding wave number). The arithmetic values of the Gold'berg number have been obtained by considering characteristic values of pressure amplitude p_0, frequency f, and corresponding atmospheric coefficient α [the latter from ANSI (Ref. 22)] for the different frequency regions of the spectrum.

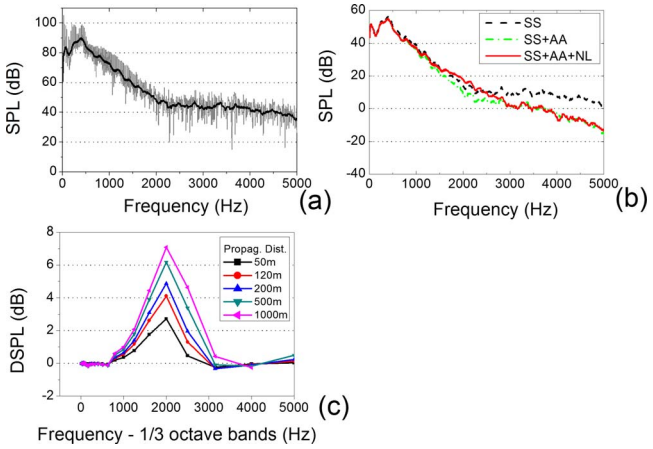


FIG. 3. (Color online) (a) Nonlinear propagation distortion of helicopter rotor BVI noise, Mic $6/x=-2.5$; noise spectrum at source. (b) Contribution of the various propagation mechanisms and comparison between linear and nonlinear calculations at propagation distance $R=500$ m (SS: spherical spreading, AA: atmospheric absorption, NL: nonlinear distortion, SS+AA: linear calculations, SS+AA+NL: nonlinear calculations). (c) DSPL for helicopter BVI noise at various propagation distances vs 1/3 octave frequency bands.

Figure 3(b) shows the evolution of the noise spectrum for Mic $6/x=-2.5$ after $R=500$ m of propagation. A difference between linear and nonlinear predictions can be observed in the region between 1000 and 3000 Hz. The observed difference is the focus of the present study.

The behavior depicted in Fig. 3(b) differs from nonlinear propagation of other noise sources, for example, helicopter rotor in supersonic flows or jet noise, where the difference between linear and nonlinear predictions increases with increased frequency. The indicated linear decay of the high frequency end of the spectrum is attributed to the initial source condition and its weak nonlinearity. The pressure amplitude at higher frequencies is a combination of (i) sound coming from the source, which is low enough in amplitude to propagate as a linear wave (e.g., without generating harmonics, recall that $\Gamma \simeq 0.01 \ll 1$), and (ii) nonlinear sound generated from interaction with the lower frequencies, which is negligible in comparison to the initial amplitude and also to the changes due to spreading and absorption. Therefore, the nonlinearity does not contribute significantly to the amplitude of these high frequency components, and so they evolve as linear waves.

Figure 3(b) also demonstrates the relative significance of the various propagation effects involved, namely, spherical spreading, atmospheric absorption, and nonlinear propagation distortion. Spherical spreading is the mechanism that mainly determines the evolution of the noise spectrum. Nonlinear distortion is less important than spherical spreading but equally important as atmospheric absorption. Atmospheric absorption[22] is routinely added to geometrical spreading to improve the accuracy of the predictions. It is the authors' opinion that nonlinear distortion should also be included.

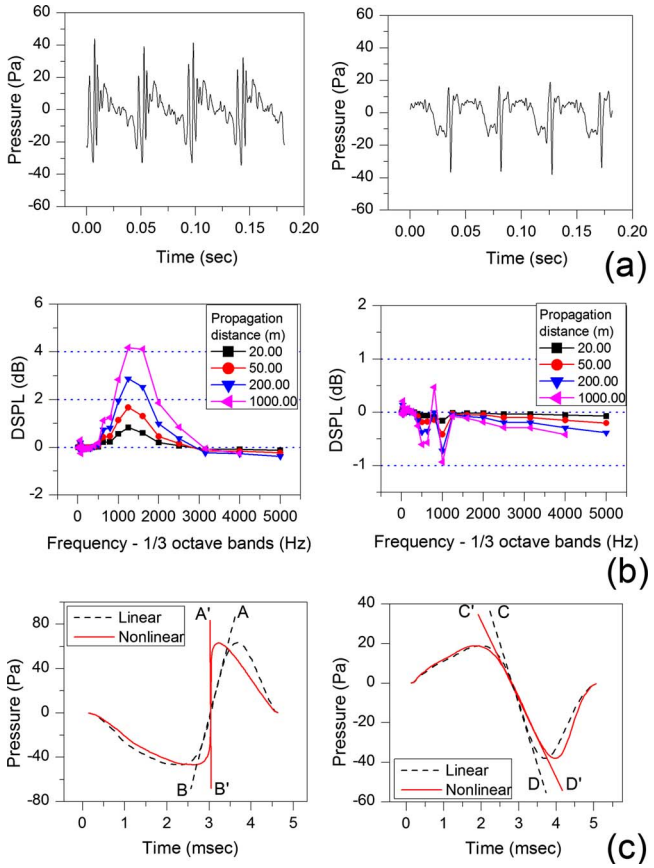


FIG. 4. (Color online) Advancing side BVI (left column) and retreating side BVI (right column): (a) signals at source. (b) Their DSPL at various propagation distances. (c) Nonlinear evolution (geometrical spreading and atmospheric absorption ignored) of main BVI pulses isolated from the signals at source.

Finally, Fig. 3(c) shows the DSPL vs one-third-octave frequency bands for Mic $6/x=-2.5$ at various propagation distances. It can be observed that DSPL increases with increased distance and that the magnitude of DSPL is sizeable, and in some cases it can be as large as 7 dB.

IV. CHARACTERISTICS OF NONLINEAR PROPAGATION DISTORTION OF BVI NOISE

In the following, it is investigated how certain characteristics of BVI pulses at source relate to their nonlinear propagation distortion. For a detailed description of BVI noise signals, the reader is referred to Schmitz[9] and Yu.[11]

A. Advancing side/retreating side BVI noise-polarity

Figure 4(a) shows two signals, one typical of advancing side BVI (Mic $3/x=0$) and one of retreating side (Mic $11/x=2$). The characteristic BVI pulses have predominantly positive pressure amplitudes in the advancing side and predominantly negative in the retreating side. This change in sign is attributed to the opposite rotational directions of the vortices during BVI on the advancing and retreating side. In the advancing side, the small negative part of the pulse comes before the large positive part, while the opposite is true for the retreating side. Also, due to geometrical considerations, the non-BVI noise between the blade passages comes, in general, after the BVI pulse in the advancing side and before the BVI pulse in the retreating side. Finally, in most cases multiple pulses at each blade passage in the advancing side are observed, as opposed to the single pulse at each blade passage in the retreating side.

The evolution of two signals yields the DPSL shown in Fig. 4(b). All BVI source signals of the database evolve either as an advancing side type signal, yielding the characteristic DSPL bell or as a retreating side type signal. The following differences between the two types are common in all receiver locations: (i) the magnitude of DSPL in the advancing side is substantially larger than that in the retreating side; (ii) DSPL values are predominantly positive in the advancing side while predominantly negative in the retreating side; (iii) the frequencies mainly affected in the advancing side are the frequencies in the octave bands of 1000 and 2000 Hz, while the frequencies mainly affected in the retreating side are frequencies from 300 to 1000 Hz; and (iv) although very small in magnitude, DSPLs in the lowest frequency bands, up to 300 Hz, are negative in the advancing side while positive in the retreating side. The magnitude of DSPL in the retreating side is so small that it can be considered zero for all practical purposes. However, as will be shown in this section, the negligible DSPL values on the retreating side should not be considered as a limiting case of advancing side type nonlinear evolution but as a different type of evolution.

The difference in the nonlinear evolution between advancing and retreating side BVI signals can be explained by observing the evolution of their main pulses, which have been isolated from the two signals and shown in Fig. 4(c). In order to clearly show the effect, only nonlinear distortion was considered. It can be observed that the segment of the signal connecting the peak with the trough steepens due to nonlinear propagation in the advancing side, while it "unsteepens" in the retreating side. Since in the advancing side BVI pulse, the small negative part of the pulse comes before the large positive part, the segment of the pulse between peak and trough will always have a positive pressure gradient and will steepen during propagation. Correspondingly, in the retreating side, it will have a negative pressure gradient and will un-steepen.

Consider the straight lines passing through points AB $(p_{(\mathrm{adv})}(0,t)=m_a t,\ m_a>0)$ and CD $(p_{(\mathrm{ret})}(0,t)=-m_r t,\ m_r>0)$ in the advancing and retreating side source signals, respectively, as shown in Fig. 4(c). For plane wave propagation in a non-dissipative fluid, the governing equation is $p_t+c_0 p_x=\beta/\rho_0 c_0^2 p p_t$ and the general solution is $p(x,t)=f(t-x/c_0+\beta x p/\rho_0 c_0^3)$, where $f(t)=p(0,t)$.[12] For the advancing side and retreating side, respectively, the solution becomes

$$p_{(\mathrm{adv})} = \frac{m_a}{1 - \dfrac{\beta m_a x}{\rho_0 c_0^3}}\tau = \frac{m_a}{1 - \dfrac{x}{\bar{x}_a}}\tau = A_{(\mathrm{adv})}\tau,$$

$$\tau = t - \frac{x}{c_0}, \quad \bar{x}_a = \frac{\rho_0 c_0^3}{\beta m_a}, \tag{3}$$

$$p_{(\mathrm{ret})} = \frac{-m_r}{1 + \dfrac{\beta m_r x}{\rho_0 c_0^3}}\tau = \frac{-m_r}{1 + \dfrac{x}{\bar{x}_r}}\tau = A_{(\mathrm{ret})}\tau,$$

$$\tau = t - \frac{x}{c_0}, \quad \bar{x}_r = \frac{\rho_0 c_0^3}{\beta m_r}. \tag{4}$$

The above solutions indicate straight lines with inclination angles $A_{(\mathrm{adv})}, A_{(\mathrm{ret})}$ that change with propagation distance. The tendency of $p_{(\mathrm{adv})}$ is to steepen until a shock is formed (at $x=\bar{x}_a$), while that of $p_{(\mathrm{ret})}$ is to un-steepen, and no shock will ever be formed at that segment of the wave form. In the frequency domain, the former is associated with generation of higher frequencies and depletion of lower frequencies, while the latter is associated with generation of lower frequencies and depletion of higher frequencies. It should be noted that $A_{(\mathrm{adv})} = \partial p_{(\mathrm{adv})}/\partial\tau > 0$ increases with increased propagation distance x ($x < \bar{x}_a$) until $A_{(\mathrm{adv})}$ becomes infinite at $x=\bar{x}_a$. Also $A_{(\mathrm{ret})} = \partial p_{(\mathrm{ret})}/\partial\tau < 0$ increases with propagation distance until it becomes zero for very long propagation distances. Pulses idealized as being made up from segments of straight lines of varying inclination angles $A \propto \partial p/\partial t$ tend to transfer energy to higher or lower frequencies depending on which tendency is overall prevailing in the entire time wave form. This tendency can be quantified by taking into account the value of the pressure gradient $A \propto \partial p/\partial t$ of each segment weighted by the local pressure amplitude.

A new quantity has been derived that provides an indicator of the tendency of the BVI signal to steepen and thus follow an advancing side type nonlinear evolution or to unsteepen and thus follow a retreating side type nonlinear evolution. The quantity is termed *polarity* (Π) and is defined as follows:

$$\Pi^3 = \frac{\sum_{i=2}^{N} (P_i^+ \times \Delta P_i^+)^2}{N^+} - \frac{\sum_{i=2}^{N} (P_i^- \times \Delta P_i^-)^2}{N^-}, \tag{5}$$

where N is the number of points in the digitized time signal, P_i is the value of the pressure at point i, $\Delta P_i = P_i - P_{i-1}$, P_i^+ is the pressure at point i, when $\Delta P_i > 0$ ($\Delta P_i = \Delta P_i^+$), P_i^- is the pressure at point i when $\Delta P_i < 0$ ($\Delta P_i = \Delta P_i^-$), N^+ is the number of points for which $\Delta P_i > 0$, and N^- is the number of points for which $\Delta P_i < 0$. According to Eq. (5), the main pulse of the advancing side BVI signal has positive polarity, while the main pulse of the retreating side BVI has negative polarity. Polarity can be used to characterize signals with mixed advancing and retreating side characteristics as either advancing or retreating. Figure 5 shows the polarity of all source signals. As far as nonlinear distortion is concerned,

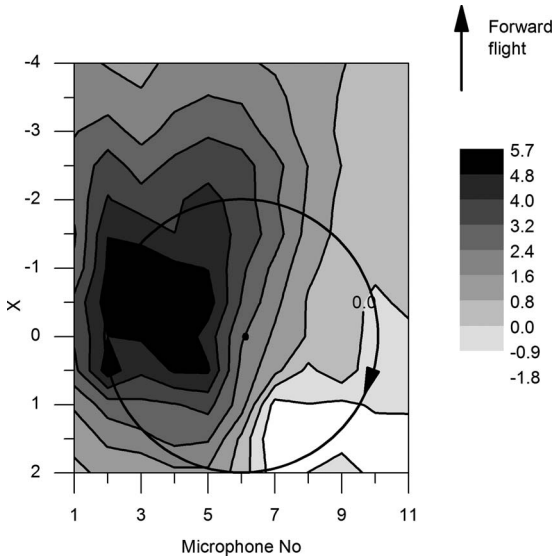


FIG. 5. Contour plots of polarity of all noise source signals on a plane underneath the rotor. Negative polarity corresponds to retreating side type nonlinear evolution; positive polarity corresponds to advancing side type nonlinear evolution.

visual observation of all DSPL plots reveals that source signals with negative polarity exhibit a retreating side type DSPL, while source signals with positive polarity show the characteristic advancing side DSPL bell. It should be further noted that the absolute value of polarity increases for signals with a strong content in advancing or retreating BVI noise. The polarity contours, therefore, resemble the traditional medium-frequency contours used to define advancing and retreating BVI regions of rotors in descending flight.[21]

It should be noted that the polarity as defined in the present work is particularly helpful for BVI pulses, as it predicts both the nonlinear evolution type of the pulse and the strong advancing BVI regions. If it is to be used for other noise source types, a different weighing might be more appropriate depending on the characteristics of the given noise source type. In any case, it should be applied in the preshock region, and only the sign of Π should be considered. Its arithmetic value might not be significant for the problem at hand. It should be emphasized that, as discussed for $A_{(\mathrm{adv})}$ and $A_{(\mathrm{ret})}$, the value of polarity Π changes with propagation distance and may change sign. An initially positive Π shall continue to be positive, while an initially negative Π might become positive during propagation. Thus, although the sign of Π at any distance (before shock formation) is indicative of the nonlinear evolution in the immediately next propagation steps, its negative sign at source is indicative only of relatively weak waves (like the ones handled here), where no drastic changes in the wave form take place.

B. Effect of impulsiveness of advancing side BVI noise-weighted rise time

Noise source signals on the advancing side vary considerably in shape. They range from signals with multiple, less impulsive peaks at each blade passage (see signal A in Fig. 6) to signals containing a single, very impulsive peak at each blade passage (see signal B in Fig. 6). The corresponding

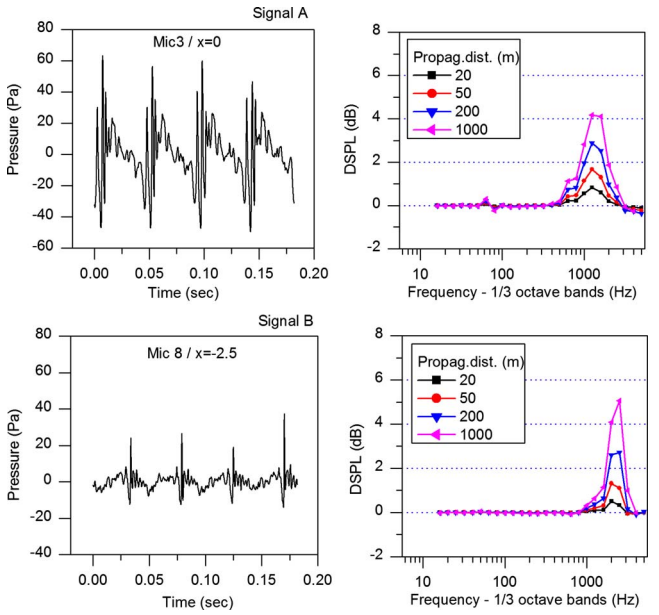


FIG. 6. (Color online) Pressure signals at source vary from containing multiple less impulsive BVI pulses to containing a single very impulsive BVI pulse (left column); corresponding DSPL plots–DSPL bells shift toward higher frequencies (right column).

DSPL bell moves to higher frequencies, as the signal transitions from one category to the other (see Fig. 6).

Signals A and B differ not only in the number of intense BVI peaks within each blade passage but also in the non-BVI noise between blade passages. Numerical experimentation showed that the shift of the DSPL bell toward higher frequencies cannot be explained by either of these differences. Further numerical experimentation showed that the starting frequency of the DSPL bell (f_{start}) is related to the rise time of the main BVI pulse. *Rise time* (DT) is the time from the trough to the peak of the main BVI pulse [see Fig. 7(a)]. The duration of the pulse, also called emission time (T), does not seem to directly affect f_{start}.

In the past, the emission time has been employed to analyze the spectral characteristics of BVI signals.[23] For the purposes of the present study, the rise time seems to be a more appropriate parameter. Consider a BVI signal idealized as a train of N-shaped pulses with $DT < T/2$. First, the local minima in the corresponding spectrum occur at the rise frequency harmonics $(1/DT, 1/2DT, \ldots)$, not the emission frequency harmonics $(1/T, 1/2T, \ldots)$ [see Fig. 7(a)]. Furthermore, as the signal undergoes nonlinear distortion, the rise time of each pulse decreases ($DT' < DT$), while the emission time remains the same. As a result, the local minima in the spectrum occur again at the rise frequency harmonics of the propagated signal $(1/DT', 1/2DT', \ldots)$, which have been displaced to higher frequencies compared to the rise frequency harmonics of the source signal $(1/DT, 1/2DT, \ldots)$ [see Fig. 7(a)]. Finally, ignoring spherical spreading and absorption, DSPL is obtained as the difference between the spectra at source and after some propagation, as shown in Fig. 7(b). It can be observed that the frequency, where nonlinear effects start to appear f_{start}, is approximately $1/DT$.

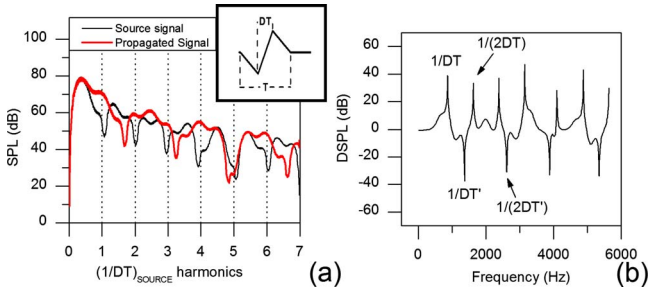


FIG. 7. (Color online) (a) Spectrum of a BVI signal idealized as a train of N-shaped pulses at source and after 150 m of propagation (nonlinear effects only); (b) corresponding DSPL plot.

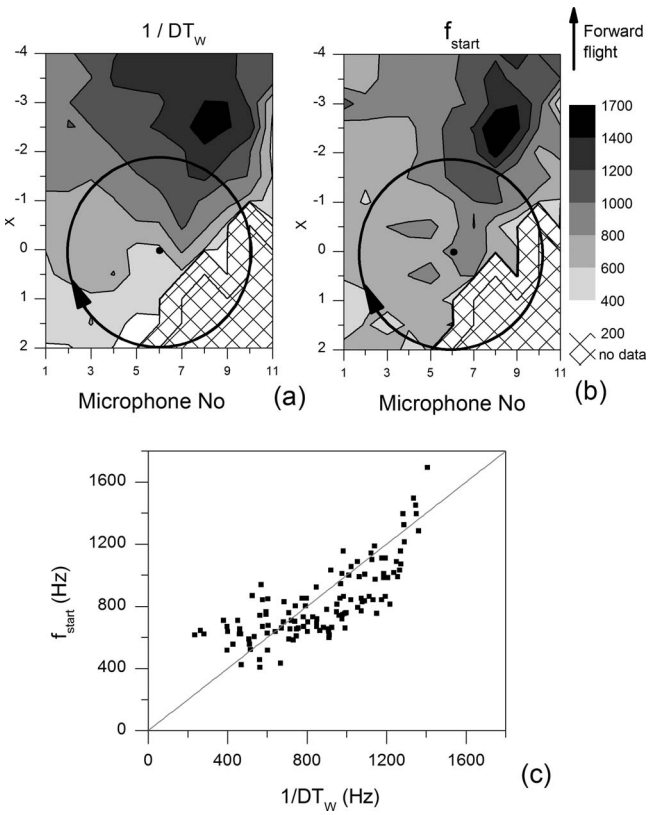


FIG. 8. (a) Contours of weighted rise frequency $(1/DT_W)$ of all noise source signals on a plane underneath the rotor. (b) f_{start} of the corresponding DSPL bells after 120 m of propagation. (c) Correlation between $1/DT_W$ at source and f_{start} after 120 m of propagation.

For the actual rotor BVI signals, the analysis is more complex, as multiple, different, non-ideal BVI pulses are contained in the pressure time signal. As a predictor of f_{start}, therefore, a weighted rise time is proposed that takes into account the rise time of all pulses in the signal. The *weighted rise time* (DT_W) is defined as follows:

$$DT_W = \frac{\sum_{i=1}^{n} DT_i \times DP_i^4}{\sum_{i=1}^{n} DP_i^4}, \tag{6}$$

where n is the number of pulses in the pressure time signal, DT_i is the rise time of the ith pulse, and $DP_i = P_{\max,i} + |P_{\min,i}|$ of the ith pulse. Figures 8(a) and 8(b) show the weighted rise frequency $(1/DT_W)$ of all source signals and the f_{start} of their corresponding DSPL bell. It can be observed

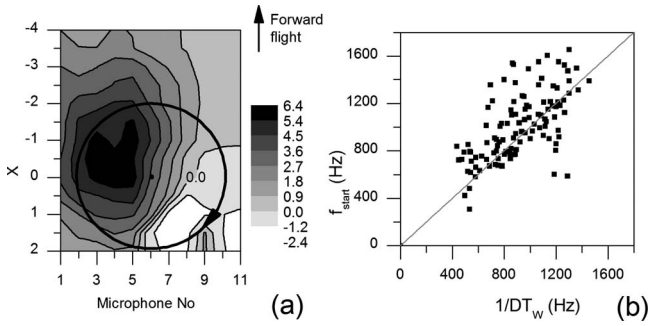


FIG. 9. Rectangular tip blade (Ref. 18): (a) polarity contour of all noise source signals on a plane underneath the rotor; (b) correlation between f_{start} of all noise signals after 120 m of propagation and weighted rise frequency $(1/DT_W)$ of all signals at source.

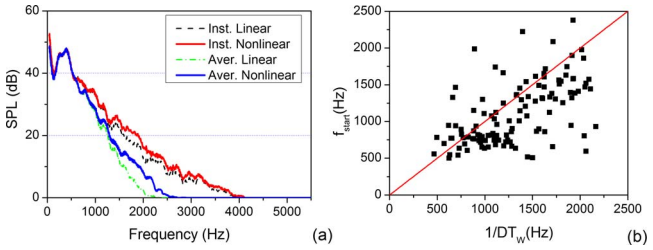


FIG. 10. (Color online) Instantaneous BVI noise signals: (a) comparison of spectra of averaged and instantaneous BVI noise signal after 1000 m of propagation for Mic $6/x=-2.5$; (b) correlation between f_{start} of all noise signals after 120 m of propagation and weighted rise frequency $(1/DT_W)$ of all signals at source.

that $1/DT_W$ can predict the areas where f_{start} has large values and also the approximate value of f_{start}. For example, in the region located upwind and on the retreating side of the rotor, f_{start} has its largest value and, accordingly, nonlinear distortion affects only the high frequencies. On the opposite side, in the region downwind and on the advancing side, nonlinear effects kick in at lower frequencies. Finally, the correlation plot of $1/DT_W$ and f_{start} indicates that $f_{\text{start}} \simeq 1/DT_W$ [Fig. 8(c)].

It should be noted that f_{start} relates only to nonlinear effects, while $1/DT_W$ is both a predictor for f_{start} and also a measure of impulsiveness of the source signal (where in the present context impulsiveness should be understood in terms of rise time). The contour plot of $1/DT_W$ indicates the areas of impulsive signals irrespective of their subsequent nonlinear evolution. The specific weighting used in the definition of $f_{\text{start}} \simeq 1/DT_W$ is specific to the signals at hand. A different weighting might be appropriate for signals of different type. Although the suitability of $1/DT_W$ to predict the arithmetic value of f_{start} depends on the specific weighting, $1/DT_W$ can be used in any case to compare two signals with regard to their impulsiveness (keeping in mind that $1/DT_W$ has been derived for signals with positive polarity and in the preshock region).

C. Effect of blade tip shape and broadband frequency noise in the nonlinear evolution of BVI noise

The blade tip shape changes the shed vortex that interacts with the succeeding blade to generate the BVI pulse. This changes the BVI signal at source and, thus, its nonlinear evolution. To investigate the effect, a second set of measurements, also taken during the HELISHAPE project,[18] was analyzed. The flight conditions and the measurement points were the same, but a rotor with a rectangular blade tip was used instead of the swept blade tip. An extended numerical analysis, similar to that of the measurements below the swept tip blade, verified similar DSPL values and frequency bands affected. Furthermore, the polarity of the source signal as a predictor of the nonlinear evolution type (advancing vs retreating) as well as the relation between weighted rise frequency $(1/DT_W)$ and f_{start} are shown to hold for the rectangular tip as well [Figs. 9(a) and 9(b)]. Figure 9(a) shows an area on the retreating side with positive polarity. Indeed, signals measured in that area evolve, following more the advancing side type than the retreating side type.

BVI noise is unsteady in nature, and instantaneous BVI noise signals can be significantly different from averaged noise signals. The environment in the DNW tunnel, however, was very steady, and averaged noise signals do not differ drastically from instantaneous. The main difference between them is the higher values at the broadband frequency region of the spectrum in the latter case. In general, the averaging procedure broadens the BVI pulse and lowers the peak-to-trough pressure amplitude.[24] It further eliminates the pressure fluctuations on the main BVI pulse and also the broadband noise, which in the time signal appears as small pulses immediately after the BVI pulses at each blade passage. As a result, the high frequency region of the power spectrum has higher values, by up to 10 dB, in the case of instantaneous signals. In terms of the Gol'dberg number Γ, the broadband frequency region is not as weak as in the case of the averaged signals. Consequently, the power spectrum of the instantaneous signal after propagation has a stronger high frequency component than the power spectrum of the averaged signal, and DSPL extends to higher frequencies [Fig. 10(a)]. The magnitude of DSPL, however, is more pronounced in the averaged signal. An extended numerical investigation of the instantaneous source signals verified the above differences in DSPL, as well as the suitability of polarity Π and weighted rise frequency $1/DT_W$ as predictors for the nonlinear evolution type and f_{start}, respectively. As expected, the polarity contours are the same for both averaged and instantaneous signals, as the BVI mechanism remains the same, while the higher values of $1/DT_W$ and f_{start} in the instantaneous signals [compare Figs. 9(b) and 10(b)] are in line with the fact that averaged signals are less impulsive due to the averaging procedure outlined before.

V. DEMONSTRATION ON NONLINEAR PROPAGATION DISTORTION IN NON-BVI NOISE

Cases of non-BVI noise are briefly examined next. Specifically, the cases of thickness noise and HSI noise, with and without shocks, are considered. Thickness noise is caused by the displacement of the air by the rotor blades and depends on the blade thickness distribution and the advancing tip Mach number (M_{AT}). Thickness noise is characterized by symmetrical negative pulses. As M_{AT} increases, the negative pulse grows dramatically in amplitude and dominates the

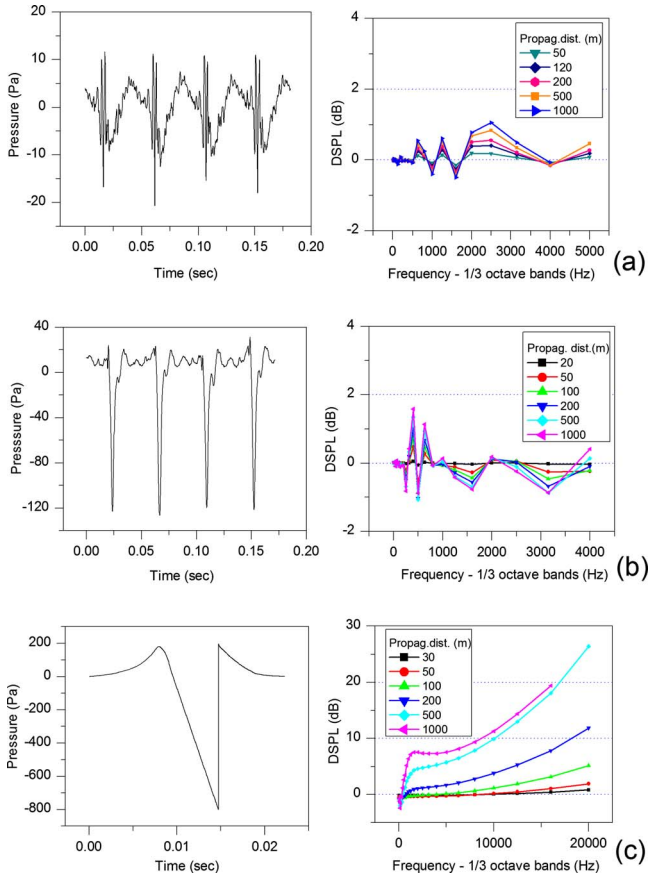


FIG. 11. (Color online) Non-BVI noise: pressure time signals at source (left column) and corresponding DSPL plots (right column). Line (a): thickness noise combined with BVI noise, M_{AT}=0.720 [signal taken during descent on the rotor plane 21.9 m from rotor head center (Ref. 18)]. Line (b): HSI noise, M_{AT}=0.881, shock-free [taken for lifting forward flight below a four-bladed rotor (Ref. 25)]. Line (c): HSI noise, M_{AT}=0.931, with shock [taken for lifting forward flight on the rotor plane of a two-bladed rotor (Ref. 10)].

time signal, but no shocks are formed. As M_{AT} increases further (above delocalization), a strong signal of very large amplitude is produced and shocks are formed. The left column of Fig. 11 shows this transition for increasing M_{AT} (albeit measured for different blades and flight conditions). Thickness noise (combined with BVI noise) is depicted on the first row, HSI noise without shocks on the second, and HSI noise with shocks on the third row. The reader is referred to Ref. 10 for a detailed description of HSI noise.

The corresponding DSPL is shown on the right column. The following can be noted: (i) When shocks are present in the signal (this being the only case that has been considered in the past), the magnitude of DSPL becomes very large. The magnitude is significantly larger than in the BVI cases considered in the previous sections, or in the case of shock-free HSI noise also shown in Fig. 11. (ii) The polarity predicts the evolution type of these non-BVI signals correctly. The polarity of the thickness noise signal and of the HSI noise after delocalization is positive. Accordingly, DSPL resembles the advancing side type. The polarity of the shock-free HSI signal is negative, and DSPL resembles the retreating side type. The latter also explains the small magnitude of DSPL despite the large negative amplitude of the noise source signal. (iii) The weighted rise frequency can be applied only for the thickness noise signal (the other two signals have either negative polarity or contain shocks). Where applicable, however, the weighted rise frequency predicts f_{start} correctly.

VI. SUMMARY

In the present work, the effect of nonlinear propagation distortion on helicopter main rotor noise was presented. The case of low-speed descent was examined, where BVI noise is the dominant noise contributor. The study was based on measured data close to the rotor and numerical calculations employing the augmented Burgers equation that predict the noise level away from the helicopter with and without nonlinear effects. It was shown that nonlinear effects in BVI noise propagation, which have been ignored in the past, can be important in certain cases. Specifically, it was shown that (i) advancing side BVI noise is affected by nonlinear distortion, while retreating side BVI noise is not, (ii) the octave frequency bands of 1000 and 2000 Hz are mostly affected, and (iii) the difference between linear and nonlinear calculations can be as high as 7 dB for the affected frequency bands. Based on the signal characteristics at source, two quantities were derived. The first quantity, termed polarity, is based on the pressure gradient of the source signal and can be used to determine whether a BVI signal will evolve as an advancing or a retreating side signal. The nonlinear evolution of advancing side BVI signals was further examined, and a second quantity, termed weighted rise time, was derived. Weighted rise time is a measure of the impulsiveness of the BVI signal at source and can be used to determine at which frequency nonlinear effects start to appear. The effects of the blade tip shape and of broadband frequency noise were also discussed in terms of polarity and weighted rise time. In addition to BVI noise, HSI noise signals (with and without shocks) and noise in the rotor plane (thickness noise combined with BVI noise) were considered and evaluated with regard to their nonlinear propagation distortion. Polarity and weighted rise time were shown to be applicable in the cases considered.

ACKNOWLEDGMENTS

The work has been partially supported by the European Integrated Project FRIENDCOPTER.

[1]C. L. Morfey and G. P. Howell, "Nonlinear propagation of aircraft noise in the atmosphere," AIAA J. **19**, 986–992 (1981).
[2]S. McInerny, K. L. Gee, M. Dowing, and M. James, "Acoustical nonlinearities in aircraft flyover data," in 13th AIAA/CEAS Aeroacoustics Conference, Rome, Italy (2007).
[3]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).
[4]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).
[5]R. L. Barger, "Theoretical prediction of nonlinear propagation effects on noise signatures generated by subsonic or supersonic propeller or rotor-blade tips," Report No. L-13388, NASA, Hampton, VA, 1980.
[6]M. P. Isom, "Acoustic shock waves generated by a transonic helicopter blade," in 36th AHS Annual Forum, Washington, D.C. (1980).
[7]K. S. Brentner and F. Farassat, "Helicopter noise prediction: The current status and future direction," J. Sound Vib. **170**, 79–96 (1994).

[8]K. S. Brentner and F. Farassat, "Modeling aerodynamically generated sound of helicopter rotors," Prog. Aerosp. Sci. **39**, 83–120 (2003).
[9]F. H. Schmitz, "Rotor noise," in *Aeroacoustics of Flight Vehicles: Theory and Practice, Vol. 1: Noise Sources*, edited by H. H. Hubbard (Acoustical Society of America, New York, 1994).
[10]F. H. Schmitz, D. A. Boxwell, W. R. Splettstoesser, and K. J. Schultz, "Model-rotor high speed impulsive noise: Full-scale comparisons and parametric variations," Vertica **8**, 395–442 (1984).
[11]Y. H. Yu, "Rotor blade-vortex interaction noise," Prog. Aerosp. Sci. **36**, 97–115 (2000).
[12]M. F. Hamilton and D. T. Blacktock, *Nonlinear Acoustics* (Academic, New York, 1998).
[13]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1994).
[14]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).
[15]Y.-S. Lee and M. F. Hamilton, "Time-domain modeling of pulsed finite-amplitude sound beams," J. Acoust. Soc. Am. **97**, 906–917 (1995).
[16]K. L. Gee, V. W. Sparrow, M. M. James, J. M. Dowing, and C. M. Hobbs, "Measurement and prediction of nonlinearity in outdoor propagation of periodic signals," J. Acoust. Soc. Am. **120**, 2491–2499 (2006).
[17]R. O. Cleveland, J. P. Chambers, H. E. Bass, R. Raspet, D. T. Blackstock, and M. F. Hamilton, "Comparison of computer codes for the propagation of sonic boom waveforms through isothermal atmospheres," J. Acoust. Soc. Am. **100**, 3017–3027 (1996).
[18]K. J. Schultz, W. Splettstoesser, B. Junker, W. Wagner, G. Arnaud, E. Schoell, and D. Fertis, "A parametric wind tunnel test on rotorcraft aerodynamics and aeroacoustics (HELISHAPE)—Test procedures and representative results," Aeronaut. J. **101**, 143–154 (1997).
[19]D. A. Boxwell, F. H. Schmitz, W. R. Splettstoesser, and K. J. Schultz, "Helicopter model rotor-blade vortex interaction impulsive noise: Scalability and parametric variations," in Tenth European Rotorcraft Forum, The Hague, The Netherlands (1984).
[20]F. H. Schmitz, D. A. Boxwell, D. A. Lewy, and C. Dahan, "Model to full-scale comparisons of helicopter blade vortex interaction noise," J. Am. Helicopter Soc. **29**, 16–25 (1984).
[21]W. R. Splettstoesser, R. Kube, W. Wagner, U. Seelhorst, A. Boutier, F. Micheli, E. Mercker, and K. Pengel, "Key results from a higher harmonic control aeroacoustic rotor test (HART)," J. Am. Helicopter Soc. **42**, 58–78 (1997).
[22]American National Standards Institute, Committee S1, Acoustics, *Method for Calculation of the Absorption of Sound by the Atmosphere, ANSI S1.26-1995* (American National Standards Institute, New York, 1995).
[23]R. M. Martin and J. C. Hardin, "Spectral characteristics of rotor blade/vortex interaction noise," J. Aircr. **25**, 62–68 (1988).
[24]O. Schneider and B. G. van der Wall, "Conditional averaging of microphone data from HART-II," Report No. IB 111-2006/37, DLR, Braunschweig, Germany, 2006.
[25]A. S. Morgans, S. A. Karabasov, A. P. Dowling, and T. P. Hynes, "Transonic helicopter noise," AIAA J. **43**, 1512–1524 (2005).

A three dimensional parabolic equation method for sound propagation in moving inhomogeneous media

Rui Cheng,[a] Philip J. Morris, and Kenneth S. Brentner
Department of Aerospace Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 9 July 2008; revised 17 April 2009; accepted 17 July 2009)

In this paper, a formulation of the Helmholtz equation for three dimensional sound propagation in a moving inhomogeneous medium in cylindrical coordinates is derived. Based on this formulation, a three dimensional parabolic equation (PE) is constructed. This PE can be used to model sound propagation in an inhomogeneous arbitrary moving medium. The method is used here to simulate three dimensional outdoor sound propagation above a rigid flat ground surface. The numerical results for two simple wind cases are presented and compared with analytical results to validate the methodology. Examples of propagation problems with more complicated wind are then included to demonstrate the importance of including the wind velocity directly in the PE method.
 [DOI: 10.1121/1.3203934]

PACS number(s): 43.28.Js, 43.28.Fp, 43.28.Gq [VWS] Pages: 1700–1710

I. INTRODUCTION

The frequency domain parabolic equation (PE) method has been used widely for noise propagation predictions due to its computational efficiency. Lentovich and Fock[1] originally introduced the PE method for electromagnetic wave propagation. Tappert[2] introduced this technique to the ocean acoustics community in the 1970s. Its main strength is the prediction of long-range, low frequency sound propagation in range-dependent environments. Considerable developments have been made over the past 3 decades and these have been reviewed by Lee *et al.*[3]

In most previous research, two dimensional (2D) models have been used to represent three dimensional (3D) acoustic propagation problems using an axisymmetric approximation. It has been assumed that the 2D models were sufficient if azimuthal variations in the medium were sufficiently weak. However, previous ocean acoustics research[4,5] has found that 3D effects should not necessarily be neglected and 3D models are required in many situations. Thus, several 3D models have been constructed for underwater acoustics research.[4–9] In these models the effect of a mean current is considered through the use of an effective sound speed (ESS) approximation, in which the current velocity component in the acoustic propagation direction is added to the local speed of sound. Yet the current velocity is treated as zero if its direction is perpendicular to the sound propagation direction. The limitation of the approach is that the ESS approach is just an approximation, which is only effective if certain conditions are satisfied,[10] even if the real current velocity is just parallel to the acoustic propagation direction. For underwater acoustics, since the sound speed is usually much larger than the magnitude of the current velocity, the ESS approximation can be effective. However, in cases where the medium velocity is a more significant fraction of the sound speed, such as in outdoor sound propagation, a PE model including the wind velocity explicitly should improve its accuracy.

Robertson *et al.*[11] gave the PEs for acoustic propagation explicitly including the current magnitude and current shear; yet it was assumed that the wind velocity only varies in the height direction. Based on the derived PE with the ESS approximation, it was identified that the current and current effects could have significant effects on sound propagation especially in the long range. (The PE method with wind velocity included explicitly was not developed). In another effort,[12] the explicit treatment of the wind velocity was applied in the fast field program (FFP) method. It was concluded that the agreement between the explicit wind velocity method and the ESS method becomes worse for long horizontal distances, especially for only a few strongly trapped modes of propagation. Meanwhile, the crosswind effect was concluded as negligible. However, in the research it was assumed that the wind velocity was constant at each vertical atmospheric layer, and that the stationary phase method and the sound propagation equations based on it are valid.

Ostashev[13,14] developed a wave equation in a moving medium with arbitrary inhomogeneities in sound speed and density. Based on this equation, a PE for 3D wave propagation in Cartesian coordinates was obtained.[15] This equation can include the wind velocity explicitly. However, long-range sound propagation for a point source involves spherical wave propagation, which cannot be represented easily in Cartesian coordinates. A PE written in cylindrical coordinates is more desirable.

In the present paper, a 3D PE for sound propagation in cylindrical coordinates for a moving inhomogeneous medium is derived. The resulting equation is then solved using the Crank–Nicholson finite difference method. Numerical simulations are conducted for several atmospheric sound propagation problems with different 3D wind profiles above a flat rigid ground surface. First, the numerical results for simple wind cases are compared with analytical results to validate the methodology. Examples of propagation prob-

[a]Present address: Advanced Rotorcraft Technology, Inc., 1330 Charleston Road, Mountain View, CA 94043.

lems with more complicated wind are then simulated to demonstrate the importance of including the wind velocity directly in the PE method.

II. GOVERNING EQUATIONS

A. The wave equation for a moving inhomogeneous medium in cylindrical coordinates

The reduced wave equation (Helmholtz equation) for a sound field $p(\mathbf{x})$ propagating in an inhomogeneous arbitrary moving medium can be written in Cartesian coordinates as[13]

$$\left[\nabla^2 + k^2(1+\epsilon) - [\nabla \ln(\rho/\rho_0)]\cdot\nabla - \frac{2i}{\omega}\frac{\partial v_i}{\partial x_j}\frac{\partial^2}{\partial x_i \partial x_j} + \frac{2ik}{c_0}\mathbf{v}\cdot\nabla\right]p(\mathbf{x}) = 0. \tag{1}$$

In this equation, Cartesian tensor notation is used, where x_1 is generally taken to be the direction of wave propagation and x_3 is the vertical or height direction. $\epsilon = c_0^2/c^2 - 1$ is the deviation from unity of the square of the refraction index in a motionless medium. Equation (1) is derived from a full set of linearized fluid dynamic equations with three assumptions: (1) internal gravity waves are neglected; (2) the mean flow is incompressible, that is, $\nabla\cdot\mathbf{v}=0$; and (3) terms of order $\mu^2 = \max(v^2/c_0^2, |v\bar{c}/c_0^2|, |v\bar{\rho}/c_0\rho_0|)$ are ignored, where $\bar{c} = c - c_0$ and $\bar{\rho} = \rho - \rho_0$ are the deviations of the sound speed and density from their mean values.[13]

Equation (1) can be transformed from Cartesian coordinates into cylindrical polar coordinates. In cylindrical coordinates (r,θ,z), r and θ define a horizontal plane. r is the range, θ is the azimuthal direction, and z is the height. The wave equation in cylindrical coordinates from Eq. (1) can be written as

$$\left[c_1\frac{\partial^2}{\partial r^2} + k^2(1+\epsilon) + c_2\frac{\partial}{\partial r} + c_3\frac{\partial}{\partial\theta} + c_4\frac{\partial^2}{\partial\theta^2} + c_5\frac{\partial^2}{\partial z^2} + c_6\frac{\partial^2}{\partial r\partial z} + c_7\frac{\partial^2}{\partial\theta\partial z} + c_8\frac{\partial^2}{\partial r\partial\theta} + c_9\frac{\partial}{\partial z}\right]p(r,\theta,z) = 0, \tag{2}$$

where $c_1, c_2, \ldots, c_9$ are relatively complicated coefficients. For example, c_1 and c_2 take the following forms:

$$c_1 = 1 - \frac{2i}{\omega}\left[(\cos\theta)^2\frac{\partial v_x}{\partial x} + (\sin\theta)^2\frac{\partial v_y}{\partial y} + \frac{\sin 2\theta}{2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right)\right], \tag{3}$$

$$c_2 = \frac{1}{r} - \frac{1}{\rho}\frac{\partial\rho}{\partial r} + \frac{2ik}{c_0}(v_x\cos\theta + v_y\sin\theta) - \frac{2i}{\omega}\left[\frac{(\sin\theta)^2}{r}\frac{\partial v_x}{\partial x} + \frac{(\cos\theta)^2}{r}\frac{\partial v_y}{\partial y} - \frac{\sin 2\theta}{2r}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right)\right]. \tag{4}$$

The exact expressions for c_3 through c_9 are given in Appendix. As can be seen, these coefficients depend on the mean velocity v_i and its derivatives. Here, the velocities and their derivatives are expressed in Cartesian coordinates since the mean media velocity is often better represented in that way. Also, these coefficients can be calculated easily before the propagation equation is solved.

If Eq. (2) is divided by c_1, we obtain

$$\left[\frac{\partial^2}{\partial r^2} + k^2(1+\eta) + \frac{c_2}{c_1}\frac{\partial}{\partial r} + \frac{c_3}{c_1}\frac{\partial}{\partial\theta} + \frac{c_4}{c_1}\frac{\partial^2}{\partial\theta^2} + \frac{c_5}{c_1}\frac{\partial^2}{\partial z^2} + \frac{c_6}{c_1}\frac{\partial^2}{\partial r\partial z} + \frac{c_7}{c_1}\frac{\partial^2}{\partial\theta\partial z} + \frac{c_8}{c_1}\frac{\partial^2}{\partial r\partial\theta} + \frac{c_9}{c_1}\frac{\partial}{\partial z}\right]p(r,\theta,z) = 0, \tag{5}$$

where $k^2(1+\eta) = k^2(1+\epsilon)/c_1$ and $\eta = (1+\epsilon-c_1)/c_1$. This is the 3D Helmholtz equation in cylindrical polar coordinates for an arbitrary moving inhomogeneous medium. The only difference between Eqs. (1) and (5) is the change in the coordinate system. Therefore, they are both based on the same three assumptions.

If there are no wind or density gradients in the medium, the expressions for the coefficients $c_1, c_2, \ldots, c_9$ can be easily reduced to the following forms:

$$c_1 = 1, \quad c_2 = \frac{1}{r}, \quad c_3 = 0, \quad c_4 = \frac{1}{r^2},$$
$$c_5 = 1, \quad c_6 = 0, \quad c_7 = 0, \quad c_8 = 0, \quad c_9 = 0. \tag{6}$$

Substituting these coefficients into Eq. (5) gives the 3D Helmholtz equation for a uniform medium at rest,

$$\left[\frac{\partial^2}{\partial r^2} + k^2 + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + \frac{\partial^2}{\partial z^2}\right]p(r,\theta,z) = 0. \tag{7}$$

Equation (7) has been used in a previous 3D propagation model (FOR3D) to describe the effects of azimuthal variations in ocean acoustics,[6] where the effect of the current velocity is added into the wavenumber k. Moreover, comparing Eqs. (7) and (5) shows that the 3D wind not only changes the coefficients of the derivative terms that exist in the truly 3D homogeneous Helmholtz equation but also introduces new cross derivative terms.

B. PE for a moving inhomogeneous medium in cylindrical coordinates

Equation (5) is the starting point for the derivation of a PE for sound propagation in a moving inhomogeneous medium in cylindrical coordinates. The approach for the new PE derivation combines the strategies used for the derivation of a homogeneous PE by Lee *et al.*[6] and for the derivation of a 3D PE in Cartesian coordinates by Ostashev *et al.*[15] First, we assume that the zeroth-order Hankel function of the second kind, $H_0^{(2)}(kr)$, is still a part of the solution of the new 3D Helmholtz equation, as in the 2D PE approach, and that it can be approximated by $\exp(ikr)/\sqrt{r}$. Setting $p = q/\sqrt{r}$ gives

$$\left[\frac{\partial^2}{\partial r^2}+k^2(1+\eta')+\left(\frac{c_2}{c_1}-\frac{1}{r}\right)\frac{\partial}{\partial r}+\left(\frac{c_3}{c_1}-\frac{1}{2r}\frac{c_8}{c_1}\right)\frac{\partial}{\partial\theta}+\frac{c_4}{c_1}\frac{\partial^2}{\partial\theta^2}+\frac{c_5}{c_1}\frac{\partial^2}{\partial z^2}+\frac{c_6}{c_1}\frac{\partial^2}{\partial r\partial z}+\frac{c_7}{c_1}\frac{\partial^2}{\partial\theta\partial z}+\frac{c_8}{c_1}\frac{\partial^2}{\partial r\partial\theta}+\left(\frac{c_9}{c_1}-\frac{1}{2r}\frac{c_6}{c_1}\right)\frac{\partial}{\partial z}\right]q=0, \tag{8}$$

where $\eta'=\eta+k^{-2}[3/(4r^2)-1/(2r)c_2/c_1]$. To simplify the above expression, we can write

$$d_1=\frac{c_2}{c_1}-\frac{1}{r},\quad d_2=\frac{c_3}{c_1}-\frac{1}{2r}\frac{c_8}{c_1},\quad d_3=\frac{c_4}{c_1},\quad d_4=\frac{c_5}{c_1},$$

$$d_5=\frac{c_6}{c_1},\quad d_6=\frac{c_7}{c_1},\quad d_7=\frac{c_8}{c_1},\quad d_8=\frac{c_9}{c_1}-\frac{1}{2r}\frac{c_6}{c_1}. \tag{9}$$

Furthermore, Eq. (8) can be expressed in the following operator notation:

$$\left(\frac{\partial^2}{\partial r^2}+k^2Q^2\right)q=0. \tag{10}$$

Here, Q is an operator including the differentiation terms in the θ and z directions, which can be written as

$$Q=(1+L)^{1/2},\quad L=F+G\frac{\partial}{\partial r}, \tag{11}$$

where

$$F=\eta'+\frac{d_2}{k^2}\frac{\partial}{\partial\theta}+\frac{d_3}{k^2}\frac{\partial^2}{\partial\theta^2}+\frac{d_4}{k^2}\frac{\partial^2}{\partial z^2}+\frac{d_6}{k^2}\frac{\partial^2}{\partial\theta\partial z}+\frac{d_8}{k^2}\frac{\partial}{\partial z} \tag{12}$$

and

$$G=\frac{d_1}{k^2}+\frac{d_5}{k^2}\frac{\partial}{\partial z}+\frac{d_7}{k^2}\frac{\partial}{\partial\theta}. \tag{13}$$

The operator Q can be represented by a Padé (1,1) approximation. This has been commonly used in previous derivations of the PE. That is,

$$Q=(1+L)^{1/2}\cong\frac{p_1+p_2L}{q_1+q_2L} \tag{14}$$

where

$$p_1=1,\quad p_2=3/4,\quad q_1=1,\quad q_2=1/4. \tag{15}$$

Equation (10) can now be split into an incoming and an outgoing wave. Here, only the outgoing wave is considered. Substituting Eq. (14) into the outgoing wave component yields

$$\frac{\partial q}{\partial r}=ik\frac{p_1+p_2(F+G\partial/\partial r)}{q_1+q_2(F+G\partial/\partial r)}q \tag{16}$$

or

$$[q_1+q_2F-ikp_2G]\frac{\partial q}{\partial r}=ik(p_1+p_2F)q-q_2G\frac{\partial^2 q}{\partial r^2}. \tag{17}$$

The right hand side of this equation contains a term involving $\partial^2 q/\partial r^2$. To obtain a PE in the range direction, this term has to be eliminated. In order to do that, Eq. (5) is written as

$$\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}+k^2+o(|v/c_0|,|\bar{c}/c_0|,|\bar{\rho}/\rho_0|)\right]p=0. \tag{18}$$

Here, the assumption that the wind velocity and sound speed deviations are small compared to the sound speed is used, and the density fluctuation is small compared to the ambient density. This assumption is reasonable for most atmospheric propagation situations. Introducing $p=q/\sqrt{r}$ into Eq. (18) gives

$$\frac{\partial^2 q}{\partial r^2}=-\left[k^2+\frac{1}{4r^2}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}\right]q. \tag{19}$$

Then, the substitution of Eq. (19) into Eq. (17) yields

$$[q_1+q_2F-ikp_2G]\frac{\partial q}{\partial r}=ik(p_1+p_2F)q+q_2G\left(k^2+\frac{1}{4r^2}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}\right)q. \tag{20}$$

Finally, with $q=\psi\exp(ikr)$, we obtain

$$[q_1+q_2F-ikp_2G]\frac{\partial\psi}{\partial r}=ik(p_1+p_2F-q_1-q_2F+ikp_2G)\psi+q_2G\left(k^2+\frac{1}{4r^2}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}\right)\psi. \tag{21}$$

Equation (21) together with Eqs. (12) and (13) represent a 3D PE in cylindrical coordinates for an arbitrary moving inhomogeneous medium.

III. NUMERICAL APPROACH

A. The Crank–Nicholson finite difference approach

To solve the resulting PE numerically, a Crank–Nicholson PE method is used in the present study for the same stability reason as for the 2D PE approach.[16] In order to facilitate the discretization, Eq. (21) is first written in an operator form,

$$A\frac{\partial\psi}{\partial r}=B\psi. \tag{22}$$

Here, A and B are operators in the z and θ directions. Applying the Crank–Nicholson finite difference algorithm in the r direction yields

$$\left(A-\frac{\Delta r}{2}B\right)\psi(r+\Delta r)=\left(A+\frac{\Delta r}{2}B\right)\psi(r). \tag{23}$$

To obtain the operators A and B, a central finite difference scheme with second-order accuracy is used,

$$\frac{\partial\psi}{\partial z}=\frac{\psi(z+\Delta z,\theta)-\psi(z-\Delta z,\theta)}{2\Delta z}, \tag{24}$$

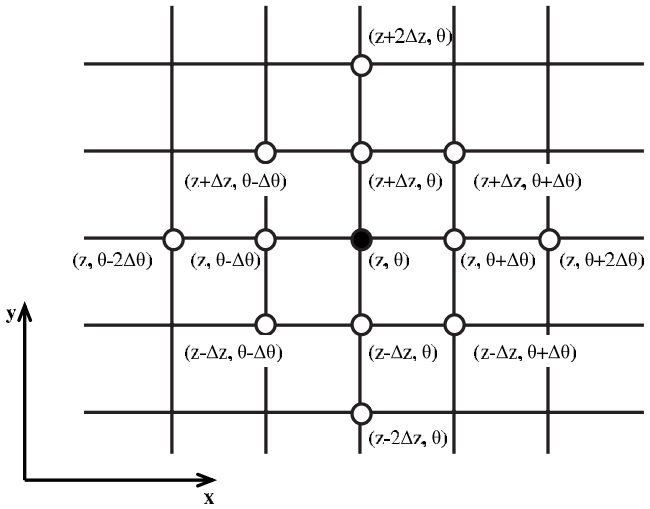


FIG. 1. The discretization strategy.

$$\frac{\partial^2 \psi}{\partial z^2} = \frac{\psi(z+\Delta z,\theta) - 2\psi(z,\theta) + \psi(z-\Delta z,\theta)}{\Delta z^2}, \tag{25}$$

$$\frac{\partial^3 \psi}{\partial z^3} = \frac{1}{2\Delta z^3}[\psi(z+2\Delta z,\theta) - 2\psi(z+\Delta z,\theta) + 2\psi(z-\Delta z,\theta) - \psi(z-2\Delta z,\theta)]. \tag{26}$$

Similar operations are applied in the θ direction and for the mixed derivative terms. After a lengthy but straightforward process, a linear system of equations can be obtained. This is the finite difference form of Eq. (23). Since third-order derivatives are included in operators A and B, five-point finite difference stencils are used in both the z and θ directions. In addition, mixed derivatives occur. Therefore, in all, there are 13 unknown variables in each of the linear system of equations. These unknown variables are $\psi(z,\theta-2\Delta\theta)$, $\psi(z-\Delta z,\theta-\Delta\theta)$, $\psi(z,\theta-\Delta\theta)$, $\psi(z+\Delta z,\theta-\Delta\theta)$, $\psi(z-2\Delta z,\theta)$, $\psi(z-\Delta z,\theta)$, $\psi(z,\theta)$, $\psi(z+\Delta z,\theta)$, $\psi(z+2\Delta z,\theta)$, $\psi(z-\Delta z,\theta+\Delta\theta)$, $\psi(z,\theta+\Delta\theta)$, $\psi(z+\Delta z,\theta+\Delta\theta)$, and $\psi(z,\theta+2\Delta\theta)$, as shown in Fig. 1. The coefficients of these variables depend on the sound speed, the density of the air, the wind speed, their gradients, and the wavenumber.

One issue is stressed here in solving the resulting linear system. To march in the range direction, it is necessary to solve a linear system of N equations with N unknown variables at each range step r. Here, the N unknown variables correspond to the acoustic pressure at each mesh point of the (z,θ) plane. It is well known that solving this system of equations requires N^3 operations if a direct solution, such as LU decomposition algorithm, is adopted. For example, if a 3D case with 2000 steps in the height direction and 181 steps in the azimuthal direction is considered, there will be $2000 \times 181 = 362\,000$ points in one cylindrical surface of the computational domain. That is, 362 000 equations corresponding to 362 000 unknown variables have to be solved. To solve such a linear system of equations requires $(362\,000)^3$ operations if the LU algorithm is used. For a 3.06 GHz CPU, this would require $(362\,000)^3/(3.06\times10^9) \cong 1.58\times10^7$ s. That is, about 183 days are required for one step of range marching. This is definitely not practical for a long-range propagation problem and it is why our preliminary calculations were limited to a very small computational domain.[17] Therefore, a more efficient solver is needed, and an iterative solver, the generalized minimum residual (GMRES) method,[18,19] is adopted in the research. Generally, an iterative algorithm requires less than N^2 calculations for a linear system with N equations. For the above computational domain setup, only 43.7 s is needed for one calculation step with the 3.06 GHz CPU. Since the GMRES method reduces the computational demands significantly, long-range marching by the 3D PE method can now be performed in a reasonable time frame.

B. The initial field

In previous research for 2D axisymmetric PE calculations, a Gaussian starting field has often been used.[2]

$$p(r=0,z) = \sqrt{ik}\exp\left[\frac{-k^2(z-z_s)^2}{2}\right]. \tag{27}$$

Here, the point source is located at $(0,0,z_s)$. This form avoids the singular point in the analytical solution if it is calculated at $r=0$. However, in the 3D PE model the coefficient of $\partial\psi/\partial\theta$ in Eq. (20) has a $1/r$ term, which could generate a potential problem for the matrix equation as $r\to0$ since a second-order central difference scheme is used. The problem is that the coefficients of $\psi(z,\theta-\Delta\theta)$ and $\psi(z,\theta+\Delta\theta)$ could be much larger than the coefficient of $\psi(z,\theta)$, which is the diagonal term of the coefficient matrix. Consequently, the resulting matrix is stiff with ensuing numerical convergence and accuracy issues. Fortunately, this problem can be avoided if the initial field is set at a finite distance away from the acoustic source, where the $1/r$ and $1/r^2$ terms are no longer large and the resulting coefficient matrix is diagonally dominant. Furthermore, since there is no singularity in this approach, the analytical solution can be used as the initial field.

The acoustic field of a point source in a uniform atmosphere, including reflection effect due to a ground surface, is given by

$$p(r,z) = \frac{\exp(ikR_1)}{R_1} + C_r\frac{\exp(ikR_2)}{R_2}. \tag{28}$$

Here, acoustic source is located at $(0,0,z_s)$, C_r is the reflection coefficient, $R_1=\sqrt{r^2+(z-z_s)^2}$, and $R_2=\sqrt{r^2+(z+z_s)^2}$. This analytical solution is only applicable to a homogeneous atmosphere, which could be used for the initial field in the following no wind example case.

The simplest moving medium sound propagation problem is that where the wind velocity is uniform. If the wind velocity is in the x-direction, the problem can be described by

$$\left(-ik + M\frac{\partial}{\partial x}\right)^2 g(\mathbf{x}|\mathbf{x}_s) - \nabla^2 g(\mathbf{x}|\mathbf{x}_s) = 2\pi\delta(\mathbf{x}-\mathbf{x}_s), \tag{29}$$

where M is the Mach number in the x-direction, k is the wavenumber, $\mathbf{x}=(x,y,z)$ is the observer location, $\mathbf{x}_s=(x_s,y_s,z_s)$ is the source location, and $g(\mathbf{x}|\mathbf{x}_s)$ is the Green's function. The solution of Eq. (29) is given by

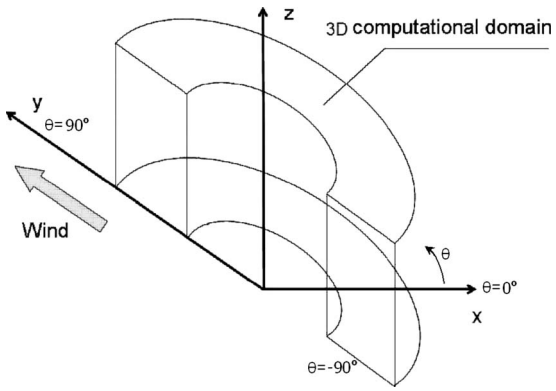


FIG. 2. The computational domain for the 3D PE calculation.

$$g(\mathbf{x}|\mathbf{x}_s) = \frac{\gamma^2}{\bar{R}} \exp[ik(\bar{R} - M\gamma^2(x - x_s))]. \quad (30)$$

Here, $\gamma = 1/\sqrt{1-M^2}$, and $\bar{R} = \gamma\sqrt{\gamma^2(x-x_s)^2 + (y-y_s)^2 + (z-z_s)^2}$. Equation (30) gives the acoustic field of a point source for a uniform wind in the x-direction. Similar to the homogeneous atmosphere case, the analytical solution for a point source in a uniform wind above a flat reflecting ground surface can be obtained by adding an image source at $(0,0,-z_s)$. This solution will also be used as an initial field later for a validation case in Sec. IV C. It has to be pointed out that the uniform wind case is just used for validation, since a uniform wind does not exist in practice due to the friction of the ground surface.

C. Boundary conditions

Besides the initial field, two kinds of boundary conditions need to be considered for a 3D computational domain: one is in the azimuthal direction and the other is in the vertical direction.

If a full circle or 360° in the azimuthal direction is considered for the computational domain, periodicity can be used for the azimuthal boundary condition. For example, the pressure field at −90° is equal to the pressure field at 270°. On the other hand, if the problem is symmetrical about the y-direction (along the −90° to 90° line) in cylindrical coordinates, as shown in Fig. 2, the 360° computational domain can be reduced to 180°, which consequently reduces the computational demands.

In setting the vertical boundary, there is a difficulty due to the second-order finite difference used for the third-order differential terms in Eq. (21). That is, the discretized PE involves $\psi(z-2\Delta z,\theta)$ and $\psi(z+2\Delta z,\theta)$ terms at the lower and upper boundaries, respectively. Consequently, two extra points are located outside of the vertical boundaries of the computational domain. This problem can be avoided if biased second-order finite difference stencils are adopted for the third-order derivative terms. They can be expressed as

$$\left.\frac{\partial^3\psi}{\partial z^3}\right|_{\text{lower boundary}} = \frac{1}{3\Delta z^3}[2\psi(z+2\Delta z,\theta) - 6\psi(z+\Delta z,\theta) + 6\psi(z,\theta) - 2\psi(z-\Delta z,\theta)], \quad (31)$$

$$\left.\frac{\partial^3\psi}{\partial z^3}\right|_{\text{upper boundary}} = \frac{1}{3\Delta z^3}[2\psi(z+\Delta z,\theta) - 6\psi(z,\theta) + 6\psi(z-\Delta z,\theta) - 2\psi(z-2\Delta z,\theta)]. \quad (32)$$

Now, only one point outside the upper and lower boundaries is needed for the discretization operations, which can be handled by using a "locally reacting" boundary condition.[20] In the research, a rigid surface with reflection coefficient of 1 (infinite impedance) is used for the ground surface in order to simplify the problem, although an outdoor ground surface's reflection coefficient is generally complex-valued and angle-dependent. For the upper boundary, a non-reflecting boundary condition is adopted at $z=z_{\max}$, with impedance equal to unity. Below this there is a buffer zone, 50 wavelengths in depth. In this buffer zone, damping is applied to the disturbances so that their amplitude is very small at the edge of the computational domain.

IV. RESULTS AND DISCUSSION

The 3D PE, Eq. (21) together with Eqs. (12) and (13), is used to simulate several example cases in this section. First, validation of the 3D PE model is conducted by comparing numerical and analytic results for a no wind and a uniform wind case. Then a logarithmic wind profile case is simulated to investigate the wind influences. At the same time, the 3D PE results are compared to the 2D PE results that use an ESS approximation for the wind velocity. Finally, a more complicated case is considered in which both the wind magnitude and direction change with height.

Unless otherwise stated, the examples shown use a unit monopole source above a flat rigid ground surface with the observer located at 1.7 m above the ground. The frequency of the acoustic source is 20 Hz, the sound speed is 340 m/s, and there are no medium density and sound speed variations throughout the computational domain.

A. Homogeneous atmosphere with no wind

In order to validate the 3D PE model, the first example case conducted is sound propagation in a homogeneous atmosphere with no wind. The acoustic source is located 3.4 m above the ground surface. Since a homogeneous atmosphere is assumed, the problem is strictly axisymmetric and there is no θ dependence. This problem can be described by Eq. (7) and it has an analytical solution given by Eq. (28). To solve this case by the 3D numerical model, both the range and height steps are chosen to be one-tenth of the sound source wavelength. Although the problem is azimuthally independent, the azimuthal direction is still included in the simulation in order to validate the new 3D PE approach. The azimuthal range is set from −90° to 90° and an azimuthal step size of 15° is used for the calculation. In general, 90° is set as the downwind direction, and −90° is the upwind direction. At 0° the wind is perpendicular to the sound propagation direction and this is called the cross wind direction as in Fig. 2. Though, in this first example, there is no wind.

Figure 3 compares the resulting real part of the acoustic pressure with the analytical results. It can be seen that the 3D PE results match the analytical solutions very well. The

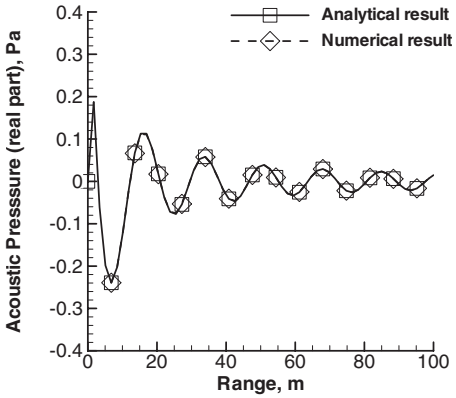


FIG. 3. Comparison of the real part of acoustic pressure from the analytical solution and the 3D PE numerical solution for the no wind case. The observer is located at $z=1.7$ m and $\theta=90°$.

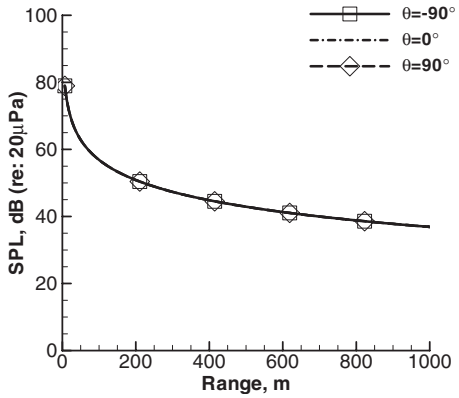


FIG. 4. Comparison of SPL at different azimuthal angles for the no wind case from a 3D PE calculation.

imaginary part of the acoustic pressure and, consequently, the resulting sound pressure level (SPL) also match between numerical results and analytical results. For brevity, these results are not shown here. Furthermore, since there is no azimuthal dependence, the 3D PE results should be independent on the azimuthal direction. This is illustrated in Fig. 4, where the three curves in different azimuthal directions are all in agreement.

B. Homogeneous atmosphere with a uniform wind

The second validation case is sound propagation in a homogeneous atmosphere with a uniform wind. The acoustic source is located at 3.4 m above the ground surface. The wind velocity is set in a direction from −90° to 90° with a magnitude of $M=0.2$. Clearly, this is not a realistic wind speed, but such a high value of M was chosen to exaggerate the effect of the wind. An analytical solution is available as given by Eq. (30). Moreover, with this equation, the initial acoustic field for the 3D PE simulation can be calculated on a cylindrical surface with a radius of 17 m, centered at the acoustic source as in Fig. 2. Since the problem is symmetrical, the semi-cylindrical computational domain is used with an azimuthal symmetry boundary condition.

Figure 5(a) compares the numerical and analytical results for the real part of the acoustic pressure in the cross wind direction, $\theta=0°$. Figure 5(b) shows a comparison of the SPL. For brevity the imaginary part of the acoustic pressure comparison is not shown here. Figure 6(a) compares the real part of the acoustic pressure for the analytical and numerical results in the upwind direction, and Fig. 6(b) conducts the same comparison in the downwind direction. As can been seen, in each case the numerical results are in good agreement with the analytical results. At the mean time, the figures show, as expected, that the wavelength is elongated in the downwind direction, and the wavelength is shortened in the upwind direction.

One issue to be addressed is the mesh resolution influence on the above calculation. For the results shown, the range and height step sizes are taken to be one-tenth of the wavelength. In the azimuthal direction, a very small step size, 0.5°, is used. This means there are 361 azimuthal steps for the 180° semi-cylindrical computational domain; that is, $m_\theta=361$, where m_θ is the total number of steps in the azimuthal direction. Note that the relative wind velocity varies with azimuthal angle in this case, and thus the azimuthal step could affect the accuracy of the numerical simulation. Figure 7 shows the influence of azimuthal mesh resolution, m_θ, on the numerical errors for the real part of the acoustic pressure. The figure shows that the finer meshes give better accuracy. A similar conclusion can be drawn from the imaginary part of the acoustic pressure. As can be seen, the $m_\theta=91$ results

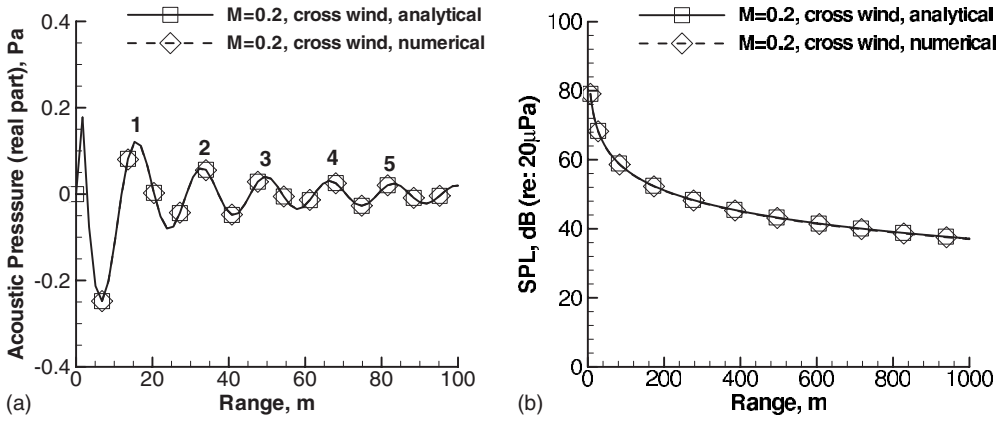


FIG. 5. Comparison of the analytical solution and the 3D PE numerical solution for the uniform wind case. The observer is located at $z=1.7$ m and $\theta=0°$. (a) Real part of the acoustic pressure and (b) SPL.

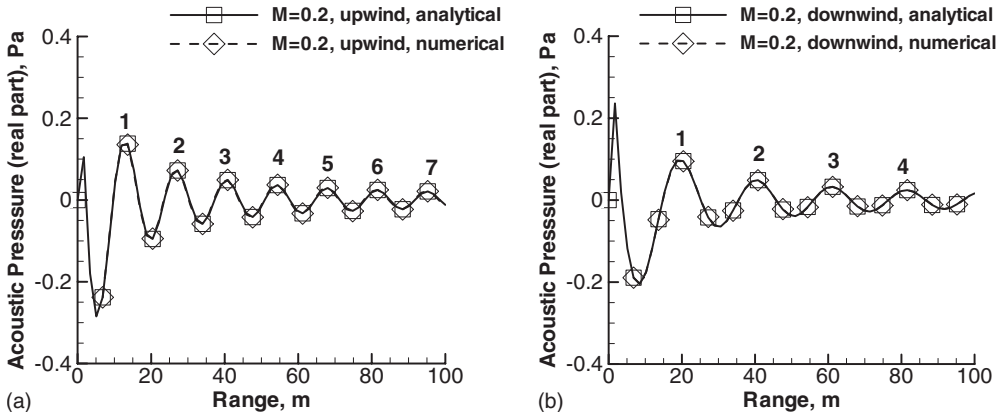


FIG. 6. Comparison of the analytical solution and the 3D PE numerical solution for the uniform wind case. The observer is located at $z=1.7$ m and $\theta=0°$. (a) In the upwind direction and (b) in the downwind direction.

are very close to the $m_\theta=361$ results. Thus, $m_\theta=91$ results with 2° resolution in azimuthal direction can be treated as the converged results, which are used for the subsequent simulations.

C. Homogeneous atmosphere with a logarithmic wind

Due to the friction of the ground surface, the wind profile is generally considered to be logarithmic in the boundary layer of the atmosphere and can be approximated in a simple formula,

$$v = A \ln(z + 1.0). \tag{33}$$

Here A is a constant, and z is the height. An example case with $A=9.424$ is modeled. The parameter gives $M \cong 0.2$ or $v=68$ m/s at the largest height (1360 m) of the computational domain. (Again, this is not a realistic wind speed, but such a high value of M was chosen to exaggerate the effect of the wind.) Since this velocity profile is relatively straightforward, it is not shown here. This case is referred to as logarithmic wind case, where the wind direction is set from −90° and 90° and only the wind magnitude varies with height. Thus, the problem is still symmetrical, and semicylindrical computational domain can be used.

No analytical solution is available for this problem. However, to set the initial field, the analytical solution, Eq. (30), for the uniform wind case is still adopted for the 3D PE calculation. The initial cylindrical surface is still located 17 m away from the acoustic source. The problem can be thought of as an initial acoustic field being propagated in an atmosphere with a logarithmic wind profile. Here, two example cases are modeled with two different source locations: $z_s=3.4$ m and $z_s=68$ m.

Figure 8 shows the resulting SPL as a function of range for the $z_s=3.4$ m case in three different azimuthal directions. The observers are located at $z=1.7$ m above the ground surface. $\theta=-90°$ is the upwind direction, $\theta=0°$ is the cross wind direction, and $\theta=90°$ is the downwind direction. Clearly, the SPL profiles are very different from each other in the different observer directions. In the downwind direction, the SPL has the largest values due to downward refraction, where peaks and valleys alternate in the SPL profile because of the multiple ground reflections. In the cross wind directions, there is no acoustic refraction in the vertical direction, and the SPL is only slightly affected by the cross wind. Multiple ground reflections do not exist either; thus the SPL profile is smooth. In the upwind direction, the SPL decreases with range due to upward refraction, and a shadow zone is formed as the range reaches about 600 m. (In an upward refracting atmosphere, a region exists where no sound rays arrive; this region is called the shadow region or shadow zone. In this

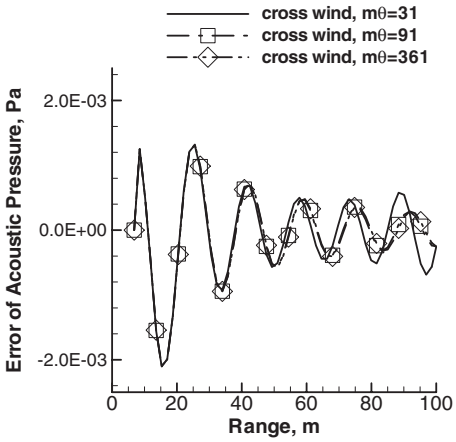


FIG. 7. Comparison of numerical errors for the real part of the acoustic pressure with different azimuthal resolution (m_θ) for the uniform wind case.

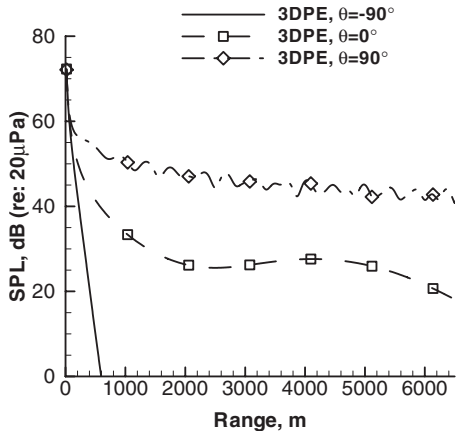


FIG. 8. Comparison of SPL for different observer azimuthal angles for the logarithmic wind case from the 3D PE calculation. Acoustic source is located at 3.4 m.

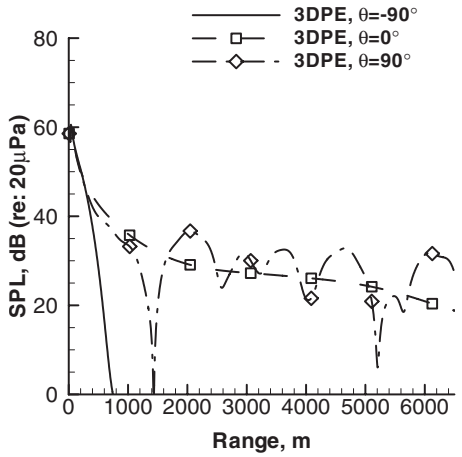


FIG. 9. Comparison of SPL for different azimuthal angles for the logarithmic wind case from the 3D PE calculation. Acoustic source is located at 68 m.

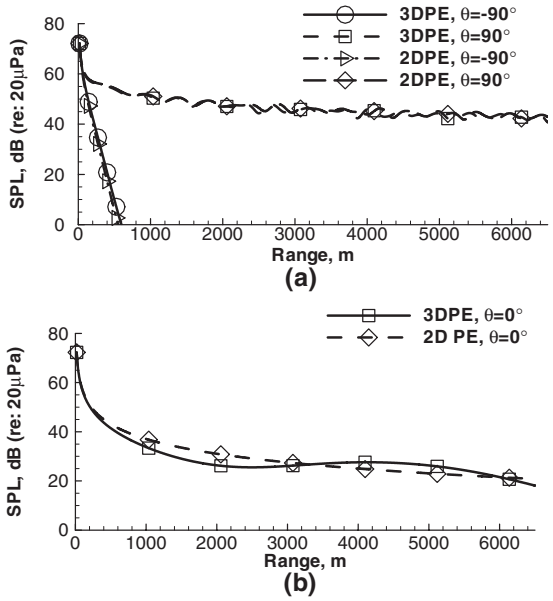


FIG. 10. SPL comparison of numerical results at $z=1.7$ m and different azimuthal locations, for calculations with the 3D PE and the 2D PE with ESS. A monopole source is located at $z_s=3.4$ m for a logarithmic wind profile.

paper, the shadow zone is taken to be the locations where the SPL is less than zero.) It should be stressed that the influence of atmospheric absorption is not considered in these simulations, but it could be added into the PE model by adding an imaginary part to the wave number.

Figure 9 shows the SPL profiles for the $z_s=68$ m case. Different from the $z_s=3.4$ m case, a shadow zone begins at a more distant range, about $r=800$ m, in the upwind direction. This is because of the larger vertical distance between the acoustic source and observer. Meanwhile, in the downwind direction, the SPL is less than in the $z_s=3.4$ m case. This is not only because the vertical distance between the source and observer increases but also because fewer ground reflections occur. Consequently, fewer peaks and valleys alternate for the $z_s=68$ m case than for the $z_s=3.4$ m case. At the observer location $\theta=0°$, the SPL for the $z_s=68$ m case is less than the $z_s=3.4$ m case at smaller ranges ($r<100$ m), because the vertical distance between acoustic source and observer is larger. Yet with an increase in range, the influence of the vertical distance between source and observer reduces, and thus the SPLs of the two cases ($z_s=68$ m case and $z_s=3.4$ m case) are similar at long ranges. One thing to be mentioned is that the downwind refraction influence increases from 0° to 90° and from 180° to 90°. On the other hand, the upwind refraction influence increases from 0° to −90° and from 180° to −90°. For brevity, these results are not shown here.

To identify the effectiveness of using ESS to model wind velocity in a 2D calculation, Fig. 10 compares the predicted SPL in different azimuthal directions for the 3D PE and the 2D PE results for the $z_s=3.4$ m case. Figure 10(a) shows the upwind and downwind comparisons. As can be seen, the differences are not more than 3 dB in either direction. However, noticeable SPL differences can be found in the cross wind direction ($\theta=0°$) in Fig. 10(b). The differences can be as large as 5 dB. Since there is no wind in the direction parallel to sound propagation at $\theta=0°$, the SPL differences can be accounted for only by the neglect of the cross wind in the 2D PE method. Figure 11 conducts the same comparison for the $z_s=68$ m case. Significant differences can be seen in the downwind direction, as shown in Fig. 11(a). The maximum difference is more than 20 dB. For observers at $\theta=0°$, small differences can be identified between 2D and 3D PE results. The maximum SPL difference is about 3 dB as in Fig. 11(b). These comparisons are similar as in Ref. 12, where "a shift in the interference pattern" was shown up as the main difference between the ESS results and the FFP method with wind velocity explicitly included. It was argued that "The difference between the two methods are most evident, when there are only a few strongly trapped modes of propagation," and "when many modes appear in

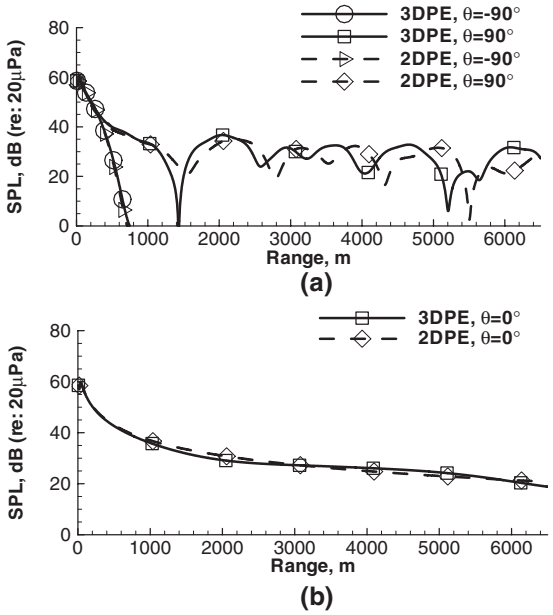


FIG. 11. SPL comparison of numerical results at $z=1.7$ m and different azimuthal locations, for calculations with the 3D PE and the 2D PE with ESS. A monopole source is located at $z_s=68$ m for a logarithmic wind profile.

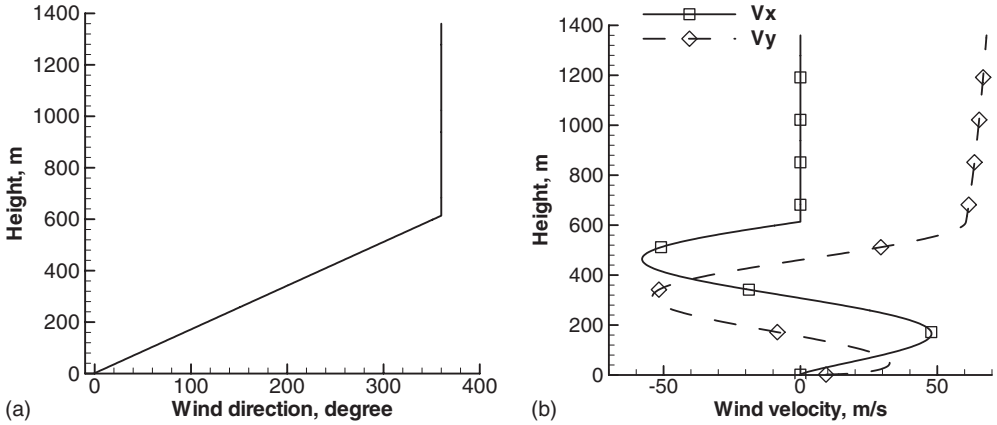


FIG. 12. Wind velocity for the rotating wind case. (a) Velocity direction and (b) velocity magnitude.

the solutions, the difference in level varies rapidly with range, but the envelopes of the solutions are quite similar." The effect is seemingly repeated here. As can be seen, there are more modes in the downwind direction of Fig. 8 than that in Fig. 9. It is reasonable since the vertical distance between source and observer is larger in the later case. Therefore, there are larger differences between the 2D and 3D results in Fig. 11(a) than that in Fig. 10(a). However, the shift in the interference pattern is not the only difference in the new results; noticeable magnitude difference can be identified, especially in Fig. 11(a). The lack of the magnitude difference in results of Li *et al.*[12] could be due to their assumption of the stationary phase method. Possibly, it is also the reason where the arguable conclusion "the crosswind component has no significant contribution" was drawn.

Furthermore, Pierce[10] gave an equation to describe the curvature of a refracted sound ray due to a wind gradient in the vertical direction on the sound propagation surface, without considering the lateral refraction caused by crosswinds,

$$r_c = \frac{c}{(dc/dz)\sin\alpha + dv_x/dz}. \tag{34}$$

Here v_x is the wind component in the sound propagation direction, and α is the incident angle between the sound ray and the vertical direction (height). It is argued that if a sound ray propagates in a nearly horizontal direction, $\sin\alpha$ is approximately unity, and the ESS is an appropriate approximation to the effects of the wind. However, if α is less than 30°, the influence of a wind-speed gradient is substantially greater than that of a sound speed gradient of the same magnitude. Under this condition, the ESS is not an appropriate approximation anymore. In the z_s=68 m case, more sound rays arrive at observer locations with large incident angles ($>30°$) than the z_s=3.4 m case. Meanwhile, there are less strong modes trapped in former than later due to the larger vertical distance. Therefore, the ESS approach for the z_s =68 m case is worse than for the z_s=3.4 m case in downwind direction. In the crosswind directions, the wind velocity is perpendicular to the sound propagation direction, and no wind velocity exists in the direction parallel to the sound propagation. The SPL difference can only come from the negligence of the cross wind in the 2D PE simulation. Particularly, at the observer height z=1.7, the wind velocity is approximately 8.3 m/s (M=0.02) and is relatively small. Thus, there is no significant SPL differences due to the lateral refraction for both z_s=3.4 m and z_s=68 m cases. For observer locations θ=−90°, since the influence of upward refraction is dominant and a shadow zone forms in a very small distance, there is not enough distance to show SPL differences between the 2D and 3D PE results.

D. Homogeneous atmosphere with a rotating wind

In all the cases considered so far, the wind velocity had a simple profile. In this subsection, a case with a more complicated 3D wind profile is considered. Here, not only the magnitude of the wind velocity changes with height as in the logarithmic wind case but also the direction of the wind changes with height. The wind direction profile is shown in Fig. 12(a), which is referred to as the rotating wind case. The resulting wind velocity components, v_x (along the 0° and 180° axes) and v_y (along the −90° and 90° axes), are shown in Fig. 12(b). For this case, a more significant wind shear is included with increased wind velocity varying with height. Notice that the problem is not symmetrical anymore. Thus, a full cylindrical computational domain has to be used with a periodic boundary condition. Again the initial field is calculated at a cylindrical surface 17 m from the acoustic source as in the logarithm wind examples. Two cases are simulated: one has the source at a height of 3.4 m and the other at a height of 68 m.

Figure 13 shows the variations in the SPL with range for three different observer directions for a source located at z_s =3.4 m, where the 2D PE results are included for comparison. Figure 13(a) shows the results at θ=−90° and θ=90°, and Fig. 13(b) shows the results at θ=0°. Since the difference between 2D and 3D PE results become significant in the long range, here only the long-range (4000–5000 m) results are shown. As can be seen, the acoustic pressure is significantly different at the different observer locations due to the wind influence. Furthermore, although the vertical distance between the sound source and observer is not large in this case, noticeable differences can be seen between the 3D and 2D results, especially at θ=90°, where much larger SPL differences are illustrated than the logarithmic wind case with the same source height as in Fig. 10. The maximum

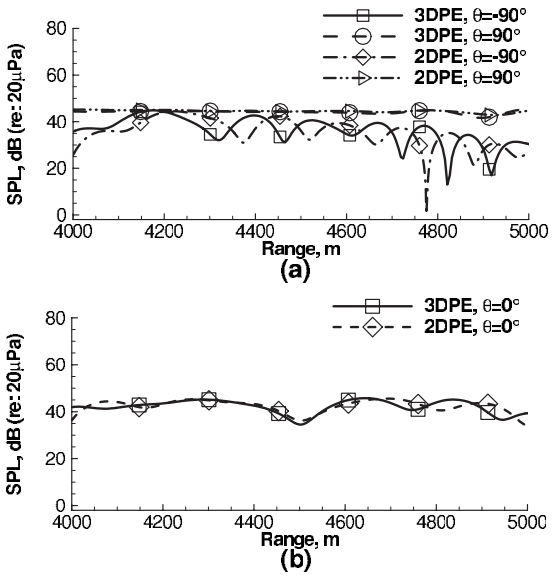


FIG. 13. SPL comparison of numerical results using 2D PE with an effective speed of sound approximation and 3D PE for a monopole source located at z_s=3.4 m and different azimuthal locations for the rotating wind case. (a) In the upwind and downwind directions and (b) in the cross wind direction.

difference is found to be more than 14 dB. It is possible that at θ=90° there are many modes trapped in the propagation; thus the results are very smooth and the 2D and 3D results look similar. At θ=−90° only two modes are trapped, and the 2D and 3D results are significantly different: At θ=0°, a few (more than two) modes are trapped, and noticeable difference is found between 2D and 3D results.

Figure 14 shows a similar comparison for a source height z_s=68 m case. Less acoustic modes are trapped in all the three propagation directions than the above case. Therefore, much more significant differences between the 3D and

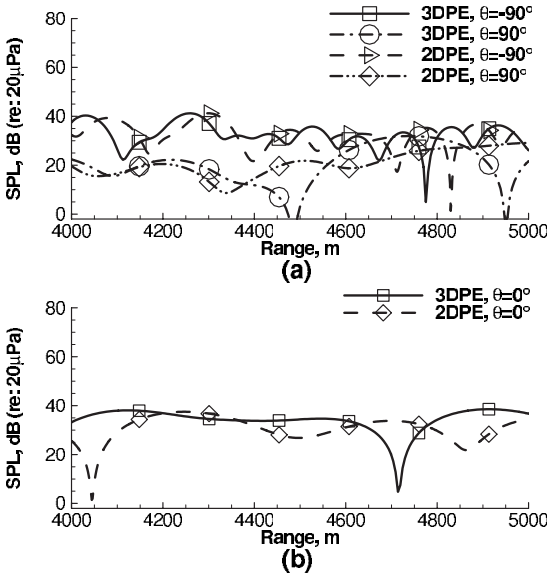


FIG. 14. SPL comparison of the numerical results using 2D PE with an effective speed of sound approximation and 3D PE for a monopole source located at z_s=68 m at different azimuthal locations for the rotating wind case. (a) In the upwind and downwind directions and (b) in the cross wind direction.

2D calculations can be noticed here than the former case, and here a difference of more than 25 dB can be found in many observer locations. Based on these comparisons, it can be concluded that, for a more complicated wind case with both direction and magnitude changes with height, the 2D PE does not predict the influence of wind on sound propagation in long distance with strong refraction accurately.

V. CONCLUSIONS

In this paper, a 3D Helmholtz equation in cylindrical coordinates is derived for an arbitrary moving inhomogeneous medium. Based on this equation, a novel 3D PE in cylindrical coordinates is developed. The PE is then discretized by the implicit Crank–Nicholson finite difference method with second-order accuracy, and the resulting system of linear equations is solved using the GMRES method with an appropriate initial field and boundary conditions.

To validate the 3D PE model, several example cases have been calculated. The results are compared to the analytical solutions for the no wind and uniform wind cases, where the numerical and analytical results match very well. Then the 3D PE model is applied to an investigation of propagation with a logarithmic wind profile. The numerical results show the atmospheric refraction due to the wind velocity gradients in height direction clearly. Then, the 3D PE results are treated as accurate solutions to identify the resulting SPL differences between the 3D PE and 2D PE calculations with an ESS approximation for wind velocity. It is illustrated that if vertical distance between the sound source and observer is small, the 2D PE errors are small. However, as the vertical distance increases, significant differences can be identified. In the downwind directions, the maximum SPL differences can be more than 20 dB. It can be concluded that as there are more sound rays arrive at observer locations with large incident angles as the vertical distance between the source and observer increases, ESS approach becomes less effective. Meanwhile, as there are less acoustic modes trapped in the larger vertical distance case, the differences between the 2D and 3D results become significant. In addition, the 3D PE has been used to simulate 3D rotating wind cases, where the wind is found to have a significant influence on sound propagation. The wind effects are more intense while the vertical distance increases, which is similar as the logarithm wind case. Based on these results, it can be concluded that for long-range sound propagation with strong 3D effects (from wind), the new 3D PE method should be used.

ACKNOWLEDGMENTS

This work is one part of a larger project entitled "Computational Aeroacoustics Analysis of Wind Turbines" sponsored by the National Renewable Energy Laboratory (Subcontract No. ZAM-3-32246-01, Paul Migliore and Patrick Moriarty, Project Managers).

APPENDIX

Expressions for the coefficients $c_3, \ldots, c_9$ in Eq. (2):

$$c_3 = -\frac{1}{\rho r^2}\frac{\partial \rho}{\partial \theta} - \frac{2i}{\omega}\left[\frac{\sin 2\theta}{r^2}\frac{\partial v_x}{\partial x} - \frac{\sin 2\theta}{r^2}\frac{\partial v_y}{\partial y} - \frac{\cos 2\theta}{r^2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right)\right], \tag{A1}$$

$$c_4 = \frac{1}{r^2} - \frac{2i}{\omega}\left[\frac{(\sin\theta)^2}{r^2}\frac{\partial v_x}{\partial x} + \frac{(\cos\theta)^2}{r^2}\frac{\partial v_y}{\partial y} - \frac{\sin 2\theta}{2r^2}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right)\right], \tag{A2}$$

$$c_5 = 1 - \frac{2i}{\omega}\frac{\partial v_z}{\partial z}, \tag{A3}$$

$$c_6 = -\frac{2i}{\omega}\left[\cos\theta\left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right) + \sin\theta\left(\frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y}\right)\right], \tag{A4}$$

$$c_7 = -\frac{2i}{\omega}\left[-\frac{\sin\theta}{r}\left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z}\right) + \frac{\cos\theta}{r}\left(\frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y}\right)\right], \tag{A5}$$

$$c_8 = -\frac{2i}{\omega}\left[-\frac{\sin 2\theta}{r}\frac{\partial v_x}{\partial x} + \frac{\sin 2\theta}{r}\frac{\partial v_y}{\partial y} + \frac{\cos 2\theta}{r}\left(\frac{\partial v_y}{\partial x} + \frac{\partial v_x}{\partial y}\right)\right], \tag{A6}$$

$$c_9 = -\frac{1}{\rho}\frac{\partial \rho}{\partial z} + \frac{2ik}{c_o}v_z. \tag{A7}$$

[1]M. A. Lentovich and V. A. Fock, "Solution of propagation of electromagnetic waves along the earth's surface by method of parabolic equations," J. Phys. (USSR) **10**, 13–23 (1946).

[2]F. D. Tappert, "The parabolic equation approximation method," in *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics Vol. **70**, edited by J. B. Keller and J. S. Papadakis (Springer-Verlag, Heildelberg, 1977).

[3]D. Lee, A. D. Pierce, and E. C. Shang, "Parabolic equation development in the twentieth century," J. Comput. Acoust. **8**, 527–637 (2000).

[4]C. F. Chen, J. J. Lin, and D. Lee, "Acoustic three-dimensional effects around the Taiwan strait: Computational results," J. Comput. Acoust. **7**, 15–26 (1999).

[5]F. Sturm and J. A. Fawcett, "On the use of higher-order azimuthal schemes in 3-D PE modeling," J. Acoust. Soc. Am. **113**, 3134–3145 (2003).

[6]D. Lee, G. Botseas, and W. L. Siegmann, "Examination of three-dimensional effects using a propagation model with azimuth-coupling capability (FOR3D)," J. Acoust. Soc. Am. **91**, 3192–3202 (1992).

[7]K. B. Smith, "A three-dimensional propagation algorithm using finite azimuthal aperture," J. Acoust. Soc. Am. **106**, 3231–3249 (1999).

[8]D. Zhu and L. Bjorno, "A three-dimensional, two-way, parabolic equation model for acoustic backscattering in a cylindrical coordinate system," J. Acoust. Soc. Am. **108**, 889–898 (2000).

[9]J. I. Arvelo and A. P. Rosenberg, "Three-dimensional effects on sound propagation and matched-field processor performance," J. Comput. Acoust. **9**, 17–39 (2001).

[10]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Melville, NY, 1989).

[11]J. S. Robertson, W. L. Siegmann, and M. J. Jacobson, "Current and current shear effects in the parabolic approximation for underwater sound channels," J. Acoust. Soc. Am. **77**, 1768–1780 (1985).

[12]Y. L. Li, M. J. White, and S. J. Franke, "New fast field programs for anisotropic sound propagation through an atmosphere with a wind velocity profile," J. Acoust. Soc. Am. **95**, 718–726 (1994).

[13]V. E. Ostashev, "Review article: Sound propagation and scattering in media with random inhomogeneities of sound speed, density and medium velocity," Waves Random Media **4**, 403–428 (1994).

[14]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (EFN SPON, London, 1997).

[15]V. E. Ostashev, D. Juve, and P. Blanc-Benon, "Derivation of a wide-angle parabolic equation for sound waves in inhomogeneous moving media," Acust. Acta Acust. **83**, 455–460 (1997).

[16]T. F. Chan, D. Lee, and L. Shen, "Stable explicit schemes for equations of the Schrödinger type," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **23**, 274–281 (1986).

[17]R. Cheng, P. J. Morris, and K. S. Brentner, "A 3D parabolic equation method for wind turbine noise propagation in moving inhomegeneous atmosphere," in 12th AIAA/CEAS Aeroacoustics Conference, Cambridge, MA 2006, AIAA Paper No. 2006-2433.

[18]V. Fraysse, L. Giraud, S. Gratton, and L. Langou, "A set of GMRES routines for real and complex arithmetics on high performance computers," Technical Report No. TR/PA/03/3, Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), Toulouse Cedex, France, 2003.

[19]http://www.cerfacs.fr/algor/softs/gmres/in-dex.html, public domain software and matrix collection by Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS) (Last viewed 7/8/2008).

[20]R. Cheng, P. J. Morris, and K. S. Brentner, "An application of the parabolic equation method to long-range propagation of wind turbine noise," in 11th AIAA/CEAS Aeroacoustics Conference, Monterey, CA (2005), AIAA Paper No. 2005-2834.

Radiative transfer theory applied to ocean bottom modeling

Jorge E. Quijano[a] and Lisa M. Zurk
Department of Electrical Engineering, Northwest Electromagnetics and Acoustics Research Laboratory, Portland State University, Portland, Oregon 97201-0751

(Received 27 October 2008; revised 30 April 2009; accepted 20 July 2009)

Research on the propagation of acoustic waves in the ocean bottom sediment is of interest for active sonar applications such as target detection and remote sensing. The interaction of acoustic energy with the sea floor sublayers is usually modeled with techniques based on the full solution of the wave equation, which sometimes leads to mathematically intractable problems. An alternative way to model wave propagation in layered media containing random scatterers is the radiative transfer (RT) formulation, which is a well established technique in the electromagnetics community and is based on the principle of conservation of energy. In this paper, the RT equation is used to model the backscattering of acoustic energy from a layered elastic bottom sediment containing distributions of independent scatterers due to a constant single frequency excitation in the water column. It is shown that the RT formulation provides insight into the physical phenomena of scattering and conversion of energy between waves of different polarizations.
 [DOI: 10.1121/1.3203992]

PACS number(s): 43.30.Ft, 43.20.Bi, 43.20.Fn, 43.30.Vh [JJM] Pages: 1711–1723

I. INTRODUCTION

Research on the interaction of acoustic energy with the sea floor has gained importance due to its role on the performance of sonar systems in shallow water and the demand for more accurate models for remote sensing of the sea floor. In the first case, a better understanding of this interaction can help to differentiate sources in the water column from bottom reverberation and improve estimates of bottom reflection loss for navy applications. In the latter case, it can enhance current techniques for inversion of geoacoustic parameters and/or provide physical understanding of the structure of the scattered field for bottom probing applications. The goal of this work is to introduce the radiative transfer (RT) formulation as an alternative method to the classic wave theory approach for computation of volume scattering.

Ocean bottom sediments are usually modeled as discrete layers with rough interfaces containing volume inhomogeneities. Acoustic backscattering is divided in rough surface scattering due to the contrast at the water-sediment interface and volume scattering due to subbottom inhomogeneities.[1] Although there are models that unify volume and rough surface scattering into a single formulation,[2,3] a simple model that handles multiple layered sediments with embedded random scatterers remains a challenge.

Several formulations based on the classic wave theory have been developed to predict sediment backscattering, and most of them are based on the *integral method*.[1,3,4] This technique consists of writing the sound speed and the density as constant mean values with random perturbations that depend on the position in the media. The resulting homogeneous wave equation for the perturbed media can be recast into a heterogeneous wave equation with a "source" term that is a function of the perturbations. With the integral method, the statement of the problem is exact,[3] but the solution of the integral equation requires approximations that might limit its range of applicability.

Examples of such approximations are the assumption of weak scatterers[2,5] or small fractional volumes.[6] Due to high sediment attenuation these assumptions simplify the mathematical formulation by considering only single scattering events. In most cases these models have shown good agreement with experimental backscattering data, but they have underestimated volume scattering in certain environments[7] and it has been suggested that this might be due to neglecting of multiple scattering.

Other common approximation in current volume scattering models is the assumption of shallow acoustic penetration, which allows to ignore contributions from deeper scattering features but it constrains the model to frequencies higher than 10 kHz and shallow grazing angles.[1] Also, most of the classic models neglect the often complicated contribution of shear waves, which is not a good approximation for certain types of consolidated elastic media.[8]

Contrary to the classic wave theory, the RT formulation works with the propagation of energy rather than the fields that carry it, and it is suggested that in general, its solution is much simpler than the integral method approach from classic wave theory.[9] RT theory has been a dynamic area of research in electromagnetic remote sensing, optic sensing, and more recently it has been adapted to acoustic waves to model the propagation of longitudinal and shear waves in elastic media, with applications in seismics[10] and ultrasound.[11] In those applications, the environment consists of parallel layers with embedded random scatterers, and this similarity with the seabed motivates this work.[12]

Despite the common presentation of RT theory as a phenomenological or heuristic statement of energy conservation, the RT theory for parallel-plane homogeneous layers with

[a]Author to whom correspondence should be addressed. Electronic mail: jorgeq@pdx.edu

random discrete scatterers can also be derived from formal principles of wave theory.[13,14] More recent developments include several approximate techniques to handle densely packed discrete scatterers,[15–17] gradients of the refractive index of background media,[18] horizontally variable layered environments,[19,20] and time-dependent broadband excitations.[21,22]

The RT equation presented in this work is the full elastic solution which includes the effect of multiple scattering, attenuation due to lossy scatterers or lossy background media, and transformation of waves of different polarizations and multiple parallel layers. Wave theory is still utilized to estimate the scattering characteristics of the random particles, but once the scattering and absorption cross sections are defined, the problem reduces to solving the RT equation. Then, the scattering media can be characterized by an extinction factor (that accounts for scattering and loss mechanisms) and an emission factor (which describes the scattering of energy from neighboring particles into the direction of observation).

In this paper, the main equations for the RT formulation for acoustics as derived by Turner and Weaver[11] are summarized and applied to layered media containing low density distributions of spherical voids. The focus of this paper is on the steady-state RT formulation, which corresponds to a monochromatic constant excitation. The notation introduced in this paper will be extended in the future to the solution of the transient RT formulation[22] required to simulate finite length pulses with the spectral characteristics commonly used in sea experiments. It is shown that the RT solution provides straightforward intuition into the physical problem and that the effect of each scattering process can be isolated.

This paper is organized as follows: Sec. II explains the scalar RT equation from a phenomenological perspective. In Sec. III, the vector RT equation utilized in acoustics is introduced for a finite layer with reflective boundaries, and Sec. IV summarizes the procedure to solve this equation. Section V presents the equations for the plane wave reflection/transmission coefficients that are utilized in this paper as boundary conditions for the water-elastic and the elastic-elastic interface. In Sec. VI, the relation between the specific intensity and the power flux is explained, and in Sec. VII simulations for different combinations of background attenuation and layering structure are presented. Section VIII corresponds to conclusions and a summary of the capabilities and limitations of the RT model for random media.

II. THE CONCEPT OF RT

The RT theory was initially formulated by astrophysicists for study of stellar spectra,[23] and several methods for the solution of the RT equation have been proposed. Some solution methods take advantage of symmetry of the random media, and in this section the RT equation is summarized for the specific case of parallel planes with discrete scatterers, as described in the literature.[11,24] With parallel-plane media, the intensity $I(\theta,\phi,z)$ propagating in the random media is a function of depth z, azimuth ϕ, and elevation angle θ.

The interaction of energy with the layered media is illustrated in Fig. 1, where the total intensity can be divided

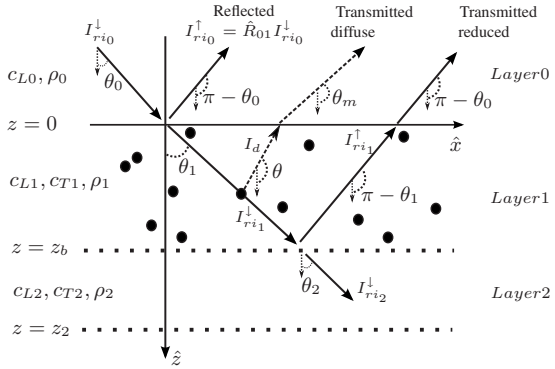


FIG. 1. Diagram of the reduced intensity I_{ri} (solid arrows) and the diffuse intensity I_d (dashed arrows) in a layer containing random scatterers. In this paper, θ_a is the diffracted angle for a wave in media a measured with respect to the $\hat{z}$ axis, with θ_0 denoting the incident angle of the energy from the source in layer 0. The reflection coefficient for specific intensity, $\hat{R}_{ab}$, corresponds to a wave traveling and bouncing back to media a when the boundary $a|b$ is reached. Layer 2 is an infinite half space with $z_2=\infty$.

into two components: the *reduced* intensity I_{ri}, which represents the coherent energy that travels with a constant direction, and the *diffuse* intensity I_d, which propagates in all directions.

The reduced intensity is attenuated due to its interaction with the scatterers and the multiple reflections in layer 1 as well as absorption within the media. From conservation laws, part of the attenuated energy will escape layer 1 and this is determined by the transmission coefficients $\hat{T}_{10}$ and $\hat{T}_{12}$ from media 1 to media 0 and 2, respectively (the subindex notation follows Tsang *et al.*[25] and the symbol $\hat{}$ refers to the transmission/reflection coefficients for specific intensities, as explained in Sec. V). Since the rest of the attenuated energy becomes the diffuse intensity, the reduced intensity can be regarded as a source of diffuse intensity, and the amount of diffuse intensity depends on the nature of the volume scatterers.

In this paper, flat interfaces are assumed and the Fresnel reflection coefficients are utilized. For a single interaction with the bottom, the total upward intensity $I(\theta,\phi,z)$ in the direction $(\theta_m > \pi/2, \phi_m)$ in layer 0 at $z=0^-$ is given by

$$I\left(\theta_m > \frac{\pi}{2}, \phi_m, 0^-\right) = \hat{T}_{10} I_d\left(\theta > \frac{\pi}{2}, \phi_m, 0^+\right) + [\hat{R}_{01} I^{\downarrow}_{\mathrm{ri}0}(\theta_0, \phi_0, 0^-) + \hat{T}_{10} I^{\uparrow}_{\mathrm{ri}1}(\pi - \theta_1, \phi_0, 0^+)]\delta(\theta_m - (\pi - \theta_0))\delta(\phi_m - \phi_0), \tag{1}$$

where (θ,ϕ) are the elevation and azimuth angles of the diffuse intensity, I_d and $I^{\uparrow}_{\mathrm{ri}1}$ are the diffuse and upward reduced intensities in layer 1, respectively, $I^{\downarrow}_{\mathrm{ri}0}$ is the incident intensity, and δ is the Kronecker operator to indicate that the coherent intensity can be observed only in the specular direction (θ_0,ϕ_0). The reduced intensity includes the reflection coefficient between layers 1 and 2, $\hat{R}_{12}$, and the attenuation of energy due to interaction with the random scatterers. Mathematically, it can be written as

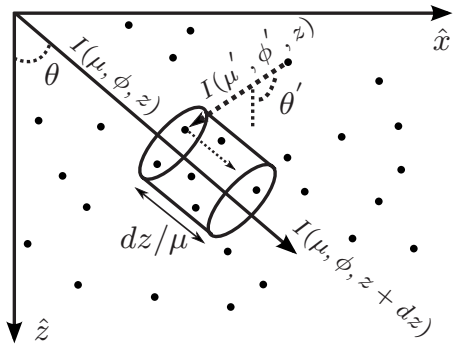


FIG. 2. Diagram of the interaction of an energy beam traveling in the (θ,ϕ) direction with scatterers within a differential volume. The intensity at the output of the volume, $I(\mu,\phi,z+dz)$, results from the attenuation of the input intensity $I(\mu,\phi,z)$ and the reinforcement due to energy coming from other directions (θ',ϕ').

$$I^{\uparrow}_{\mathrm{ri}_1}(\pi-\theta_1,\phi_0,0)=\hat{T}_{01}\hat{R}_{12}I^{\downarrow}_{\mathrm{ri}_0}(\theta_0,\phi_0,0)\exp\left[-2\frac{\eta\sigma z_b}{\cos\theta_1}\right], \tag{2}$$

where z_b is the thickness of the layer with random scatterers, η is the density of scatterers, and σ is known as the extinction coefficient, which is a characteristic of the scatterers and the background attenuation and will be defined later. The angle θ_1 results from diffraction when energy in layer 0 with an incidence angle θ_0 is transmitted to layer 1, according to Snell's law. As expected, the intensity related to volume scattering (I_d) will be zero if θ_0 is greater than the critical angle at the fluid-water interface since no energy can penetrate into layer 1.

In the RT literature it is a common practice to define the specific intensity as $I(\mu,\phi,z)$, where $\mu=\cos\theta$ is negative for upward going intensity and positive for downward going intensity. This convention will be adopted in this paper.

Figure 2 shows a differential volume containing a distribution of η scatterers per m^3.

The change in the total intensity $I(\mu,\phi,z)$ as a function of the traveled distance dz/μ and the energy that arrives from other directions (θ',ϕ') can be expressed as[11,24]

$$\frac{\partial I(\mu,\phi,z)}{\partial(z/\mu)}=-[\eta(\kappa+\nu)+2\alpha]I(\mu,\phi,z)+\frac{\eta}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}p(\mu,\phi;\mu',\phi')I(\mu',\phi',z)d\mu'd\phi', \tag{3}$$

where $I(\mu,\phi,z)=I_{\mathrm{ri}}(\mu,\phi,z)+I_d(\mu,\phi,z)$ includes the reduced and diffuse intensities, κ and ν are the scattering and absorption cross sections of a single particle normalized by 4π, respectively, α is the attenuation of the background media in Np/m, and $p(\mu,\phi;\mu',\phi')$ is the scattering strength into direction (θ,ϕ) for a single particle when the incident energy arrives from direction (θ',ϕ'). Following the definition by Turner and Weaver,[11] $\kappa=(1/4\pi)\int_{\Omega=4\pi}p(\mu,\phi;\mu',\phi')d\Omega$ where $\int_{\Omega=4\pi}\langle\cdot\rangle d\Omega$ indicates an integral over a solid angle of 4π.

The form of Eq. (3) assumes independent scattering, where the total scattering is the product $\eta(\kappa+\nu)$. If the fractional volume is large enough (typically 5%–10% or greater), the dense media radiative transfer formulation is used instead, where the total scattering is substituted by an *effective* scattering parameter.[15,16]

From Eq. (3), the intensity within the volume is decreased by $[\eta(\kappa+\nu)+2\alpha]$ and increased by the contribution of energy coming from all directions and coupling into the (θ,ϕ) direction. This is represented by the integral over all possible θ' and ϕ' angles.

The extinction cross section $\sigma=(\kappa+\nu)+2\alpha/\eta$ is a measurement of the amount of coherent energy that is transformed into diffuse energy due to scattering and absorption of the background media and the particles.

III. RT EQUATION IN ELASTIC MEDIA

In this section, the procedure developed by Turner and Weaver[11] to obtain a RT for ultrasound is outlined and modified for the case of a finite layer with reflecting boundaries at the top and bottom.

For acoustics, the specific intensity $I(\theta,\phi,z)$ can be divided in longitudinal and shear waves. From the definition of Stokes parameters, five specific intensities can be defined: I_L, I_x, and I_y represent the total energy due to longitudinal, shear horizontal, and shear vertical waves, respectively, and I_U and I_V represent correlations between the shear waves. As in Sec. II the total intensity is the summation of the reduced and the diffuse intensities. Since these intensities have different polarizations, they are orthogonal to each other and Eq. (3) can be applied to each one independently:

$$\mu\frac{\partial I_L(\mu,\phi,z)}{\partial z}=-\eta\sigma_L I_L(\mu,\phi,z)+\frac{\eta}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}\left[\sum_b P_{Lb}(\mu,\phi;\mu',\phi')\times I_b(\mu',\phi',z)\right]d\mu'd\phi' \tag{4}$$

for the longitudinal component and

$$\mu\frac{\partial I_a(\mu,\phi,z)}{\partial z}=-\eta\sigma_T I_a(\mu,\phi,z)+\frac{\eta}{4\pi}\int_{-1}^{1}\int_{0}^{2\pi}\left[\sum_b P_{ab}(\mu,\phi;\mu',\phi')\times I_b(\mu',\phi',z)\right]d\mu'd\phi' \tag{5}$$

for any of the four transversal components, where $b\in[L,x,y,U,V]$ and a is any of the shear specific intensities (x, y, U, or V). The longitudinal and shear extinction cross sections are defined as $\sigma_L=\kappa_L+\upsilon_L+2\alpha_L/\eta$ and $\sigma_T=\kappa_T+\upsilon_T+2\alpha_T/\eta$, respectively. The term $P_{ab}(\mu,\phi;\mu',\phi')$ represents the coupling of energy from a wave with b polarization into a wave with a polarization. Conceptually, it is similar to the scattering function $p(\mu,\phi;\mu',\phi')$ for the scalar case, except that $P_{ab}(\mu,\phi;\mu',\phi')$ includes the rotation of Stokes parameters about an axis, which is a step required to align the

incident and scattered intensities with an arbitrary scattering plane.[11,13]

Equations (4) and (5) have the same form, and it is convenient to define a vector RT equation that can be solved more efficiently. Also, in RT literature the variable $\tau = \eta \kappa_T z$ is defined as a normalized depth. After this substitution,

$$\mu \frac{\partial \underline{I}(\mu,\phi,\tau)}{\partial \tau} = -\tilde{\underline{\sigma}}\underline{I}(\mu,\phi,\tau) + \frac{1}{4\pi\kappa_T}\int_{-1}^{1}\int_{0}^{2\pi} \underline{P}(\mu,\phi;\mu',\phi') \times \underline{I}(\mu',\phi',\tau)d\mu' d\phi', \tag{6}$$

where the underbar denotes vector/matrix quantities and

$$\underline{I}(\mu,\phi,\tau) = \begin{pmatrix} I_L \\ I_y \\ I_x \\ U \\ V \end{pmatrix}, \tag{7}$$

$$\tilde{\underline{\sigma}} = \begin{pmatrix} \tilde{\sigma}_L & 0 & 0 & 0 & 0 \\ 0 & \tilde{\sigma}_T & 0 & 0 & 0 \\ 0 & 0 & \tilde{\sigma}_T & 0 & 0 \\ 0 & 0 & 0 & \tilde{\sigma}_T & 0 \\ 0 & 0 & 0 & 0 & \tilde{\sigma}_T \end{pmatrix}, \tag{8}$$

with $\tilde{\sigma}_L = \sigma_L / \kappa_L$ and $\tilde{\sigma}_T = \sigma_T / \kappa_L$. The term $\underline{P}(\mu,\phi;\mu',\phi')$ is the 5×5 Mueller matrix[11] for a single scatterer:

$$\underline{P}(\mu,\phi;\mu',\phi') = \begin{pmatrix} P_{LL} & P_{Ly} & P_{Lx} & P_{LU} & P_{LV} \\ P_{yL} & P_{yy} & P_{yx} & P_{yU} & P_{yV} \\ P_{xL} & P_{xy} & P_{xx} & P_{xU} & P_{xV} \\ P_{UL} & P_{Uy} & P_{Ux} & P_{UU} & P_{UV} \\ P_{VL} & P_{Vy} & P_{Vx} & P_{VU} & P_{VV} \end{pmatrix} \tag{9}$$

[the angular dependence of the elements in the matrix $\underline{P}(\mu,\phi;\mu',\phi')$ has been suppressed for brevity]. Except for the addition of the background attenuation α and the assumption of steady-state conditions, Eq. (6) is identical to Eq. (48) from Turner and Weaver.[11] In analogy to the scalar case,

$$\kappa_L = \frac{1}{4\pi}\int_{\Omega=4x} [P_{LL} + P_{xL} + P_{yL}]d\Omega,$$

$$\kappa_T = \frac{1}{8\pi}\int_{\Omega=8x} [P_{Ly} + P_{yy} + P_{xy} + P_{Lx} + P_{yx} + P_{xx}]d\Omega. \tag{10}$$

For a finite layer 1 of thickness z_b and reflecting boundaries, the boundary conditions are defined as

$$\underline{I}(\mu>0,\phi,0^+) = \underline{I}_{src} + \underline{I}_{10}(\mu>0,\phi,0^+), \quad z=0^+ \text{ downward},$$

$$\underline{I}(\mu<0,\phi,\tau_b^-) = \underline{I}_{12}(\mu<0,\phi,\tau_b^-), \quad z=z_b^- \text{ upward}, \tag{11}$$

where

$$\underline{I}_{src}(\mu,\phi,0^+) = \begin{bmatrix} \hat{T}_{01}^{LL} K_{L0}\delta(\theta-\theta_1^{L_1}) \\ \hat{T}_{01}^{Ly} K_{L0}\delta(\theta-\theta_1^{y_1}) \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{12}$$

results from the coupling of energy from the water column into the sediment. K_{L0} is the amplitude of the incident power flux at $z=0$ due to an acoustic source in the water column. This flux is assumed to be a collimated beam in the direction $(\theta_o^{L_1}, \phi_o)$ and it diffracts into longitudinal and shear specific intensities in the sediment, with amplitude determined by the specific intensity transmission coefficients $\hat{T}_{01}^{LL}$ (longitudinal-to-longitudinal) and $\hat{T}_{01}^{Ly}$ (longitudinal-to-shear vertical); there is no longitudinal-to-shear horizontal coupling, so $\hat{T}_{01}^{Lx}=0$. As a convention, the superindex in the reflection/transmission coefficients indicates the change in polarization and the subindex indicates the layers at the interface.

The angle of the diffracted waves is indicated by $\mu_1^{L_1} = \cos\theta_1^{L_1}$ and $\mu_1^{y_1} = \cos\theta_1^{y_1}$, where as in Fig. 1, the subindex corresponds to the layer. The superindex has been introduced to indicate whether the angle describes a longitudinal or a shear vertical intensity. As explained in Sec. IV, multiple interactions of the intensity with the boundaries of the layer result in the alignment of the coherent intensity along several directions of propagation that are indicated with a numerical value in the superindex. For example, $\theta_1^{L_1}$ and $\theta_1^{L_2}$ correspond both to longitudinal coherent intensities in layer 1, traveling in two different angles that are labeled as L_1 and L_2. Similarly, $\theta_1^{y_1}$ and $\theta_1^{y_2}$ are shear vertical coherent intensities in layer 1 along two angles labeled as y_1 and y_2.

The vectors I_{10} and I_{12} are the reflected specific intensities at $\tau=0$ and $\tau=\tau_b$, respectively, and are defined as

$$\underline{I}_{10}(\mu>0,\phi,0^+) = \begin{bmatrix} \hat{R}_{10}^{LL} I_L(-\mu,\phi,0^+) + \hat{R}_{10}^{yL} I_y(-\mu_1^{y_A},\phi,0^+) \\ \hat{R}_{10}^{Ly} I_L(-\mu_1^{L_A},\phi,0^+) + \hat{R}_{10}^{yy} I_y(-\mu,\phi,0^+) \\ I_x(-\mu,\phi,0^+) \\ \hat{R}_{10}^{UU} I_U(-\mu,\phi,0^+) + \hat{R}_{10}^{VU} I_V(-\mu,\phi,0^+) \\ \hat{R}_{10}^{UV} I_U(-\mu,\phi,0^+) + \hat{R}_{10}^{VV} I_V(-\mu,\phi,0^+) \end{bmatrix}, \tag{13}$$

$$\underline{I}_{12}(\mu<0,\phi,\tau_b^-) = \begin{bmatrix} \hat{R}_{12}^{LL} I_L(-\mu,\phi,\tau_b^-) + \hat{R}_{12}^{yL} I_y(\mu_1^{y_A},\phi,\tau_b^-) \\ \hat{R}_{12}^{Ly} I_L(\mu_1^{L_A},\phi,\tau_b^-) + \hat{R}_{12}^{yy} I_y(-\mu,\phi,\tau_b^-) \\ I_x(-\mu,\phi,\tau_b^-) \\ \hat{R}_{12}^{UU} I_U(-\mu,\phi,\tau_b^-) + \hat{R}_{12}^{VU} I_V(-\mu,\phi,\tau_b^-) \\ \hat{R}_{12}^{UV} I_U(-\mu,\phi,\tau_b^-) + \hat{R}_{12}^{VV} I_V(-\mu,\phi,\tau_b^-) \end{bmatrix}, \tag{14}$$

where $\hat{R}_{cd}^{ab}$ is the reflection coefficient for an incident wave with polarization a into a wave with polarization b at the boundary between media c and d. The variables $\mu_1^{L_A} = \cos\theta_1^{L_A}$ and $\mu_1^{y_A} = \cos\theta_1^{y_A}$ indicate off-axis contribution due

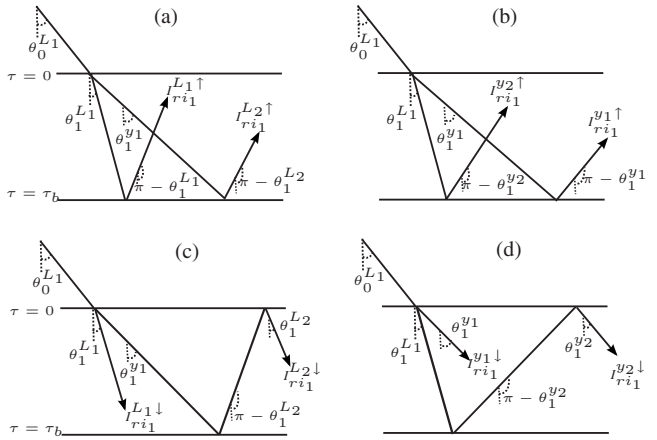


FIG. 3. The reduced intensity can be collimated in eight different directions. (a) The upward longitudinal intensities $I_{\mathrm{ri}_1}^{L_1\uparrow}$ and $I_{\mathrm{ri}_1}^{L_2\uparrow}$ result from the transformations $L{-}L{-}L$ and $L{-}y{-}L$, respectively, where L stands for *longitudinal* and y for *shear vertical*, (b) Similar to (a) for $I_{\mathrm{ri}_1}^{y_1\uparrow}$ $(L{-}y{-}y)$ and $I_{\mathrm{ri}_1}^{y_2\uparrow}$ $(L{-}L{-}y)$; (c) The downward longitudinal intensities $I_{\mathrm{ri}_1}^{L_1\downarrow}$ and $I_{\mathrm{ri}_1}^{L_2\downarrow}$ correspond to the transformations $L{-}L$ and $L{-}y{-}L{-}L$, respectively, (d) Similar to (c) for $I_{\mathrm{ri}_1}^{y_1\downarrow}$ $(L{-}y)$ and $I_{\mathrm{ri}_1}^{y_2\downarrow}$ $(L{-}L{-}y{-}y)$.

to conversion from longitudinal and shear vertical polarizations and are defined as

$$\theta_1^{L_A} = \sin^{-1}\left(\frac{c_{L1}}{c_{T1}}\sin\theta\right), \quad \theta_1^{y_A} = \sin^{-1}\left(\frac{c_{T1}}{c_{L1}}\sin\theta\right),$$

$$\theta = \cos^{-1}|\mu|. \tag{15}$$

IV. SOLUTION TO THE RT EQUATION

In general, Eq. (6) must be solved numerically but in some cases (like single scattering solutions[11]), analytical solutions can also be obtained. For any of those approaches, the reduced intensity must be found first so it can be used as a source for the diffuse intensity.

A. The reduced intensity for a finite layer

For the general case of a finite layer with reflecting boundaries, the reduced intensity has upward and downward components.[24] Due to the refraction of longitudinal waves into shear waves and vice versa, the resulting reduced intensity can be collimated along eight different angles, as shown in Fig. 3.

Each diagram in Fig. 3 represents a possible combination of multiple reflections of the energy within the sediment. The notation for the reduced intensity is similar to the notation for angles in Eq. (12): the subindex ri_1 stands for *reduced intensity in layer 1*, and the superindices $L_1\downarrow$ and $L_2\downarrow$ stand for downward longitudinal intensities collimated with two different angles (L_1 and L_2, respectively), and a similar interpretation applies to the other terms. A closed form expression for each combination of multiple reflections can be found by writing an infinite summation of terms and simplifying the resulting geometric series,[24] which is a useful result for the steady-state solution that is the subject of this paper. However, writing each of the interactions as a separate term will allow the extension of the notation in preparation for the transitory solution of the RT equation, which can be accomplished by following the procedure found in the literature[22] and it will be the subject of a future communication.

The first eight interactions of the coherent excitation with the layer boundaries are

$$I_{\mathrm{ri}_1}^{L_1\uparrow}(\mu,\phi,\tau) = A_1^{L_1\uparrow}(\mu)e^{-\tilde{\sigma}_L\tau/\mu}\delta(\theta-(\pi-\theta_1^{L_1}))\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{L_2\uparrow}(\mu,\phi,\tau) = A_1^{L_2\uparrow}(\mu)e^{-\tilde{\sigma}_L\tau/\mu}\delta(\theta-(\pi-\theta_1^{L_2}))\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{y_1\uparrow}(\mu,\phi,\tau) = A_1^{y_1\uparrow}(\mu)e^{-\tilde{\sigma}_T\tau/\mu}\delta(\theta-(\pi-\theta_1^{y_1}))\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{y_2\uparrow}(\mu,\phi,\tau) = A_1^{y_2\uparrow}(\mu)e^{-\tilde{\sigma}_T\tau/\mu}\delta(\theta-(\pi-\theta_1^{y_2}))\delta(\phi-\phi_o), \tag{16}$$

where

$$A_1^{L_1\uparrow}(\mu) = K_{L0}\hat{T}_{01}^{LL}\hat{R}_{12}^{LL}\exp\left[\frac{2\tilde{\sigma}_L\tau_b}{\mu}\right],$$

$$A_1^{L_2\uparrow}(\mu) = K_{L0}\hat{T}_{01}^{Ly}\hat{R}_{12}^{yL}\exp\left[-\tau_b\left(\frac{\tilde{\sigma}_T}{\mu_1^{y_1}}-\frac{\tilde{\sigma}_L}{\mu}\right)\right],$$

$$A_1^{y_1\uparrow}(\mu) = K_{L0}\hat{T}_{01}^{Ly}\hat{R}_{12}^{yy}\exp\left[\frac{2\tilde{\sigma}_T\tau_b}{\mu}\right],$$

$$A_1^{y_2\uparrow}(\mu) = K_{L0}\hat{T}_{01}^{LL}\hat{R}_{12}^{Ly}\exp\left[-\tau_b\left(\frac{\tilde{\sigma}_L}{\mu_1^{L_1}}-\frac{\tilde{\sigma}_T}{\mu}\right)\right] \tag{17}$$

are the amplitude terms. Similarly, the expressions for the downward reduced intensities are

$$I_{\mathrm{ri}_1}^{L_1\downarrow}(\tau,\mu,\phi) = A_1^{L_1\downarrow}(\mu)e^{-\tilde{\sigma}_L\tau/\mu}\delta(\theta-\theta_1^{L_1})\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{L_2\downarrow}(\tau,\mu,\phi) = A_1^{L_2\downarrow}(\mu)e^{-\tilde{\sigma}_L\tau/\mu}\delta(\theta-\theta_1^{L_2})\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{y_1\downarrow}(\tau,\mu,\phi) = A_1^{y_1\downarrow}(\mu)e^{-\tilde{\sigma}_T\tau/\mu}\delta(\theta-\theta_1^{y_1})\delta(\phi-\phi_o),$$

$$I_{\mathrm{ri}_1}^{y_2\downarrow}(\tau,\mu,\phi) = A_1^{y_2\downarrow}(\mu)e^{-\tilde{\sigma}_T\tau/\mu}\delta(\theta-\theta_1^{y_2})\delta(\phi-\phi_o), \tag{18}$$

where

$$A_1^{L_1\downarrow}(\mu) = K_{L0}\hat{T}_{01}^{LL},$$

$$A_1^{L_2\downarrow}(\mu) = K_{L0}\hat{T}_{01}^{Ly}\hat{R}_{12}^{yL}\hat{R}_{11}^{LL}\exp\left[-\tau_b\left(\frac{\tilde{\sigma}_T}{\mu_1^{y_1}}+\frac{\tilde{\sigma}_L}{\mu}\right)\right],$$

$$A_1^{y_1\downarrow}(\mu) = K_{L0}\hat{T}_{01}^{Ly},$$

$$A_1^{y_2\downarrow}(\mu) = K_{L0}\hat{T}_{01}^{LL}\hat{R}_{12}^{Ly}\hat{R}_{11}^{yy}\exp\left[-\tau_b\left(\frac{\tilde{\sigma}_T}{\mu_1^{L_1}}+\frac{\tilde{\sigma}_L}{\mu}\right)\right]. \tag{19}$$

The variables $\mu_1^{L_1},\mu_1^{y_1},\mu_1^{L_2},\mu_1^{y_2}$ are the cosines of the angles

$$\theta_1^{L_1} = \sin^{-1}\left[\frac{c_{L1}}{c_{L0}}\sin\theta_0^{L_1}\right], \quad \theta_1^{y_1} = \sin^{-1}\left[\frac{c_{T1}}{c_{L0}}\sin\theta_0^{L_1}\right],$$

$$\theta_1^{L_2} = \sin^{-1}\left[\frac{c_{L1}}{c_{T1}}\sin\theta_1^{y_1}\right], \quad \theta_1^{y_2} = \sin^{-1}\left[\frac{c_{T1}}{c_{L1}}\sin\theta_1^{L_1}\right], \tag{20}$$

where c_{La} and c_{Ta} are the longitudinal and shear sound speeds in the ath layer, respectively.

B. The RT equation for the diffuse specific intensity

The procedure to solve Eq. (6) consists of writing the total specific intensity as the summation of the reduced and the diffuse intensity:

$$\underline{I}(\mu,\phi,\tau) = \underline{I}_{\mathrm{ri}}(\mu,\phi,\tau) + \underline{I}_d(\mu,\phi,\tau), \tag{21}$$

where

$$\underline{I}_{\mathrm{ri}}(\mu,\phi,\tau) = \begin{bmatrix} I_{\mathrm{ri}_1}^{L_1\downarrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{L_2\downarrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{L_1\uparrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{L_2\uparrow}(\mu,\phi,\tau) \\ I_{\mathrm{ri}_1}^{y_1\downarrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{y_2\downarrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{y_1\uparrow}(\mu,\phi,\tau) + I_{\mathrm{ri}_1}^{y_2\uparrow}(\mu,\phi,\tau) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{22}$$

and

$$\underline{I}_d(\mu,\phi,\tau) = \begin{bmatrix} I_{d_1}^{L}(\mu,\phi,\tau) \\ I_{d_1}^{y}(\mu,\phi,\tau) \\ I_{d_1}^{x}(\mu,\phi,\tau) \\ I_{d_1}^{U}(\mu,\phi,\tau) \\ I_{d_1}^{V}(\mu,\phi,\tau) \end{bmatrix} \tag{23}$$

is the vector of diffuse intensities to be found. Note that for any specific angle μ only one term in the sum (22) will be non-zero due to the δ operands in Eqs. (16) and (18). Substitution of Eq. (21) into Eq. (6) results in an expression for $\underline{I}_d(\mu,\phi,\tau)$ with sources of longitudinal and transversal specific intensities:

$$\begin{aligned} \mu\frac{\partial \underline{I}_d(\mu,\phi,\tau)}{\partial\tau} &= -\tilde{\sigma}\underline{I}_d(\mu,\phi,\tau) \\ &+ \frac{1}{4\pi\kappa_T}\left[\int_{-1}^{1}\int_{0}^{2\pi}\underline{\mathrm{P}}(\mu,\phi;\mu',\phi')\underline{I}_d(\mu',\phi',\tau)d\mu'd\phi'\right] \\ &+ \underline{\mathrm{S}}_1^{L_1\downarrow}(\mu,\phi)e^{-\tilde{\sigma}_L\tau/\mu_1^{L_1}} + \underline{\mathrm{S}}_1^{L_2\downarrow}(\mu,\phi)e^{-\tilde{\sigma}_L\tau/\mu_1^{L_2}} \\ &+ \underline{\mathrm{S}}_1^{y_1\downarrow}(\mu,\phi)e^{-\tilde{\sigma}_T\tau/\mu_1^{y_1}} + \underline{\mathrm{S}}_1^{y_2\downarrow}(\mu,\phi)e^{-\tilde{\sigma}_T\tau/\mu_1^{y_2}} \\ &+ \underline{\mathrm{S}}_1^{L_1\uparrow}(\mu,\phi)e^{\tilde{\sigma}_L\tau/\mu_1^{L_1}} + \underline{\mathrm{S}}_1^{L_2\uparrow}(\mu,\phi)e^{\tilde{\sigma}_L\tau/\mu_1^{L_2}} \\ &+ \underline{\mathrm{S}}_1^{y_1\uparrow}(\mu,\phi)e^{\tilde{\sigma}_T\tau/\mu_1^{y_1}} + \underline{\mathrm{S}}_1^{y_2\uparrow}(\mu,\phi)e^{\tilde{\sigma}_T\tau/\mu_1^{y_2}}, \end{aligned} \tag{24}$$

where

$$\underline{\mathrm{S}}_1^{L1\downarrow}(\mu,\phi) = \frac{1}{4\pi\kappa_T}\underline{\mathrm{P}}(\mu,\phi;\mu_1^{L_1},\phi_o)\begin{pmatrix} A_1^{L1\downarrow}(\mu_1^{L_1}) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{25}$$

$$\underline{\mathrm{S}}_1^{L_2\downarrow}(\mu,\phi) = \frac{1}{4\pi\kappa_T}\underline{\mathrm{P}}(\mu,\phi;\mu_1^{L_2},\phi_o)\begin{pmatrix} A_1^{L_2\downarrow}(\mu_1^{L_2}) \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{26}$$

$$\underline{\mathrm{S}}_1^{y_1\downarrow}(\mu,\phi) = \frac{1}{4\pi\kappa_T}\underline{\mathrm{P}}(\mu,\phi;\mu_1^{y_1},\phi_o)\begin{pmatrix} 0 \\ A_1^{y_1\downarrow}(\mu_1^{y_1}) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{27}$$

$$\underline{\mathrm{S}}_1^{y_2\downarrow}(\mu,\phi) = \frac{1}{4\pi\kappa_T}\underline{\mathrm{P}}(\mu,\phi;\mu_1^{y_2},\phi_o)\begin{pmatrix} 0 \\ A_1^{y_2\downarrow}(\mu_1^{y_2}) \\ 0 \\ 0 \\ 0 \end{pmatrix}, \tag{28}$$

and the upward source terms $\underline{\mathrm{S}}_1^{L_1\uparrow}(\mu,\phi)$, $\underline{\mathrm{S}}_1^{L_2\uparrow}(\mu,\phi)$, $\underline{\mathrm{S}}_1^{y_1\uparrow}(\mu,\phi)$, and $\underline{\mathrm{S}}_1^{Ly_2\uparrow}(\mu,\phi)$ can be obtained from Eqs. (25)–(28) by substituting "↓" by "↑" and $\mu_1^{a_b}$ by $-\mu_1^{a_b}$ in the phase matrix $\underline{\mathrm{P}}(\mu,\phi;\mu_1^{a_b},\phi_o)$.

For the diffuse intensity in Eq. (24) the same boundary conditions as in Eqs. (13) and (14) can be used, and the solution of the diffuse intensity is outlined in the Appendix.

V. PLANE WAVE REFLECTION COEFFICIENTS FOR AN ELASTIC-ELASTIC INTERFACE

For simplicity, in this paper the plane wave reflection and transmission coefficients for elastic media are utilized to define the boundary conditions for the RT differential equation. For the water-sediment interface, the incident longitudinal wave from the water column results in a reflected longitudinal wave and transmitted longitudinal and shear vertical waves. Expressions for the reflection and transmission coefficients at the water-sediment interface were presented elsewhere.[12] For the sand-limestone interface, shear and longitudinal waves are supported in both interfaces. From Brekhovskikh,[26] the reflection/transmission coefficients for a longitudinal wave in media 1 with incidence angle θ_1^L are

$$\begin{aligned} \Delta R_{12}^{LL} &= \Upsilon_1^2 - \beta_1\Upsilon_2^2/\alpha_1 + (\alpha_2/\beta_2)(B_1^2 - \beta_1 B_2^2/\alpha_1) \\ &+ m(\beta_2/\beta_1 - \alpha_2/\alpha_1)(\kappa_1^4/4\zeta^4), \end{aligned} \tag{29}$$

$$-\Delta/2R_{12}^{Ly} = \Upsilon_1\Upsilon_2 + (\alpha_2/\beta_2)B_1B_2, \tag{30}$$

$$\Delta T_{12}^{LL} = (\kappa_2^2/\zeta^2)(\Upsilon_1 - B_2), \tag{31}$$

$$\Delta T_{12}^{Ly} = (\kappa_2^2/\zeta^2)(\Upsilon_2 + \alpha_2 B_1/\beta_2), \tag{32}$$

where

$$\Delta = \Upsilon_1^2 + \beta_1 \Upsilon_2^2/\alpha_1 + (\alpha_2/\beta_2)(B_1^2 + \beta_1 B_2^2/\alpha_1) + m(\beta_2/\beta_1 + \alpha_2/\alpha_1)(\kappa_1^4/4\zeta^4), \tag{33}$$

$$\Upsilon_1 = n^2 - mp_2/\zeta,$$

$$\Upsilon_2 = (n^2 p_1 - mp_2)/\beta_1,$$

$$B_1 = (n^2 - m)\beta_2/\zeta,$$

$$B_2 = (\beta_2/\beta_1)(n^2 p_1/\zeta - m), \tag{34}$$

and

$$n = \frac{c_{T1}}{c_{T2}}, \quad m = \frac{\rho_1}{\rho_2}, \quad \zeta = \left(\frac{\omega}{c_{L1}}\right)\sin\theta_1^L,$$

$$\beta_1 = \left(\frac{\omega}{c_{T1}}\right)\cos\theta_1^y, \quad \beta_2 = \left(\frac{\omega}{c_{T2}}\right)\cos\theta_2^y,$$

$$\alpha_1 = \left(\frac{\omega}{c_{L1}}\right)\cos\theta_1^L, \quad \alpha_2 = \left(\frac{\omega}{c_{L2}}\right)\cos\theta_2^L,$$

$$p_1 = \left[\zeta^2 - \left(\frac{\omega}{c_{T1}}\right)^2 \frac{1}{2}\right]\zeta^{-1}, \quad p_2 = \left[\zeta^2 - \left(\frac{\omega}{c_{T2}}\right)^2 \frac{1}{2}\right]\zeta^{-1}. \tag{35}$$

The angles θ_1^L, θ_1^y, θ_2^L, and θ_2^y are related by Snell's law:

$$\frac{\sin\theta_1^L}{c_{L1}} = \frac{\sin\theta_1^y}{c_{T1}} = \frac{\sin\theta_2^L}{c_{L2}} = \frac{\sin\theta_2^y}{c_{T2}}. \tag{36}$$

A vertical shear wave will also excite longitudinal and vertical shear waves in both elastic media, and it will be characterized by the corresponding R_{12}^{yL}, R_{12}^{yy}, T_{12}^{yL}, and T_{12}^{yy} coefficients.[26] Finally, a shear horizontal wave in the sediment can only excite reflected and transmitted shear horizontal waves, so it is completely characterized by two coefficients R_{12}^{xx} and T_{12}^{xx}.

Since the main quantity for the RT equation is the specific intensity, the coefficients for the specific intensity must be utilized instead. Table I shows the relationship between the plane wave coefficients and its counterpart for specific intensity in an elastic-elastic interface. For the water-sediment interface, the corresponding coefficients can be found in the literature.[27]

Table II shows the value of the geoacoustic parameters for a sandy sediment and a limestone layer.

Simulations for the power transmission and reflection coefficients from a sand-limestone interface are shown in Figs. 4(a) and 4(b). The relations $|\tilde{R}_{12}^{LL}| + |\tilde{R}_{12}^{Ly}| + |\tilde{T}_{12}^{LL}| + |\tilde{T}_{12}^{Ly}| = 1$ and $|\tilde{R}_{12}^{yL}| + |\tilde{R}_{12}^{yy}| + |\tilde{T}_{12}^{yL}| + |\tilde{T}_{12}^{yy}| = 1$ can be verified at any angle, but this relation does not hold for the specific intensity coefficients. Similar simulations for the water-sediment interface were presented in a previous paper.[12]

As discussed in Sec. IV the transmitted/reflected reduced intensity (shown in terms of transmission/reflection coefficients in Fig. 4) gives rise to additional sources of coherent upward and downward intensities that have the potential to be transformed into diffuse intensity. The complex behavior of the transmission and reflection of energy at the sand-limestone interface modulates the amplitude of those sources. This will be evident in Sec. VII when the volume scattering from a finite layer on a water-sand-limestone arrangement is compared to the scattering levels with the limestone layer replaced by sand.

TABLE I. Power ($\tilde{}$) and specific intensity ($\hat{}$) reflection/transmission coefficients for the elastic-elastic interface.

Inc.	Power coefficient	Specific intensity
L	$\tilde{R}_{12}^{LL} = \lvert R_{12}^{LL}\rvert^2$	$\hat{R}_{12}^{LL} = \tilde{R}_{12}^{LL}$
	$\tilde{R}_{12}^{Ly} = \lvert R_{12}^{Ly}\rvert^2 \dfrac{\tan\theta_1^L}{\tan\theta_1^y}$	$\hat{R}_{12}^{Ly} = \tilde{R}_{12}^{Ly} \dfrac{c_{L1}^2}{c_{T1}^2}$
	$\tilde{T}_{12}^{LL} = \lvert T_{12}^{LL}\rvert^2 \dfrac{\rho_2 \tan\theta_1^L}{\rho_1 \tan\theta_2^L}$	$\hat{T}_{12}^{LL} = \tilde{T}_{12}^{LL} \dfrac{c_{L1}^2}{c_{L2}^2}$
	$\tilde{T}_{12}^{Ly} = \lvert T_{12}^{Ly}\rvert^2 \dfrac{\rho_2 \tan\theta_1^L}{\rho_1 \tan\theta_2^y}$	$\hat{T}_{12}^{Ly} = \tilde{T}_{12}^{Ly} \dfrac{c_{L1}^2}{c_{T2}^2}$
SV	$\tilde{R}_{12}^{yy} = \lvert R_{12}^{yy}\rvert^2$	$\hat{R}_{12}^{yy} = \tilde{R}_{12}^{yy}$
	$\tilde{R}_{12}^{yL} = \lvert R_{12}^{yL}\rvert^2 \dfrac{\tan\theta_1^y}{\tan\theta_1^L}$	$\hat{R}_{12}^{yL} = \tilde{R}_{12}^{yL} \dfrac{c_{T1}^2}{c_{L1}^2}$
	$\tilde{T}_{12}^{yy} = \lvert T_{12}^{yy}\rvert^2 \dfrac{\rho_2 \tan\theta_1^y}{\rho_1 \tan\theta_2^y}$	$\hat{T}_{12}^{yy} = \tilde{T}_{12}^{yy} \dfrac{c_{T1}^2}{c_{T2}^2}$
	$\tilde{T}_{12}^{yL} = \lvert T_{12}^{yL}\rvert^2 \dfrac{\rho_2 \tan\theta_1^y}{\rho_1 \tan\theta_2^L}$	$\hat{T}_{12}^{yL} = \tilde{T}_{12}^{yL} \dfrac{c_{T1}^2}{c_{L2}^2}$
SH	$\tilde{R}_{12}^{xx} = \lvert R_{12}^{xx}\rvert^2$	$\hat{R}_{12}^{xx} = \hat{R}_{12}^{xx}$
	$\tilde{T}_{12}^{xx} = \lvert T_{12}^{xx}\rvert^2 \dfrac{\rho_2 \tan\theta_1^x}{\rho_1 \tan\theta_2^x}$	$\hat{T}_{12}^{xx} = \tilde{T}_{12}^{xx} \dfrac{c_{T1}^2}{c_{T2}^2}$

VI. CONSERVATION OF POWER FOR AN INCIDENT COHERENT BEAM

One of the main features of the RT model is the conservation of power. This can be used to test the accuracy of the RT computer simulation and to provide the contribution of longitudinal and shear waves in units of power rather than specific intensity. The relationship between specific intensity $I_{f_g}^{e\uparrow}(\mu, \phi, z)$ and the upward power flux $F_{f_g}^{e\uparrow}(z)$ normal to layer g at any depth z is[13]

TABLE II. Acoustic properties of the sediment, limestone, and water column used in this paper.

Variable	Sediment[a]	Water	Limestone[b]
ρ_s (kg/m^3)	2023.2	1027	2200
c_L (m/s)	1689	1500	4390
c_T (m/s)	117	0	2570
α_L (dB/m at 10 kHz)	5.24	0	0.2
α_T (dB/m at 10 kHz)	178	0	0.7

[a]References 28 and 29.
[b]References 30 and 31.

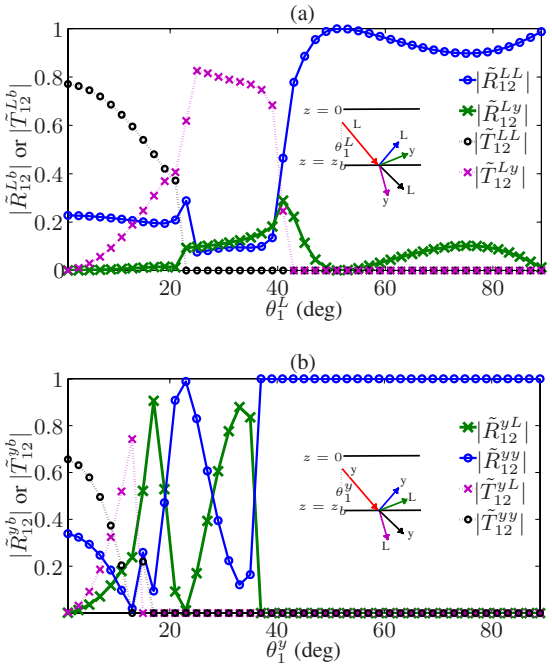


FIG. 4. (Color online) Power reflection/transmission coefficients at the sand-limestone interface. (a) Coefficients $\tilde{R}_{12}^{LL}$, $\tilde{R}_{12}^{Ly}$, $\tilde{T}_{12}^{LL}$, and $\tilde{T}_{12}^{Ly}$ for an incident longitudinal wave, as shown in the inserted diagram. (b) Coefficients $\tilde{R}_{12}^{yy}$, $\tilde{R}_{12}^{yL}$, $\tilde{T}_{12}^{yy}$, and $\tilde{T}_{12}^{yL}$ for an incident shear vertical wave, as shown in the inserted diagram.

$$F_{f_g}^{e\uparrow}(z) = \int_{2\pi} I_{f_g}^{e\uparrow}(\mu,\phi,z)\mu d\Omega, \quad -1 < \mu < 0, \tag{37}$$

where $d\Omega$ is a solid angle and $0<\phi<2\pi$, the subindex f refers to reduced (ri) or diffuse (d) and the superindex e could be L_1, L_2, y_1, etc. Similarly, the downward power flux normal to layer g at any depth z is

$$F_{f_g}^{e\downarrow}(z) = \int_{2\pi} I_{f_g}^{e\downarrow}(\mu,\phi,z)\mu d\Omega, \quad 0 < \mu < 1. \tag{38}$$

As an example,

$$F_{\mathrm{ri}_1}^{L_1\uparrow}(z) = \int_{2\pi} I_{\mathrm{ri}_1}^{L_1\uparrow}(\mu,\phi,z)\mu d\Omega, \quad -1 < \mu < 0 \tag{39}$$

is the upward power flux due to the coherent longitudinal specific intensity $I_{\mathrm{ri}_1}^{L_1\uparrow}(\mu,\phi,z)$, while

$$F_{d_1}^{L\downarrow}(z) = \int_{2\pi} I_{d_1}^{L\downarrow}(\mu,\phi,z)\mu d\Omega, \quad 0 < \mu < 1 \tag{40}$$

is the downward flux due to the diffuse specific intensity $I_{d_1}^{L\downarrow}(\mu,\phi,z)$. If the attenuation of the background media in layer 1 is zero and the scatterers are lossless, the conservation of the normal component of power flux can be stated as

$$\begin{aligned}
&(F_{\mathrm{ri}_0}^{L_1\downarrow} - F_{\mathrm{ri}_0}^{L_1\uparrow})|_{z=0} && \text{(incident-reflected)}\\
&= (F_{d_1}^{L\uparrow} + F_{d_1}^{T\uparrow})|_{z=0} + (F_{d_1}^{L\downarrow} + F_{d_1}^{T\downarrow})|_{z=z_b} && \text{(diffuse/volume scat.)}\\
&+ (F_{\mathrm{ri}_1}^{L1\uparrow} + F_{\mathrm{ri}_1}^{L2\uparrow})|_{z=0} + (F_{\mathrm{ri}_1}^{L1\downarrow} + F_{\mathrm{ri}_1}^{L2\downarrow})|_{z=z_b} && \text{(coherent } L\text{)}\\
&+ (F_{\mathrm{ri}_1}^{y1\uparrow} + F_{\mathrm{ri}_1}^{y2\uparrow})|_{z=0} + (F_{\mathrm{ri}_1}^{y1\downarrow} + F_{\mathrm{ri}_1}^{y2\downarrow})|_{z=z_b} && \text{(coherent } T\text{)},
\end{aligned} \tag{41}$$

where

$$F_{\mathrm{ri}_0}^{L_1\downarrow}(z) = \int_{2\pi} I_{\mathrm{ri}_0}^{L_1\downarrow}(\mu,\phi,z)\cos\theta d\Omega,$$

$$F_r^{L_1\uparrow}(z) = \int_{2\pi} I_{\mathrm{ri}_0}^{L_1\uparrow}(\mu,\phi,z)\cos\theta d\Omega \tag{42}$$

are the normal fluxes due to the coherent incident and reflected specific intensities in the water column. The relation in Eq. (41) states that all energy provided to the sediment must manifest itself in either diffuse or upward/downward reduced intensities. The incident and reflected specific intensities are defined as

$$I_{\mathrm{ri}_0}^{L_1\downarrow}(\mu,\phi,\tau) = K_{L0}\delta(\theta-\theta_0^{L_1})\delta(\phi-\phi_0),$$

$$I_{\mathrm{ri}_0}^{L_1\uparrow}(\mu,\phi,\tau) = \hat{R}_{01}^{LL}K_{L0}\delta(\theta-(\pi-\theta_0^{L_1}))\delta(\phi-\phi_0). \tag{43}$$

In Sec. VII, simulated results are presented in terms of the normal power fluxes in Eq. (41).

VII. RESULTS

Simulations were run for $K_{L0}=1$ W/m^2 with different combinations of sediment attenuation and boundary conditions. The scatterers are spherical cavities, and the scattering terms in Eq. (9) were computed using the Mie analytical solution [see Eq. (86) from Turner and Weaver[27]]. In all simulations, the frequency $f=10$ kHz, the particle radius $a=0.01$ m, $z_b=1$ m, and the particle density $\eta=2388$ scatterers/m^3.

Figures 5 and 6 were computed to show the conservation of the outward power flux that is normal to the boundaries of the layer containing scatterers when $\sigma_L=\kappa_L$ and $\sigma_T=\kappa_T$ (no background attenuation or scatterer absorption). In both cases, the conservation of the normal power flux established in Eq. (41) can be confirmed at each incidence angle $\theta_0^{L_1}$.

Comparison between Figs. 5 and 6 also shows the effect of the contrast between layer 1 and the bottom halfspace. Figure 5 corresponds to a sand bottom half space (no contrast between layers 1 and 2), while Fig. 6 shows the resulting power flux when the limestone bottom half space is introduced.

In Fig. 5, the selection of the same background material for layers 1 and 2 results in the suppression of the upward coherent specific intensities since $R_{12}^{LL}=R_{12}^{Ly}=0$, and for this simulation, the amount of volume scattering coming out of the layer at $z=z_b$, $(F_{d_1}^{L\downarrow}+F_{d_1}^{T\downarrow})|_{z=z_b}$, is higher than the volume scattering at $z=0$, $(F_{d_1}^{L\uparrow}+F_{d_1}^{T\uparrow})|_{z=0}$. The critical angle for the longitudinal energy that couples into the sediment is 62°, and above this angle the volume scattering is negligible. Because

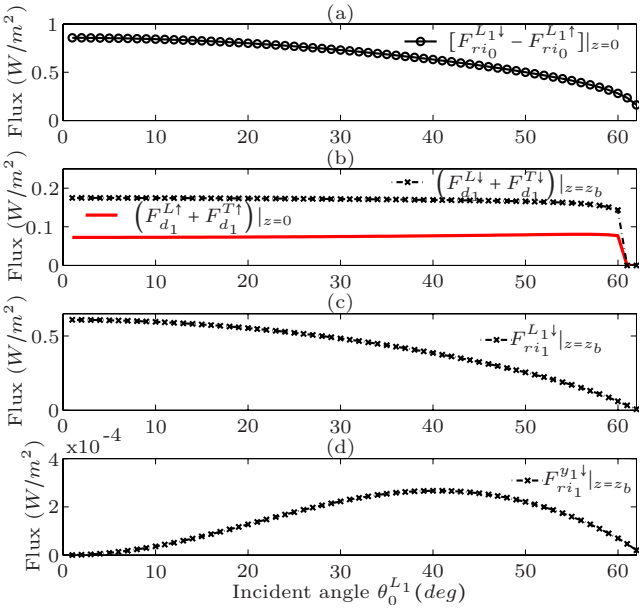


FIG. 5. (Color online) Normal component of the outgoing power flux in layer 1 for a configuration water-sand-sand and no background attenuation. (a) The difference $F^{L_1\downarrow}_{\mathrm{ri}_0} - F^{L_1\uparrow}_{\mathrm{ri}_0}$ at $z=0$ indicates the power flux that is transmitted into layer 1 from the water column. (b) Total upward and downward diffuse power fluxes at $z=0$ and $z=z_b$ m, respectively, (c) and (d) show the power fluxes from coherent longitudinal and shear vertical energies, respectively [note different scale on (d)]. Due to the transparent boundary condition at $z=z_b$, $F^{L_1\uparrow}_{\mathrm{ri}_1}$, $F^{L_2\uparrow}_{\mathrm{ri}_1}$, $F^{L_2\downarrow}_{\mathrm{ri}_1}$, $F^{y_1\uparrow}_{\mathrm{ri}_1}$, $F^{y_2\uparrow}_{\mathrm{ri}_1}$, and $F^{y_2\downarrow}_{\mathrm{ri}_1}$ are zero.

the shear sound speed in the sediment is smaller than the sound speed in the water, there is always some minimum amount of transversal energy penetrating into the finite layer. Nevertheless, the total contribution to volume scattering related to the shear polarization is small [note scale in Fig. 5(d)] since the transmission coefficient from the water column into the sediment, T^{Ly}_{01}, is very small compared to T^{LL}_{01}.

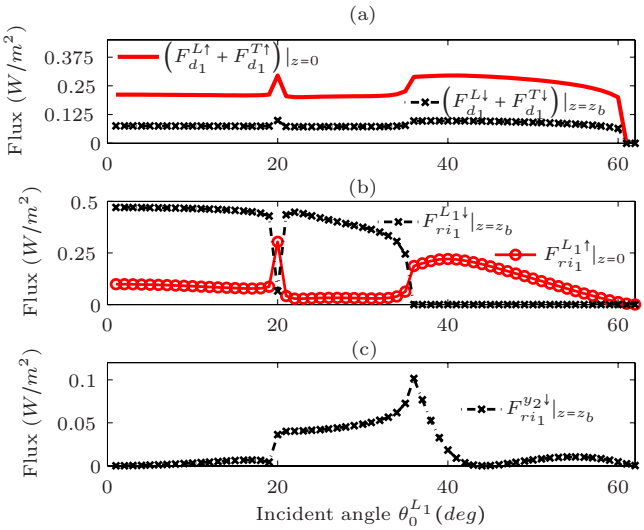


FIG. 6. (Color online) Normal component of the outgoing power flux in layer 1 for a configuration water-sand-limestone and no background attenuation. (a) Total upward and downward diffuse power fluxes at $z=0$ and $z=z_b$ m, respectively. The diffuse power flux has more variability with respect to Fig. 5(b) due to the reflection/transmission coefficients at the sand-limestone interface, (b) and (c) show the power fluxes from coherent longitudinal and shear vertical energies, respectively ($F^{L_2\uparrow}_{\mathrm{ri}_1}$, $F^{L_2\downarrow}_{\mathrm{ri}_1}$, $F^{y_1\uparrow}_{\mathrm{ri}_1}$, $F^{y_2\uparrow}_{\mathrm{ri}_1}$, and $F^{y_1\downarrow}_{\mathrm{ri}_1}$ are negligible). Note that $F^{L_1\downarrow}_{\mathrm{ri}_0} - F^{L_1\uparrow}_{\mathrm{ri}_0}$ at $z=0$ (not shown) is the same as in Fig. 5.

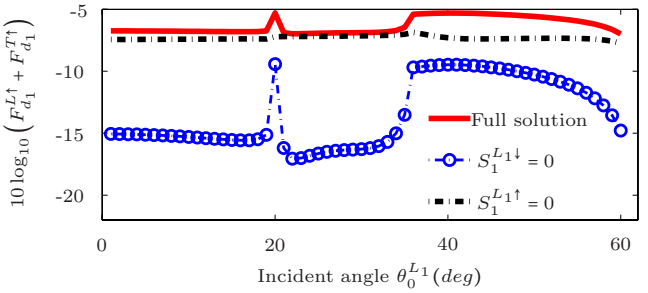


FIG. 7. (Color online) Computation of $(F^{L\uparrow}_{d_1} + F^{T\uparrow}_{d_1})|_{z=0}$ (upward volume scattering) with the sources $S_1^{L_1\uparrow}$ and $S_1^{L_1\downarrow}$ set to zero, as indicated. The full solution corresponds to $(F^{L\uparrow}_{d_1} + F^{T\uparrow}_{d_1})|_{z=0}$ from Fig. 6(a), in which all eight sources were included. See text for discussion.

In Fig. 6, the high acoustic contrast between sand and limestone results in a complex structure for R^{LL}_{12} and R^{Ly}_{12}, as shown in the reflection-transmission coefficients in Figs. 4(a) and 4(b), and the dominant coherent excitations are $I^{L_1\downarrow}_{\mathrm{ri}_1}$, $I^{L_1\uparrow}_{\mathrm{ri}_1}$, and $I^{y_2\downarrow}_{\mathrm{ri}_1}$. Note that since T^{Ly}_{01} is small for the water-sediment interface,[12] the coherent excitations $I^{L_2\uparrow}_{\mathrm{ri}_1}$, $I^{y_1\uparrow}_{\mathrm{ri}_1}$, $I^{L_2\downarrow}_{\mathrm{ri}_1}$, and $I^{y_1\downarrow}_{\mathrm{ri}_1}$ are negligible.

The upward source term $S_1^{L_1\uparrow}$ results in an increase in the volume scattering at $z=0$ with respect to the simulation in Fig. 5. Some of the features of these coefficients can be observed in the diffuse power flux in Fig. 6(a). For example, the sharp increase in the power flux $(F^{L\uparrow}_{d_1} + F^{T\uparrow}_{d_1})|_{z=0}$ at the incidence angle $\theta_0^{L_1}=20°$ is caused by the transformation from coherent into diffuse energy of the source $I^{L_1\uparrow}_{\mathrm{ri}_1}$. The amplitude of this source is determined by the reflection coefficient R^{LL}_{12}, which exhibits a peak at $\theta_1^{L_1}=22.7°$ [see Eq. (20) and Fig. 4(a)].

The contribution of each of the coherent sources can also be explored by comparing the full solution of the RT equation in Eq. (24) to the solution obtained by forcing the selected coherent source to zero. For example, Fig. 7 shows the outgoing diffuse flux $(F^{L\uparrow}_{d_1} + F^{T\uparrow}_{d_1})|_{z=0}$ when source terms $S_1^{L_1\uparrow}$ or $S_1^{L_1\downarrow}$ set to zero, compared to the full solution taken from Fig. 6(a). When $S_1^{L_1\downarrow}$ is zeroed, most of the power decreases, but the sharp peak at $\theta_0^{L_1}=20°$ is still present. This indicates that most of the volume scattering is caused by the coherent intensity $I^{L_1\downarrow}_{\mathrm{ri}_1}$, while $I^{L_1\uparrow}_{\mathrm{ri}_1}$ contributes mostly at $\theta_0^{L_1}=20°$ and for $\theta_0^{L_1}>36°$, which is when the reflection coefficient R^{LL}_{12} starts to pick up at $\theta_1^{L_1}=45°$. When $S_1^{L_1\uparrow}$ is forced to zero, most of the power remains the same except at the angles previously indicated.

A similar exploration can be done by removing the sources of coherent shear vertical intensity $S_1^{y_2\uparrow}$ or $S_1^{y_2\downarrow}$ with the result shown in Fig. 8. From this figure one can conclude that the main contribution of the shear vertical coherent intensity to the total diffuse power $(F^{L\uparrow}_{d_1} + F^{T\uparrow}_{d_1})|_{z=0}$ is between $\theta_0^{L_1}=21°$ and $\theta_0^{L_1}=39°$. Even though for this particular simulation this contribution is small, it illustrates the importance of considering shear propagation since it could be relevant for other kinds of background media such as sedimentary rock.[8]

The volume scattering levels observed in Figs. 5 and 6 are unrealistic for a field measurement due to the lack of background attenuation. Figure 9 includes the attenuation

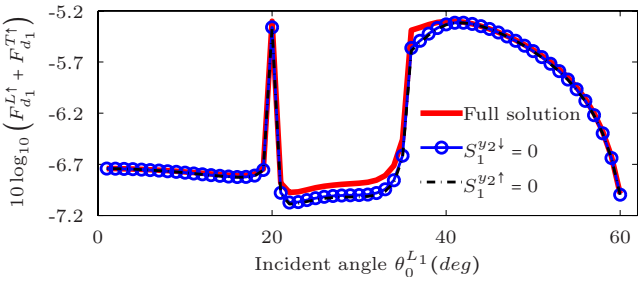


FIG. 8. (Color online) Similar to Fig. 7, with the shear vertical source terms $S_1^{y_2\uparrow}$ and $S_1^{y_2\downarrow}$ set to zero, as indicated. Again, the full solution corresponds to $(F_{d_1}^{L\uparrow}+F_{d_1}^{T\uparrow})|_{z=0}$ from Fig. 6(a).

from Table I for the sand layer, and this strongly reduces the amount of volume scattering as it would be expected with experimental data. In this simulation, the effect of the attenuation on the source $S_1^{L_1\uparrow}$ is stronger than for $S_1^{L_1\downarrow}$ because of a larger traveling path through the finite layer. This can be observed by comparing $F_{\mathrm{ri}1}^{L_1\uparrow}$ and $F_{\mathrm{ri}1}^{L_1\downarrow}$ in Fig. 9 with respect to Fig. 6. Also, the attenuation smoothes some of the sharp features in $F_{d_1}^{L\uparrow}+F_{d_1}^{T\uparrow}$ and $F_{d_1}^{L\downarrow}+F_{d_1}^{T\downarrow}$, introduced by the reflection off the sand-limestone interface.

VIII. CONCLUSION

The transport equation for acoustics has been solved for a finite layer with random scatterers and fully elastic flat boundaries, allowing multiple conversions of longitudinal and shear coherent energy. Each interaction of the coherent energy with the boundaries of the finite layer can be used as a source for diffuse intensity, which increases the amount of volume scattering. The RT equation has been extended to include eight sources that represent all possible propagation angles for the reduced intensity when flat boundaries are assumed. The preliminary simulations presented in this paper show the flexibility of the RT formulation in isolating and interpreting the effect of geoacoustic parameters such as layering structure and background attenuation.

In this paper, the results provided by the RT theory in units of specific intensity have been converted to power flux, which is required for further comparison with experimental data. The conservation of the normal component of this flux was demonstrated with simulations of the propagation of energy within a lossless finite layer of sand.

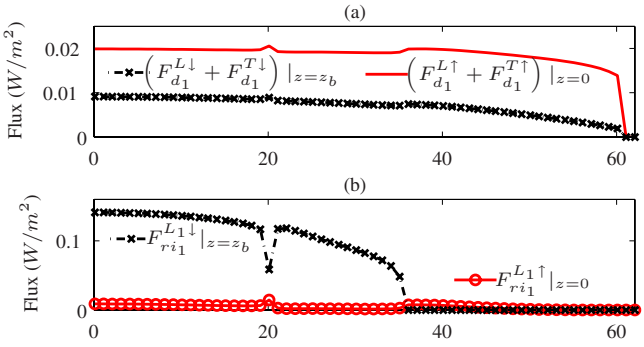


FIG. 9. (Color online) Simulation similar to Fig. 6, with attenuation $\alpha_L=5.24$ and $\alpha_T=178$ dB/m in the background sand layer. Due to a longer traveling path, $F_{\mathrm{ri}_1}^{L_1\uparrow}$ is more strongly affected by the sediment attenuation than $F_{\mathrm{ri}_1}^{L_1\downarrow}$. The downward coherent power flux $F_{\mathrm{ri}_1}^{y_2\downarrow}|_{z=z_b}$ (not shown) is negligible due to the high shear attenuation α_T.

For simplicity, the RT formulation presented in this work corresponds to an environment of parallel layers of infinite extent with no lateral variation in statistical properties in the x-y plane. Some of the features of the RT formulation can be appreciated under this ideal scenario, such as the incorporation of multiple discrete layers with random discrete scatterers and transformations of longitudinal and shear polarizations at the boundaries of the layers and at the scatterers. More complex scenarios of finite layers with horizontal variability have already been considered in other disciplines of electromagnetic remote sensing using the RT approach and will be used to extend the proposed model.

The simulations presented here correspond to single frequency and steady-state excitations. Nevertheless, the RT formulation can also handle arbitrary broadband excitations by means of the transient RT solution. Research on the transitory behavior of the RT is still on development in electromagnetics[21,32] and acoustics[22] applications. The solution of the transient RT equation is interpreted as the response of the random media to a very short impulse of energy,[22] and it is argued that the solution to more complicated excitations such as linear chirps can be obtained by convolution with this "impulse" response. The extension of the presented model to transitory excitations is the subject of future research.

ACKNOWLEDGMENTS

This project has been sponsored by the Office of Naval Research (ONR), Grant No. N000140510886.

APPENDIX: SOLUTION OF THE RT EQUATION BY A NUMERICAL METHOD

Equation (24) can be solved by applying a Fourier azimuthal decomposition in $2M+1$ terms of the Mueller matrix and the specific intensity. The corresponding Fourier series representation is

$$\underline{\mathrm{P}}_m(\mu,\mu_o)=\frac{1}{2\pi}\int_0^{2\pi}\underline{\mathrm{P}}(\mu,\mu_0,\phi-\phi_o)e^{im(\phi-\phi_o)}d(\phi-\phi_o),$$

$$\underline{I}_{dm}(\mu,\tau)=\frac{1}{2\pi}\int_0^{2\pi}\underline{I}_d(\mu,\phi-\phi_o,\tau)e^{im(\phi-\phi_o)}d(\phi-\phi_o), \tag{A1}$$

which allows the representation of the source terms as

$$\underline{\mathrm{S}}_{1m}^{L_1\downarrow}(\mu)=\frac{1}{4\pi\kappa_T}\underline{\mathrm{P}}_m(\mu,\mu_1^{L_1})\begin{pmatrix}A_1^{L_1\downarrow}(\mu_1^{L_1})\\0\\0\\0\\0\end{pmatrix} \tag{A2}$$

for Eq. (25) and in a similar way for Eqs. (26)–(28). The θ dependency can be simplified with the Gaussian quadrature method by discretizing the variable θ in $2N$ angles, as described by Turner and Weaver.[11] Following their notation, Eq. (24) is transformed to

$$\frac{\partial \mathbf{I}_{dm}(\tau)}{\partial \tau} + \mathbf{W}_m \mathbf{I}_{dm}(\tau)$$
$$= \mathbf{S}_{1m}^{L_1\downarrow} \exp\left[-\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_1}}\right] + \mathbf{S}_{1m}^{L_2\downarrow} \exp\left[-\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_2}}\right]$$
$$+ \mathbf{S}_{1m}^{y_1\downarrow} \exp\left[-\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_1}}\right] + \mathbf{S}_{1m}^{y_2\downarrow} \exp\left[-\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_2}}\right]$$
$$+ \mathbf{S}_{1m}^{L_1\uparrow} \exp\left[\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_1}}\right] + \mathbf{S}_{1m}^{L_2} \exp\left[\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_2}}\right]$$
$$+ \mathbf{S}_{1m}^{y_1\uparrow} \exp\left[\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_1}}\right] + \mathbf{S}_{1m}^{y_2\uparrow} \exp\left[\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_2}}\right], \tag{A3}$$

where

$$\mathbf{I}_{dm}(\tau) = \begin{bmatrix} \underline{I}_{dm}(\mu_{-N}, \tau) \\ \underline{I}_{dm}(\mu_{-N+1}, \tau) \\ \cdot \\ \cdot \\ \cdot \\ \underline{I}_{dm}(\mu_{N-1}, \tau) \\ \underline{I}_{dm}(\mu_N, \tau) \end{bmatrix}, \tag{A4}$$

$$\mathbf{S}_{1m}^{(L_{1,2}/y_{1,2})(\uparrow,\downarrow)} = \begin{bmatrix} \underline{S}_{1m}^{(L_{1,2}/y_{1,2})(\uparrow,\downarrow)}(\mu_{-N})/\mu_{-N} \\ \underline{S}_{1m}^{(L_{1,2}/y_{1,2})(\uparrow,\downarrow)}(\mu_{-N+1})/\mu_{-N+1} \\ \cdot \\ \cdot \\ \cdot \\ \underline{S}_{1m}^{(L_{1,2}/y_{1,2})(\uparrow,\downarrow)}(\mu_{N-1})/\mu_{N-1} \\ \underline{S}_{1m}^{(L_{1,2}/y_{1,2})(\uparrow,\downarrow)}(\mu_N)/\mu_N \end{bmatrix}, \tag{A5}$$

and $\mathbf{W}_m$ is defined by Turner and Weaver.[11] The solution to Eq. (A3) consists of a particular solution for each of the eight sources and a homogeneous solution. The particular solution is

$$\mathbf{I}_{pm}(\tau) = \mathbf{H}_{1m}^{L_1\downarrow} \exp\left[-\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_1}}\right] + \mathbf{H}_{1m}^{L_2\downarrow} \exp\left[-\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_2}}\right]$$
$$+ \mathbf{H}_{1m}^{y_1\downarrow} \exp\left[-\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_1}}\right] + \mathbf{H}_{1m}^{y_2\downarrow} \exp\left[-\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_2}}\right]$$
$$+ \mathbf{H}_{1m}^{L_1\uparrow} \exp\left[\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_1}}\right] + \mathbf{H}_{1m}^{L_2\uparrow} \exp\left[\frac{\tilde{\sigma}_L \tau}{\mu_1^{L_2}}\right]$$
$$+ \mathbf{H}_{1m}^{y_1\uparrow} \exp\left[\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_1}}\right] + \mathbf{H}_{1m}^{y_2\uparrow} \exp\left[\frac{\tilde{\sigma}_T \tau}{\mu_1^{y_2}}\right], \tag{A6}$$

where $\mathbf{D}$ is the $10N \times 10N$ identity matrix,

$$\mathbf{H}_{1m}^{L_1\downarrow} = \left(\mathbf{W}_m - \mathbf{D}\frac{\tilde{\sigma}_L}{\mu_1^{L_1}}\right)^{-1} \mathbf{S}_{1m}^{L_1\downarrow},$$

$$\mathbf{H}_{1m}^{L_2\downarrow} = \left(\mathbf{W}_m - \mathbf{D}\frac{\tilde{\sigma}_L}{\mu_1^{L_2}}\right)^{-1} \mathbf{S}_{1m}^{L_2\downarrow},$$

$$\mathbf{H}_{1m}^{y_1\downarrow} = \left(\mathbf{W}_m - \mathbf{D}\frac{\tilde{\sigma}_T}{\mu_1^{y_1}}\right)^{-1} \mathbf{S}_{1m}^{y_1\downarrow},$$

$$\mathbf{H}_{1m}^{y_2\downarrow} = \left(\mathbf{W}_m - \mathbf{D}\frac{\tilde{\sigma}_T}{\mu_1^{y_2}}\right)^{-1} \mathbf{S}_{1m}^{y_2\downarrow} \tag{A7}$$

correspond to the downward source terms and

$$\mathbf{H}_{1m}^{L_1\uparrow} = \left(\mathbf{W}_m + \mathbf{D}\frac{\tilde{\sigma}_L}{\mu_1^{L_1}}\right)^{-1} \mathbf{S}_{1m}^{L_1\uparrow},$$

$$\mathbf{H}_{1m}^{L_2\uparrow} = \left(\mathbf{W}_m + \mathbf{D}\frac{\tilde{\sigma}_L}{\mu_1^{L_2}}\right)^{-1} \mathbf{S}_{1m}^{L_2\uparrow},$$

$$\mathbf{H}_{1m}^{y_1\uparrow} = \left(\mathbf{W}_m + \mathbf{D}\frac{\tilde{\sigma}_T}{\mu_1^{y_1}}\right)^{-1} \mathbf{S}_{1m}^{y_1\uparrow},$$

$$\mathbf{H}_{1m}^{y_2\uparrow} = \left(\mathbf{W}_m + \mathbf{D}\frac{\tilde{\sigma}_T}{\mu_1^{y_2}}\right)^{-1} \mathbf{S}_{1m}^{y_2\uparrow} \tag{A8}$$

correspond to the upward terms.

The homogeneous solution can be found by solving an eigenvalue problem with eigenvectors g_{mn} and eigenvalues λ_{mn}. The full solution for the mth Fourier expansion term of the diffuse intensity is written as

$$\mathbf{I}_{dm}(\tau) = \mathbf{I}_{pm}(\tau) + \sum_{n=1}^{10N} C_{mn} g_{mn} e^{-\lambda_{mn}\tau}, \tag{A9}$$

where the constants C_{mn} must be found from the boundary conditions.

The intensity vector $\mathbf{I}_{dm}$ can be divided into upward $(\mathbf{I}_{dm}^+, \mu < 0)$ and downward $(\mathbf{I}_{dm}^-, \mu > 0)$ intensities, and it is evaluated at $\tau = 0^+$ and $\tau = \tau_b^-$, so the top boundary condition $\mathbf{I}_{dm}^-(\tau = 0^+) = \hat{\mathbf{R}}_{10}\mathbf{I}_{dm}^+(\tau = 0^+)$ yields the equation

$$\sum_{n=1}^{10N} C_{mn} g_{mn}^- + \mathbf{I}_{pm}^-(0^+) = \hat{\mathbf{R}}_{10}\left(\sum_{n=1}^{10N} C_{mn} g_{mn}^+ + \mathbf{I}_{pm}^+(0^+)\right), \tag{A10}$$

and the bottom boundary condition $\mathbf{I}_m^+(\tau_b^-) = \hat{\mathbf{R}}_{12}\mathbf{I}_m^-(\tau_b^-)$ yields

$$\sum_{n=1}^{10N} C_{mn} g_{mn}^+ e^{-\lambda_{mn}\tau_b} + \mathbf{I}_{pm}^+(\tau_b)$$
$$= \hat{\mathbf{R}}_{12}\left(\sum_{n=1}^{10N} C_{mn} g_{mn}^- e^{-\lambda_{mn}\tau_b} + \mathbf{I}_{pm}^-(\tau_b)\right), \tag{A11}$$

where

$$\hat{\mathbf{R}}_{10} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & \underline{\hat{R}}_{10}(\mu_{-1}) \\ 0 & 0 & 0 & \cdots & 0 & \underline{\hat{R}}_{10}(\mu_{-2}) & 0 \\ & & & & \cdot & & \\ & & & & \cdot & & \\ & & & & \cdot & & \\ 0 & \underline{\hat{R}}_{10}(\mu_{-N+1}) & 0 & 0 & 0 & 0 & 0 \\ \underline{\hat{R}}_{10}(\mu_{-N}) & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{A12}$$

$$\hat{\mathbf{R}}_{12} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & \hat{\underline{R}}_{12}(\mu_N) \\ 0 & 0 & 0 & \cdots & 0 & \hat{\underline{R}}_{12}(\mu_{N-1}) & 0 \\ & & & & \cdot & & \\ & & & \cdot & & & \\ & & & \cdot & & & \\ 0 & \hat{\underline{R}}_{12}(\mu_2) & 0 & 0 & 0 & 0 & 0 \\ \hat{\underline{R}}_{12}(\mu_1) & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{A13}$$

with the matrix of reflection coefficients $\hat{\underline{R}}_{ab}$ defined as

$$\hat{\underline{R}}_{ab} = \begin{pmatrix} \hat{R}_{ab}^{LL} & \hat{R}_{ab}^{yL} & 0 & 0 & 0 \\ \hat{R}_{ab}^{Ly} & \hat{R}_{ab}^{yy} & 0 & 0 & 0 \\ 0 & 0 & \hat{R}_{ab}^{xx} & 0 & 0 \\ 0 & 0 & 0 & \hat{R}_{ab}^{UU} & \hat{R}_{ab}^{VU} \\ 0 & 0 & 0 & \hat{R}_{ab}^{UV} & \hat{R}_{ab}^{VV} \end{pmatrix}. \tag{A14}$$

Equations (A10) and (A11) can be written in matrix form as

$$G^{-}(0)C_m + \mathbf{I}_{pm}^{-}(0) = \widehat{\mathbf{R}_{10}}(G^{+}(0)C_m + \mathbf{I}_{pm}^{+}(0)),$$

$$G^{+}(\tau_b)C_m + \mathbf{I}_{pm}^{+}(\tau_b) = \widehat{\mathbf{R}_{12}}(G^{-}(\tau_b)C_m + \mathbf{I}_{pm}^{-}(\tau_b)), \tag{A15}$$

where

$$G^{-}(\tau) = \begin{pmatrix} g_{m1}^{1}e^{-\lambda_1\tau} & g_{m2}^{1}e^{-\lambda_2\tau} & \cdots & g_{m_{10}N}^{1}e^{-\lambda_{10N}\tau} \\ g_{m1}^{2}e^{-\lambda_1\tau} & g_{m2}^{2}e^{-\lambda_2\tau} & \cdots & g_{m10N}^{2}e^{-\lambda_{10N}\tau} \\ \cdot & \cdot & \cdot & \cdot \\ g_{m1}^{N}e^{-\lambda_1\tau} & g_{m2}^{N}e^{-\lambda_2\tau} & \cdots & g_{mN}^{N}e^{-\lambda_{10N}\tau} \end{pmatrix}, \tag{A16}$$

$$G^{+}(\tau) = \begin{pmatrix} g_{m1}^{-N}e^{-\lambda_1\tau} & g_{m2}^{-N}e^{-\lambda_2\tau} & \cdots & g_{m_{10}N}^{-N}e^{-\lambda_{10N}\tau} \\ g_{m1}^{-N+1}e^{-\lambda_1\tau} & g_{m2}^{-N+1}e^{-\lambda_2\tau} & \cdots & g_{m10N}^{-N+1}e^{-\lambda_{10N}\tau} \\ \cdot & \cdot & \cdot & \cdot \\ g_{m1}^{-1}e^{-\lambda_1\tau} & g_{m2}^{-1}e^{-\lambda_2\tau} & \cdots & g_{mN}^{-1}e^{-\lambda_{10N}\tau} \end{pmatrix}, \tag{A17}$$

$$C_m = \begin{pmatrix} C_{m1} \\ C_{m2} \\ \cdots \\ C_{m10N} \end{pmatrix}. \tag{A18}$$

The equations in Eq. (A15) are grouped to form a single matrix equation

$$G_1 C_m + P = R_G C_m + R_P, \tag{A19}$$

where

$$G_1 = \begin{pmatrix} G^{-}(0) \\ G^{+}(\tau_b) \end{pmatrix}, \tag{A20}$$

$$P = \begin{pmatrix} \mathbf{I}_{pm}^{-}(0) \\ \mathbf{I}_{pm}^{+}(\tau_b) \end{pmatrix}, \tag{A21}$$

$$R_G = \begin{pmatrix} \hat{\mathbf{R}}_{10}G^{+}(0) \\ \hat{\mathbf{R}}_{12}G^{-}(\tau_b) \end{pmatrix}, \tag{A22}$$

$$R_P = \begin{pmatrix} \hat{\mathbf{R}}_{10}\mathbf{I}_{pm}^{+}(0) \\ \hat{\mathbf{R}}_{12}\mathbf{I}_{pm}^{-}(\tau_b) \end{pmatrix}. \tag{A23}$$

Solving for C_m,

$$C_m = (G_1 - R_G)^{-1}(R_P - P). \tag{A24}$$

[1] D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics*, 1st ed. (Springer, New York, 2007).
[2] D. R. Jackson, D. P. Winebrenner, and A. Ishimaru, "Application of the composite roughness model to high-frequency bottom scattering," J. Acoust. Soc. Am. **79**, 1410–1422 (1986).
[3] A. N. Ivakin, "A unified approach to volume and roughness scattering," J. Acoust. Soc. Am. **103**, 827–837 (1998).
[4] D. Tang, "Acoustic wave scattering from a random ocean bottom," Ph.D. thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Massachusetts (1991).
[5] P. D. Mourad and D. R. Jackson, "A model/data comparison for low-frequency bottom backscatter," J. Acoust. Soc. Am. **94**, 344–358 (1993).
[6] A. P. Lyons, "The potential impact of shell fragment distributions on high-frequency seafloor backscatter," J. Oceanic Eng. **30**, 843–851 (2005).
[7] D. Chu, K. L. Williams, D. Tang, and D. R. Jackson, "High-frequency bistatic scattering by sub-bottom gas bubbles," J. Acoust. Soc. Am. **102**, 806–814 (1997).
[8] A. N. Ivakin and D. R. Jackson, "Effects of shear elasticity on sea bed scattering: Numerical examples," J. Acoust. Soc. Am. **103**, 346–354 (1998).
[9] A. A. Kokhanovsky, *Optics of Light Scattering Media*, 2nd ed. (Springer–Praxis, U.K., 2001).
[10] L. Margerin, M. Campillo, and B. van Tiggelen, "Radiative transfer and diffusion of waves in a layered medium: New insight into coda Q," Geophys. J. Int. **134**, 596–612 (1998).
[11] J. A. Turner and R. L. Weaver, "Radiative transfer of ultrasound," J. Acoust. Soc. Am. **96**, 3654–3672 (1994).
[12] J. E. Quijano and L. M. Zurk, "Application of radiative transfer theory to acoustic propagation in the ocean bottom," in Proceedings of the Oceans '07, Vancouver, BC, Canada (2007), pp. 1–7.
[13] A. Ishimaru, *Wave Propagation and Scattering in Random Media*, 1st ed. (Academic, New York, 1978), Vol. **1**.
[14] M. I. Mishchenko and L. D. Travis, *Multiple Scattering of Light by Particles*, 1st ed. (Cambridge University Press, Cambridge, 2006).
[15] L. Tsang and A. Ishimaru, "Radiative wave and cyclical transfer equations for dense non tenuous media," J. Opt. Soc. Am. **2**, 2187–2194 (1985).
[16] B. Wen, L. Tsang, D. P. Winebrenner, and A. Ishimaru, "Dense medium radiative transfer theory: Comparison with experiment and application to microwave remote sensing and polarimetry," IEEE Trans. Geosci. Remote Sens. **28**, 46–59 (1990).
[17] L. M. Zurk, L. Tsang, and D. P. Winebrenner, "Scattering properties of dense media from Monte Carlo simulations with application to active remote sensing of snow," Radio Sci. **31**, 803–819 (1996).
[18] G. Bal, "Radiative transfer equations with varying refractive index: A mathematical perspective," J. Opt. Soc. Am. **23**, 1639–1644 (2006).
[19] K. N. Liou, *Introduction to Atmospheric Radiation*, 2nd ed. (Academic, San Diego, CA, 2002).
[20] G. A. Titov, "Radiative horizontal transport and absorption in stratocumulus clouds," J. Atmos. Sci. **55**, 2549–2560 (1998).
[21] T. Okutucu, Y. Yener, and A. A. Busnaina, "Transient radiative transfer in participating media with pulse-laser irradiation: An approximate Galerkin solution," J. Quant. Spectrosc. Radiat. Transf. **103**, 118–130 (2007).
[22] J. A. Turner and R. L. Weaver, "Time dependence of multiply scattered diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **97**, 2639–2644 (1995).
[23] S. Chandrasekhar, *Radiative Transfer*, 1st ed. (Dover, New York, 1960).
[24] R. T. Shin and J. A. Kong, "Radiative transfer theory for active remote sensing of a homogeneous layer containing spherical scatterers," J. Appl. Phys. **52**, 4221–4230 (1981).
[25] L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*, 1st ed. (Wiley-Interscience, New York, 1985).
[26] L. M. Brekhovskikh, *Waves in Layered Media*, 2nd ed. (Academic, Orlando, FL, 1980).

[27]J. A. Turner and R. L. Weaver, "Ultrasonic radiative transfer in polycrystalline media: Effects of a fluid-solid interface," J. Acoust. Soc. Am. **98**, 2801–2808 (1995).
[28]D. R. Jackson, K. B. Briggs, K. L. Williams, and M. D. Richardson, "Tests of models for high-frequency seafloor," J. Oceanic Eng. **21**, 458–470 (1996).
[29]M. J. Buckingham, "Wave propagation, stress relaxation, and grain-to-grain shearing in saturated, unconsolidated marine sediments," J. Acoust. Soc. Am. **108**, 2796–2815 (2000).
[30]E. L. Hamilton, "Geoacoustic modeling of the sea floor," J. Acoust. Soc. Am. **68**, 1313–1339 (1980).
[31]E. L. Hamilton, "V_p/V_s and Poisson's ratios in marine sediments and rocks," J. Acoust. Soc. Am. **66**, 1093–1101 (1979).
[32]C. Wu, "Propagation of scattered radiation in a participating planar medium with pulse irradiation," J. Quant. Spectrosc. Radiat. Transf. **64**, 537–548 (2000).

Riverbed sediment classification using multi-beam echo-sounder backscatter data

AliReza Amiri-Simkooei,[a] Mirjam Snellen, and Dick G. Simons
Acoustic Remote Sensing Group, Delft Institute of Earth Observation and Space Systems (DEOS), Faculty of Aerospace Engineering, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands

(Received 31 March 2009; revised 21 July 2009; accepted 24 July 2009)

A method has recently been developed that employs multi-beam echo-sounder backscatter data to both obtain the number of sediment classes and discriminate between them by applying the Bayes decision rule to multiple hypotheses [Simons and Snellen, Appl. Acoust. **70**, 1258–1268 (2009)]. In deep water, the number of scatter pixels within the beam footprint is large enough to ensure Gaussian distributions for the backscatter strengths and to increase the discriminative power between acoustic classes. In very shallow water (<10 m), however, this number is too small. This paper presents an extension of this high-frequency methodology for these environments, together with a demonstration of its performance using backscatter data from the river Waal, The Netherlands. The objective of this work is threefold. (i) Increasing the discriminating power of the classification method: high-resolution bathymetry data allow precise bottom slope corrections for obtaining the true incident angle, and the high-resolution backscatter data reduce the statistical fluctuations via an averaging procedure. (ii) Performing a correlation analysis: the dependence of acoustic backscatter classification on sediment physical properties is verified by observing a significant correlation of 0.75 (and a disattenuated correlation of 0.90) between the classification results and sediment mean grain size. (iii) Enhancing the statistical description of the backscatter intensities: angular evolution of the *K*-distribution shape parameter indicates that the riverbed is a rough surface, in agreement with the results of the core analysis.
 [DOI: 10.1121/1.3205397]

PACS number(s): 43.30.Hw, 43.30.Xm, 43.30.Ma, 43.30.Vh [RCG] Pages: 1724–1738

I. INTRODUCTION

It is widely accepted that multi-beam echo-sounder (MBES) data can be used to measure the bathymetry of rivers, seas, and oceans. In addition, backscatter data acquired from MBES systems are employed to obtain information about the physical properties of the riverbed and seafloor. The main advantage of the method is its high spatial coverage capability at limited costs. Proper analysis and subsequent interpretation of the backscatter data are still challenging problems. The ultimate goal of acoustic classification methods is to remotely measure physical properties of the surficial sediments such as porosity, permeability, and grain-size distribution.

Methods exist that base the classification on the backscatter data from which the angular dependence has been taken out. However, complications in eliminating the angular dependence can arise, e.g., due to local bottom slopes and the unknown MBES directivity pattern. In addition, there is an intrinsic variation in the backscatter intensity with incident angle. To eliminate this angular dependence, one can, for instance, apply Lambert's law,[1] which states that the intensity of acoustic backscatter is proportional to the square of the cosine of the incident angle ($\pi/2-$grazing angle). Lambert's law does not always correctly represent the angular dependence. In fact, the statistical distribution of backscatter data changes with incident angle, and therefore removing only the mean values does not completely compensate the angular effects.

In an earlier work, a method was proposed for the classification of the seabed sediment that accounts for backscatter variability.[2] The method, based on the Bayesian decision rule, was applied to MBES backscatter data for the classification in a test area in the North Sea with well-known lithology. This method employs the backscatter strength collected at a certain incident angle instead of studying the angular behavior of the backscatter strength. The backscatter data employed are the averaged backscatter strengths per beam, i.e., obtained from averaging over backscatter strength for a large number of signal footprints or scatter pixels. The classification is performed per angle, separately from other angles, and hence it is considered to be an angle-independent method. This method needs to be adopted when applied to MBES data taken in very shallow waters.

There are two issues involved when applying a classification method such as that described in Ref. 2 to very shallow water (e.g., riverbed) areas. The shallower water depths correspond to smaller beam footprints, resulting in a smaller number of scatter pixels per beam footprint. Because the standard deviation of the averaged backscatter data is inversely proportional to the square root of the number of scatter pixels, the averaged backscatter data are subject to higher

[a]Author to whom correspondence should be addressed. Electronic mail: a.amirisimkooei@tudelft.nl. Also at: Department of Surveying Engineering, The University of Isfahan, 81744, Isfahan, Iran.

ping-to-ping variability in shallow water areas. The discriminating power between sediments will accordingly decrease. In addition, significant bottom slopes (e.g., up to 30°) exist in river environments. These will affect the incident angle and the backscatter data, and therefore the classification results. In this contribution we elaborate these two issues in detail and improve the above-mentioned classification method for a river environment. The improved classification method will be applied to a very shallow water river environment where significant slopes occur.

This paper is built up as follows. Section II briefly reviews previous work. In Sec. III, the MBES classification method proposed in Ref. 2 is briefly described. The data considered in Ref. 2 were acquired in an area where no significant slopes exist. But, for the work described in this paper the slope effects need to be accounted for. Simons and Snellen[2] used a single beam of all MBES beams available, thereby employing only a small part of the available data. Here we describe the extension of the method where almost all MBES backscatter data are accounted for in the classification. In Sec. IV we discuss our methodology to estimate the bottom slopes using the precise bathymetry data and to apply corrections to the backscatter data. Section V presents the classification results applied to the MBES data of the river Waal in The Netherlands. Extensive sediment grabbing (analyzed for grain-size distribution) is also available, which allows one to assess the performance of the classification method. The acoustic classification results are correlated with the mean grain sizes of the core data. Pearson and disattenuated correlation coefficients will be compared. We further study the problem with another class of distributions, namely, non-Rayleigh distributions. The possible application of the *K*-distribution for the classification of the data is assessed. We conclude the paper in Sec. VI.

II. BACKGROUND AND PREVIOUS WORK

Many studies of remote sediment classification use sidescan sonar systems and MBES systems[3–7] and directly compare backscatter data to the physical properties of sediments. Recent studies showing that the high-frequency (e.g., >40 kHz as we use for our classification method) acoustic backscatter data depend on the sediment physical properties in general and the grain-size distribution, in particular, can be found in Refs. 8 and 9. Our impression is that the low (insignificant) correlation coefficient (in earlier studies) between the backscatter data and the mean grain size of sediments is due to the high variability of the backscatter data, which attenuates the correlation coefficients.

Some research is going on in the field of self-organizing maps (SOMs) and artificial neural networks (ANNs) applied to the problem of seafloor classification using MBES backscatter data.[10,11] More recent studies perform seabed classification using the ANN method that preserves the backscatter angular information and incorporates both backscatter and bathymetric data.[12] The SOM is a type of ANN algorithm based on unsupervised learning. It provides a tool for visualizing the multidimensional numerical data to produce a low-dimensional map. Also similar studies are ongoing in the field of angular range analysis (ARA) using multi-beam sonar systems.[13,14] The method, based on the normalized acoustic backscatter mosaic, aims to estimate the acoustic impedance and roughness of the insonified area on the seafloor. The methods described above adjust themselves to the data without any explicit specification of the distributional form for the underlying model.[15] It is therefore difficult to statistically interpret the classification results using such methods.

The backscatter data change with the angle of incidence. The angular dependence of the backscatter data can potentially be used as a tool for classification.[16,17] A problem in this approach arises for areas where the seafloor type varies along the swath. It is therefore difficult to discriminate between the angular variation and the real seafloor type variation along the swath. In addition, the approach requires a good calibration of the MBES system, i.e., its sensitivity should be equal for all steering angles. The classification method described in Ref. 2 takes the data at a single angle and it is considered to be less sensitive to the seafloor type variation along the swath and the calibration of the MBES.

It is widely known that the backscatter data are subject to statistical fluctuations.[18–20] The classical Rayleigh distribution is not applicable to backscatter data when the deterministic number of scatterers within the signal footprint (also called size of scatter pixel or ensonified area) is not large enough and hence the central limit theorem does not hold. Non-Rayleigh distributions such as *K*, Weibull, Rayleigh mixture, or log-normal distributions occur when the conditions of the central limit theorem are violated.[20–24] Among them the *K*-distribution provides a good fit to the skewed distributions of experimental data for all sediment types.[25–28] It also offers physical insights into the backscatter data.[29,30] Previous studies have shown that the statistical characteristics of backscatter data strongly depend on the incident angle. More recent studies use angular evolution of the *K*-distribution shape parameter as a tool for seafloor characterization.[31,32]

III. ACOUSTIC CLASSIFICATION METHOD

A. Fluctuation of backscatter data

The MBES systems that typically operate at a few hundred kHz permit seafloor backscatter imaging with high resolution. The echo amplitudes (here backscatter strength) measured by the MBES can be employed for seafloor and riverbed classification. Since the signal footprint A_f (or scatter pixel) is small compared with the beam footprint for beams away from nadir, many scatter pixels are expected to fall within the footprint of the receiving beam.

It is traditionally assumed that the backscatter intensity of the *i*th scatter pixel in a beam, denoted by the random variable I, is exponentially distributed,[19] i.e., distributed as chi-square with two degrees of freedom. This is based on the validity of the central limit theorem where the number of scatterers N_s inside the signal footprint is large enough. The normalized amplitude $\sqrt{I}$ has a Rayleigh distribution. The corresponding backscatter strength in decibels obtained by applying $10\log_{10}$ to the intensities I has a Gumbel

distribution,[33] which is a special case of the log-Weibull (or Fisher–Tippett) distribution.[34] The theory of Rayleigh distributions is not applicable to backscatter data when the number of scatterers within the signal footprint is not large enough and hence the central limit theorem does not hold (see Sec. V D).

The data employed for the classification method consist of backscatter values (in decibels) per receiver beam, i.e., backscatter values obtained from averaging over N independent scatter pixels. Such values—given for each beam angle—are corrected for propagation loss, the area A_f of the signal footprint, and the local bottom slopes. Ping-to-ping variability masks the influence of the seafloor type on the backscatter strength. The averaged backscatter value is still subject to statistical fluctuations. For large N's, the averaged backscatter is normally distributed (central limit theorem) for one sediment class. The classification method described in Ref. 2 fully employs this knowledge about the backscatter strength probability density function (PDF). It thus assumes that both N_s and N are sufficiently large.

In very shallow water (depth typically 5 m) the number of scatter pixels N is not large enough to use the central limit theorem to ensure the Gaussianity of the averaged backscatter strength. To ensure a Gaussian distribution of the backscatter strength, we here propose to use backscatter strengths averaged over small surface patches (Sec. IV). For further characterization of the riverbed sediment we may alternatively use the original backscatter intensities and apply the K-distribution explained in Sec. V D. Both will be addressed in this contribution.

B. Classification methodology

Let the backscatter intensity of the ith scatter pixel be distributed as $f_I(I)$, which is an arbitrary distribution (e.g., classically an exponential distribution). When the number N of the independent and identically distributed scatter pixels (per beam footprint) is large enough, the central limit theorem states that the averaged backscatter strength BS has a Gaussian distribution. The classification approach, described in great detail in Ref. 2 employs the averaged backscatter data at a single beam angle. Without going into detail, the method is summarized here and comprises the following steps.

Step 1 (nonlinear curve fitting): The algorithm starts by fitting r number of Gaussian PDFs, i.e., $\mathrm{BS}\sim f_{\mathrm{BS}}(\mathrm{BS})=\sum_{i=1}^{r} c_i N(\mathrm{BS},\mu_i,\sigma_i^2)$, to the histogram of measured backscatter strengths BS for a selected single angle. r is the number of sediment types. Each Gaussian PDF, with unknown mean μ_i and variance σ_i^2, represents one sediment type. The coefficients c_i of the linear combination of the PDFs are not known. This leads to the total number of unknown parameters as $3r$ (i.e., c_i the contribution of individual PDFs, μ_i the mean of PDFs, and σ_i the standard deviation of PDFs, $i=1,\ldots,r$).

Let the equally binned (e.g., with bin size of 0.5 dB) backscatter strength at $\mathrm{BS}_1,\ldots,\mathrm{BS}_m$ be denoted $y=[n_1,\ldots,n_m]^T$, where n_i, $i=1,\ldots,m$ is the frequency of the samples in the bin. The model of observation equation can then be written as

$$E(y)=A(x)=A(c,\mu,\sigma),\quad D(y)=Q_y, \tag{1}$$

where $A(c,\mu,\sigma)$ is expressed as an unknown linear combination of r Gaussian PDFs and the covariance matrix $Q_y=\mathrm{diag}(n_1,\ldots,n_m)$ is based on the Poisson-distributed random variables y_i's with variances n_i's. E and D are the *expectation* and *dispersion* operators, respectively.

The nonlinear least-squares problem is formulated as

$$\hat{c},\hat{\mu},\hat{\sigma}=\arg\min_{c,\mu,\sigma}\|y-A(c,\mu,\sigma)\|^2_{Q_y^{-1}}, \tag{2}$$

subject to the non-equality constraints $c\geq 0$, $\mu^l\leq\mu\leq\mu^u$, and $\sigma^l\leq\sigma\leq\sigma^u$, where $\|\cdot\|^2_{Q_y^{-1}}=(\cdot)^TQ_y^{-1}(\cdot)$ and the subscripts l and u denote the lower and upper bounds of the variables, respectively. The nonlinear least-squares subject to bounds on variables[35] is used to obtain the μ_i's and σ_i's, and the non-negative least-squares[36] is used to obtain the contributions of the individual PDFs by constraining the coefficients c_i's to be positive. (In the mathematical computer package, MATLAB one may, respectively, use lsqnonlin.m and lsqnonneg.m.)

Because in practice the number r of the sediment types is not known, it has also to be determined. The number r of Gaussian PDFs can be determined by using a goodness of fit criterion based on a chi-square distributed test statistic,

$$\chi^2=\|\hat{e}\|^2_{Q_y^{-1}}=\|y-A(\hat{c},\hat{\mu},\hat{\sigma})\|^2_{Q_y^{-1}}\sim\chi^2(m-3r), \tag{3}$$

where $\hat{e}$ is the least-squares residual vector, and $m-3r$ is the degrees of freedom. The curve fitting procedure is executed in an iterative manner for different values of r (starting from $r=1$) such that no further decrease in the test statistic is observed by increasing the number of Gaussian PDFs. For such a case and also the case when the test statistic falls below a critical value [i.e., $\chi^2_{\alpha,(m-3r)}$, with α the significance level of the test] the procedure will be stopped.

Step 2 (identification of acoustic classes): For the classification, when we know the PDF for each seafloor type i, we can apply the Bayes decision rule. We have r hypotheses H_i, $i=1,\ldots,r$, and therefore there exist r possible decisions. We choose the hypothesis that, given the backscatter observations, maximizes the likelihood function. The intersections of the r Gaussian PDFs result in r non-overlapping acceptance regions. Each interval in backscatter strength now corresponds to one acoustic class (Fig. 1).

Step 3 (assigning seafloor types): The goal is to correspond the Gaussian distributions to the grain-size distribution of the sediments. We need to assign a seafloor type to each of the r acceptance regions (acoustic classes) obtained in Step 2. There might exist different ways to approach this goal. One can, for instance, rely on the previous work,[3,4,8,9,37] where the estimated $\hat{\mu}_i$'s can directly be associated with the seafloor mean grain size. An alternative, followed in this contribution, is to use the results from the core analysis for comparison and to perform a correlation analysis afterward. Three grab samples per kilometer (a total of 30 samples for

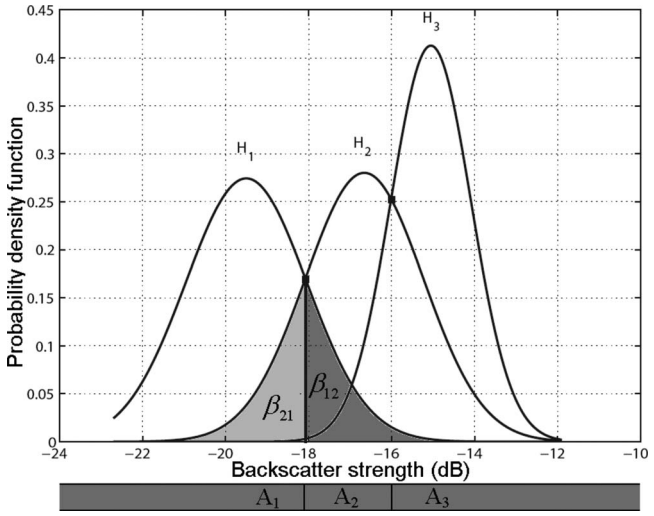


FIG. 1. Three Gaussian PDFs (H_1, H_2, and H_3) represent three sediment classes. Intersection of consecutive PDFs gives non-overlapping acceptance regions A_1, A_2, and A_3. Also indicated are examples of probability of incorrect decision β_{12} and β_{21}.

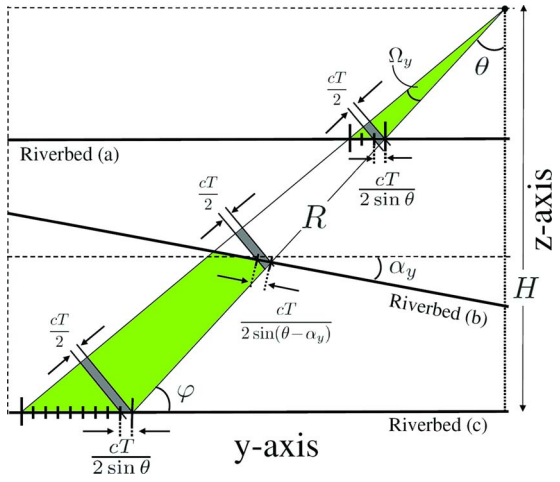


FIG. 2. (Color online) Across-track cross section (y-z plane) for signal footprint of an oblique beam for three configurations: shallow water (a), non-flat bottom (b), and deep water (c).

10 km) have been taken for such comparison (see Sec. V C).

Step 4 (quality assessment): The quality of the classification algorithm can be assessed by calculating the decision matrix of the multiple-hypothesis testing problem. This matrix contains the probabilities of correct and incorrect decisions. The decision matrix provides us with a measure for the quality of the classification method. The probability of incorrect decision is proportional to the overlap area of the Gaussian PDFs (Fig. 1). If the probability of incorrect decision decreases, the overlap area will decrease and consequently the power of the discrimination (classification) will increase.

Step 5 (presentation and mapping): This final step of the algorithm comprises the actual mapping, i.e., allocation of seafloor type (e.g., a color) to all measured backscatter strengths. As the MBES system provides a position to each backscatter strength measurement, we can map seafloor type versus position. For better presentation of the results, an interactive three-dimensional data visualization system such as FLEDERMAUS software can be used.[38] It allows, for instance, to further smooth the classification results by using a weighted moving average method.

For the analysis described above, the backscatter strengths per beam are assumed to have a Gaussian PDF. For shallow water applications N might not be sufficiently large. In addition, bottom slopes can be significant in the river environment considered in this paper. Therefore, two intermediate steps are added to the approach in Ref. 2. These steps are as follows.

Step I (correcting and averaging procedure): This step is performed before Step 1 to prepare the backscatter data (average over small patches and correct for local slopes) for the classification method described above. In shallow water environments such as rivers, the number N of scatter pixels inside the beam footprint is not large because N is proportional to the water depth [Fig. 2(a)]. The current application of the classification method, to result in the normality restoration by means of the central limit theorem, is based on the average backscatter values over the small surface patches. Each patch consists of a few beams in the across-track direction and a few pings in the along-track direction. It also allows one to apply the slope corrections to the backscatter data, namely, correction due to the changes in the area of the signal footprint and correction due to the true beam grazing angle. The details of this step are explained in Sec. IV. Therefore, for angle θ the "averaged corrected" (over patches) backscatter data will be used.

Step II (combination of different angles): This step is performed and iterated after Steps 1 and 2 to combine the results from different angles. The methodology of Simons and Snellen[2] takes observations from one single angle only. In practice, to use the full high-resolution mapping potential of the method, we may consider multiple beams and individually perform the classification. This consequently allows one to obtain a continuous map over the whole area. The classification method at angles close to nadir (e.g., θ=20°), however, becomes less efficient as the backscatter values of different sediment types have values close to each other. One remedy, followed in this contribution, is to first use the backscatter data at a few low grazing angles (e.g., reference angles of θ=64°, 62°, 60°) and apply the classification method. This analysis gives the number r of the sediment types, the means μ_i, the variances σ_i^2, and the coefficients c_i. The nonlinear curve fitting in Step 1 is based on the bounds on the variables. Based on this information, the curve fitting procedure is then executed and extended to all other angles ranging from θ=60°, θ=58°,…, θ=20°.

- For a fixed number r of the Gaussian PDFs, where r has been determined from the application of the classification method to the backscatter data of the low grazing reference angles (say, θ=64, 62, 60°).
- By obtaining a good initial guess for the mean parameters, i.e., $\mu_i^0(i=1,\ldots,r)$, of the backscatter data at the angle under study. This is achieved by using the means $\mu_i(i=1,\ldots,r)$ of the reference angles and equally shifted by the differences between the mean backscatter values at the angle under study (of entire histogram) and the mean backscatter values at the reference angles.

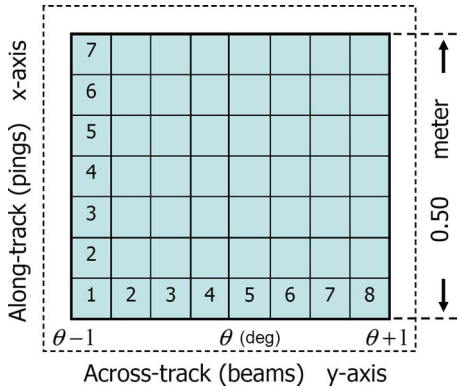


FIG. 3. (Color online) Schematic surface patch at "nominal" incident angle of θ, which consists of $m=8\times7=56$ beams along- and cross-track directions.

- By using more strict bounds (reduce bounds) on the mean parameters $\mu_i(i=1,\ldots,r)$ for the classification of backscatter data at the angle under study (e.g., $\mu_i^l=\mu_i^0-0.5$ dB and $\mu_i^u=\mu_i^0+0.5$ dB). The bounds considered are still wide enough to compensate for the angular dependence of the statistical distributions for the backscatter data.

IV. LOCAL SLOPE CORRECTION

The significant local slopes of the riverbed will affect the classification results. To compensate for these effects, one has to estimate the along- and across-track slopes. MBESs provide detailed bathymetry information from which the local slopes can be estimated. This allows one to improve the seabed classification results by applying the corrections to the backscatter data. The literature has paid little attention to the question of how such corrections should be estimated and taken into account. We may refer, for example, to Ref. 13. We develop a methodology that compensates for the effect of bottom slopes, both in along- and across-track directions. Such effects are of high importance especially for river environments as considered in this study. Two effects are discussed: (1) correction due to the changes in the ensonified area (signal footprint) to which the backscatter data refers and (2) correction due to the true beam grazing angle. Both corrections can be applied when the along- and across-track slopes of the seafloor (riverbed) are available. The least-squares method is employed to estimate the local slopes using the precise bathymetry data.

A. Estimation of slopes

A discrete surface patch $z_i=f(x_i,y_i)$, $i=1,\ldots,m$ includes a few angles around the central beam angle (e.g., with deviation of 1°), where the angular dependence of the statistical distribution of the backscatter data is negligible. Also, because the ping rate is high (40 Hz), we may in addition include a set of neighboring pings to make a surface patch and hence to be able to estimate the along- and across-track slopes. This results in a window (e.g., 0.5×0.5 m^2) that contains, say, $m=56$ beams (Fig. 3). The average backscatter data and the average depth in this small patch will be used. Using this strategy to divide the area under survey into small surface patches and to use the average backscatter values, (a) one can compute the along- and across-track slopes and correct for the true grazing angle and the backscatter data; (b) One can assure that the normality assumption is achieved by means of the central limit theorem. This is a prerequisite for using the classification method (see Sec. III). (c) One can decrease the variance and hence increase the discriminating power between sediments. This makes the classification method more discriminative.

A bi-quadratic function consisting of six unknown coefficients is used to model (estimate) the surface patch. This subsequently allows one to obtain the along-track (x) and across-track (y) slopes

$$z=f(x,y)=a_0+a_1x+a_2y+a_3x^2+a_4y^2+a_5xy. \tag{4}$$

The least-squares method can be used to estimate the parameters of the polynomial, and a procedure called "data snooping"[39] can be used to test for the presence of outliers in the bathymetry data (see below). When the local surface is small enough, further simplification of $f(x,y)$ is possible by using a plane instead of a bi-quadratic polynomial ($a_3=a_4=a_5=0$). The subsequent formulations can thus be simplified accordingly.

For the linear model of observation equations $E(z)=Aa$, the least-squares estimate of the vector of unknown coefficients $a=[a_0,\ldots,a_5]^T$ is

$$\hat{a}=(A^TQ_z^{-1}A)^{-1}A^TQ_z^{-1}z, \tag{5}$$

where A is the known $m\times6$ design matrix (its ith row is $A_i=[1\ x_i\ y_i\ x_i^2\ y_i^2\ x_iy_i]$), $z=[z_1,\ldots,z_m]^T$ is an m-vector of depth measurements, and $Q_z=\sigma^2I$ is the covariance matrix of z, with σ^2 the variance of the data and I an identity matrix. Note that for this special structure of $Q_z=\sigma^2I$ (independent and identically distributed errors), Eq. (5) simplifies to $\hat{a}=(A^TA)^{-1}A^Tz$. It indicates that the unknown coefficients a can be estimated independent of the (un)known variance of the data. The least-squares estimate of the variance component is $\hat{\sigma}^2=\hat{e}^T\hat{e}/(m-6)$, where $\hat{e}=A\hat{a}-z$ is the m-vector of the least-squares residual.[40] The covariance matrix of the unknown coefficients a is given as: $Q_{\hat{a}}=\hat{\sigma}^2(A^TA)^{-1}$. Also the covariance matrix of the residuals is $Q_{\hat{e}}=\hat{\sigma}^2(I-A(A^TA)^{-1}A)$.

The data snooping procedure[39] for the detection, identification, and adaptation of possible outliers and anomalies in the measurements can be applied in an iterative manner to screen the observations from the presence of such errors. The normalized entries of the residual vector $\hat{e}$, i.e., $\hat{w}_i=\hat{e}_i/\sigma_{\hat{e}_i}$ (i runs from 1 to m), is a test statistic used for data snooping. In this statistic, $\sigma_{\hat{e}_i}=(Q_{\hat{e}})_{ii}^{1/2}$ is the standard deviation of the least-squares residuals obtained as the square root of the ith diagonal entry of $Q_{\hat{e}}$. The test statistic w_i has a standard normal distribution when σ^2 is known. It has a Student t distribution when σ^2 is unknown. When the test for observation i is rejected, one may conclude that observation i is affected by some extraordinary large errors. By letting i run from 1 to m, one can screen the data on the presence of potential outliers in the individual observations. The test statistic $w_{\max}$ (one value out of m values: $w_{\max}=\max(w_i)$, $i=1,\ldots,m$) that has the largest (in absolute sense) value refers to the observation which is most likely corrupted with a

outlier. The corresponding observation is excluded from the list of observations and the same procedure is applied for identifying yet other potential outliers.

The estimated bathymetry data, which express the estimated local surface, are $\hat{z}=\hat{f}(x,y)=A\hat{a}$. At point (x,y) the partial derivatives of $\hat{z}=\hat{f}(x,y)$ with respect to x, $\hat{a}_x=\hat{f}_x(x,y)=\hat{a}_1+2\hat{a}_3x+\hat{a}_5y$, and with respect to y, $\hat{a}_y=\hat{f}_y(x,y)=\hat{a}_2+2\hat{a}_4y+\hat{a}_5x$, give the tangent planes (or slope) at these two directions. We may now obtain the average local slopes at the discrete points $1,\ldots,m$ as

$$\hat{a}_x=\frac{1}{m}\sum_{i=1}^{m}\hat{f}_x(x_i,y_i),\quad \hat{a}_y=\frac{1}{m}\sum_{i=1}^{m}\hat{f}_y(x_i,y_i), \tag{6}$$

for along- and across-track, respectively. The average angles α_x and α_y that the tangent plane makes with the positive x and y axes are $\alpha_x=\tan^{-1}a_x$ and $\alpha_y=\tan^{-1}a_y$, respectively. The least-squares method can provide us with the precision of the estimates. The covariance matrix of the estimates $\hat{a}$ is $Q_{\hat{a}}=(A^TQ_z^{-1}A)^{-1}$. One can thus obtain the standard deviation of a_x and a_y, and hence the standard deviation of α_x and α_y.

Finally, a practical comment on the coordinates transformation is in order. For the local surfaces usually the geographic coordinates north (N) and east (E) in the Universal Transverse Mercator (UTM) coordinate system are available and not directly the vessel frame coordinates x and y. These two sets of coordinate systems can be transformed to each other using the transformation

$$\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{bmatrix}\begin{bmatrix}N\\E\end{bmatrix}, \tag{7}$$

where α is the heading angle of the vessel. One way to estimate the along- and across-track slopes is to transform the coordinates to the vessel frame system using Eq. (7) and use the previous formulation. An alternative is to estimate the slopes in the E-N system (i.e., $\hat{a}_N=\hat{a}_1'+2\hat{a}_3'N+\hat{a}_5'E$ and $\hat{a}_E=\hat{a}_2'+2\hat{a}_4'E+\hat{a}_5'N$) and then transform them into the x-y system using (Appendix A)

$$\hat{a}_x=\hat{a}_N\cos\alpha+\hat{a}_E\sin\alpha,$$

$$\hat{a}_y=-\hat{a}_N\sin\alpha+\hat{a}_E\cos\alpha. \tag{8}$$

This equation is similar to the transformation of Eq. (7). Therefore, one can either transform the coordinates first and then estimate the slopes in the vessel frame system, or estimate the slopes first and then transform them using Eq. (8).

B. Grazing angle correction

Suppose that the local surface is estimated as $\hat{z}=\hat{f}(x,y)$. The average local slopes $\hat{a}_x$ and $\hat{a}_y$ of the surface are given by Eq. (6). The normal vector to this surface patch is (the gradient of the surface)

$$\vec{n}=[\hat{a}_x\ \hat{a}_y\ -1]^T=[\tan\alpha_x\ \tan\alpha_y\ -1]^T. \tag{9}$$

On the other hand, the nominal receiving-beam direction, which is based on the flat surface in the z-y plane, is $\vec{m}=[0-\cos\varphi\ \sin\varphi]^T=[0-\sin\theta\ \cos\theta]^T$, where θ is the nominal incident angle and $\varphi=\pi/2-\theta$ is the grazing angle (see Fig. 2). The angle between the two vectors $\vec{n}$ and $\vec{m}$ is the true incident angle and is given as

$$\cos\theta_t=\frac{\vec{n}\cdot\vec{m}}{\|\vec{n}\|\|\vec{m}\|}=\frac{\sin\varphi+a_y\cos\varphi}{\sqrt{1+a_x^2+a_y^2}}. \tag{10}$$

This equation can thus be used to obtain the true grazing angle $(\varphi_t=90-\theta_t)$ when both the along- and across-track slopes are available. In a special case when $a_x=0$ it follows, with $a_y=\tan\alpha_y$, from Eq. (10) that $\theta_t=90-(\varphi+\alpha_y)$ and hence $\varphi_t=\varphi+\alpha_y$.

C. Backscatter correction

Another correction due to the local slopes a_x and a_y is the fact that the signal footprint (ensonified area) will change if the surface is not flat. We now aim to correct the backscatter data for the local bottom slopes. The backscattering strength is obtained from the echo signal using the sonar equation[1]

$$\mathrm{BS}(\theta)=\mathrm{EL}(\theta)-\mathrm{SL}(\theta)+2\mathrm{TL}-10\log A_f(\theta), \tag{11}$$

where EL is the echo level, SL is the source level, TL is the transmission loss, and A_f is the true area of the signal footprint. The relation between true area A_f and nominal area A_f' (based on a flat surface) is $A_f=\lambda A_f'$, where λ is a scaling factor. This results in $\log A_f(\theta)=\log A_f'(\theta)+\log\lambda(\theta)$. The correction $C=-10\log\lambda(\theta)$ is then obtained as a function of local slopes in along- and across-track directions.

For angles away from, nadir the area A_f' is given as[1]

$$A_f'(\theta)=\frac{cTR\Omega_x}{2\sin\theta}=\frac{cTR\Omega_x}{2\cos\varphi}, \tag{12}$$

where Ω_x is the beam aperture in the along-track direction and φ is the grazing angle. The term $\delta_y=cT/2\sin\theta$ is the across-track resolution (size of the scatter pixel) of the backscatter imaging, and $\delta_x=R\Omega_x$ is the along-track resolution. When there exist significant bottom slopes, the area of the signal footprint may be modified to (Fig. 2)

$$A_f(\theta)=\frac{cTR\Omega_x}{2\sin(\theta-\alpha_y)\cos\alpha_x}. \tag{13}$$

The term $\cos\alpha_x$ in the denominator of the preceding equation indicates that the area is always larger in the along-track direction. This however, does not hold for the across-track direction as it depends on θ. The correction $C=-10\log\lambda(\theta)$ is then

$$C=10\log\left(\frac{\sin(\theta-\alpha_y)\cos\alpha_x}{\sin\theta}\right), \tag{14}$$

which is expressed in decibels. We now assume that the local slopes α_x and α_y are uncorrelated and have the same standard deviation σ_α. Application of the error propagation law[41] to the linearized form of Eq. (14) gives the standard deviation of the correction as

$$\sigma_C = \frac{10\,\sigma_\alpha}{\ln 10}(\tan^2(\varphi + \alpha_y) + \tan^2\alpha_x)^{1/2}, \tag{15}$$

which is obtained in decibels. One can further simplify this equation by using the approximate values of $\alpha_x = 0$ and $\alpha_y = 0$ as

$$\sigma_C = \frac{10\cot\theta\sigma_\alpha}{\ln 10}. \tag{16}$$

Equation (16) shows that the correction is not significant at low grazing angles, but it may not be neglected at high grazing angles.

For the near nadir beams (high grazing angles with $\theta \approx 0$), the area A_f' of a flat surface is given as

$$A_f' = R^2\Omega_x\Omega_y, \tag{17}$$

where Ω_y is the beam aperture in the across-track direction. With the presence of bottom slopes, the area A_f at normal incidence is modified as

$$A_f = \frac{R^2\Omega_x\Omega_y}{\cos\alpha_x\cos\alpha_y}. \tag{18}$$

V. RESULTS AND DISCUSSIONS

A. Experiment description

The river Waal is the main distributary branch of river Rhine flowing to the central Netherlands for about 80 km. It is a major river that serves as the main waterway connecting the Rotterdam harbor and Germany for commercial activities. Along several parts of this river, the bottom is subsiding. Since the subsidence varies along the river, dangerous shoals can occur. Appropriate sediment suppletion—it is a stable layer of concrete blocks—is planned to counteract the subsidence and to keep the riverbed more stable. To monitor the suppletion effectiveness, regular MBES measurements are planned, allowing for simultaneous estimation of bathymetry and sediment composition.

In October 2007, as a first step, MBES measurements were acquired at the Waal, accompanied with extensive sediment grabbing. The MBES used for the measurements is a Kongsberg EM3002, which was hull-mounted at a depth of 70 cm in the water. It typically works at a frequency of 300 kHz for shallow water (1–150 m). The depths of the area under survey range from 2 to 10 m. The EM3002 system used has a single sonar head with left and right transducers. Other technical specifications of this system are as follows: (1) The pulse length is 150 μs. (2) The maximum number of beams per ping is 254. (3) The maximum ping rate is 40 Hz. (4) The maximum angular coverage is 130°. (5) The beamwidth is $1.5° \times 1.5°$ at nadir. (6) The beam pattern is equidistant or equiangular. (7) The transducer geometry is mills cross.

The bathymetry of this study area is shown in Fig. 4. Except for the flat area (sediment suppletion to prevent deformation in the outer part of the bend) in the middle of the area, the river exhibits significant bottom slopes. This section presents the results of the acoustic sediment classification based on the methodology developed in Secs. IV and III. To assess the MBES classification results, a comparison is made with the analysis of the grab samples.

B. Acoustic classification results

We now apply the classification method of Sec. III to the above-described set of backscatter data. Figure 5 shows the histogram along with its best Gaussian fit for the original backscatter values at $\theta = 60°$ and $\theta = 62°$. The two Gaussian PDFs indicate that there exist two sediment types for the riverbed. Note that a third PDF was also found which has a small contribution of 0.3%. This third PDF is likely due to the heterogenity in the sediments or the violation of the normality assumption because we deal with a small number of scatter pixels per beam footprint. One may consider to apply the classification method based on these results.

There are, however, two issues that need to be addressed: (1) water depths are very shallow and (2) significant bottom slopes exist. The low depth results in a small number of scatterers within the signal footprint (N_s is low) and a small number of scatter pixels per beam footprint (N is low). In both cases, the central limit theorem is not valid, and consequently neither is the normality assumption of the backscatter data. These will affect the classification results. Also, the lower water depths correspond to higher ping-to-ping variability and hence higher fluctuation for backscatter data. Therefore, the discriminating power of the classification decreases as the two Gaussian PDFs are highly overlapped (Fig. 5). The second issue is the bottom slopes, which can significantly affect the backscatter data and the grazing angle.

One way out of this dilemma is to increase the number of samples for each beam considered (Secs. IV and III). This is achieved by including more angles around the central beam angle (e.g., with deviation of 1° as $\theta - 1° < \theta < \theta + 1°$). For such close angles, the angular dependence of the backscatter distribution can be ignored. One can also average over a few consecutive pings (e.g., 7 pings) because the ping rate (40 Hz) is high in shallow water. This results in a small surface patch that contains, say, 56 beams (Fig. 3). The averaged corrected (over patches) backscatter data will then be used.

The number of bottom types is unknown and needs to be determined according to the method described in Sec. III B. This is achieved by increasing the number of Gaussian functions to well describe the histogram of the averaged backscatter strength values after applying the slope corrections and averaging over the small patches. Figure 6 shows a plot of test statistic χ^2 in Eq. (3) versus the number r of the Gaussian PDFs. The value of r at which either the test statistic falls below the critical value or no further significant decrease in test statistic is obtained is the optimal value for r. This value is set to be $r = 3$ (Fig. 6), which is the "estimated" number of bottom types based on the acoustic data.

Figure 7 shows the histogram and its best Gaussian fit for the averaged backscatter values. Three Gaussian PDFs, indicating three acoustic classes, are identified. The previously detected small PDF (Fig. 5) is averaged out over the

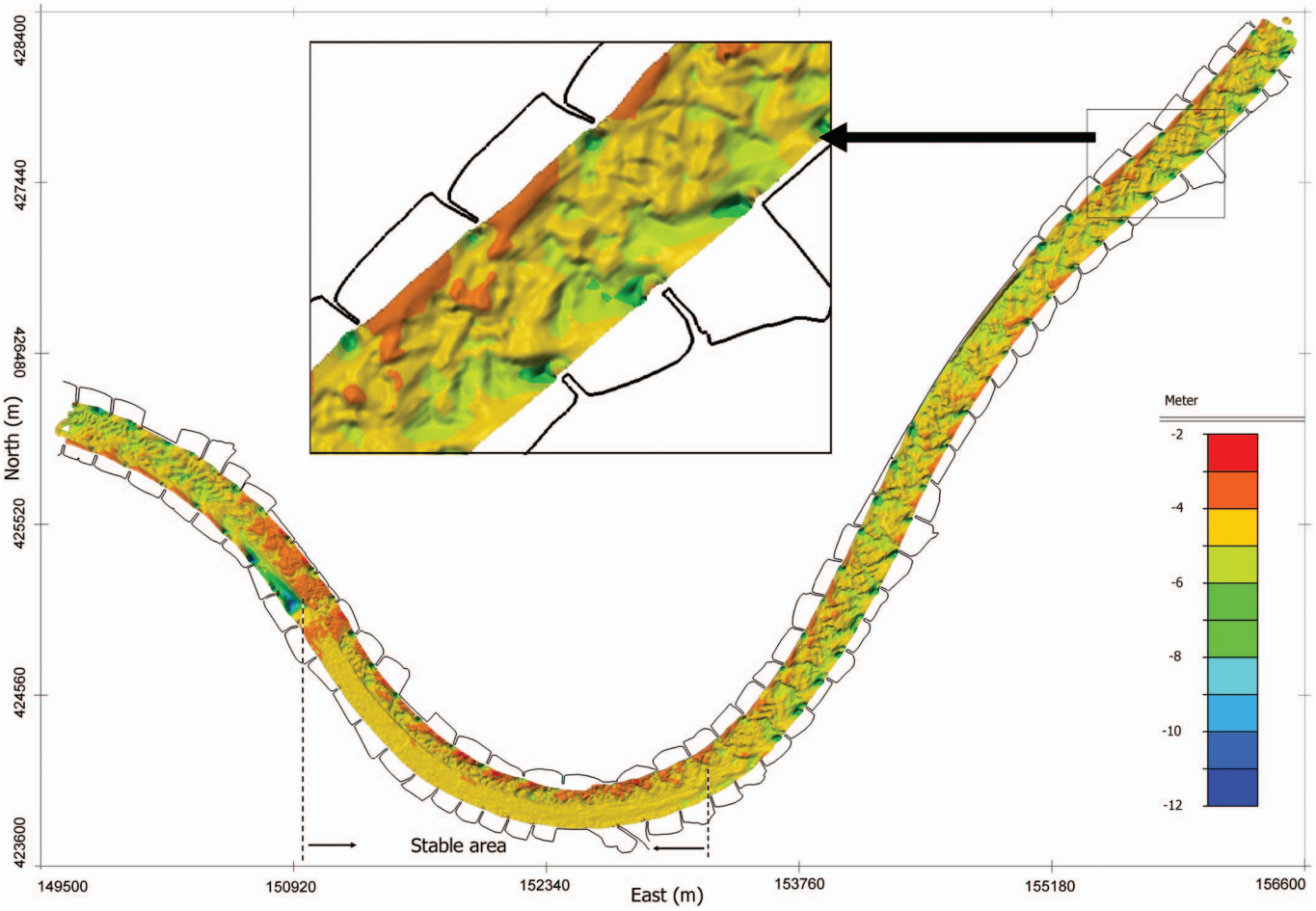


FIG. 4. Bathymetry map of river Waal, The Netherlands; km 920–930 [km 0 refers to a bridge in Constance, Switzerland (Ref. 42)].

small patches and it is not detectable anymore. However, a new PDF (a third one) is now detected and the Gaussian PDF distributions are better separated. The histograms are more peaked than those in Fig. 5, indicating lower variance and higher discriminating power. Also, the left tail of the histogram is now longer (left-skewed), indicating that the mass of the distribution is concentrated on the right hand side. The contributions of the PDFs are roughly 5%, 40%, and 55%. It is worthwile mentioning that the classification method is in-

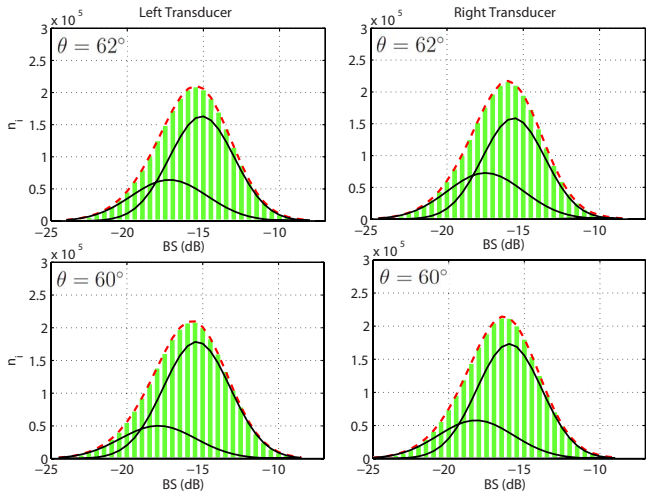


FIG. 5. (Color online) Histograms (light bar) of original measured backscatter data, its three Gaussians (solid line), and its best fit (dashed line) at angles θ=62° (top) and θ=60° (bottom) over the whole area; left and right transducers; number of Gaussians r=3 (third PDF is very small and located at μ_3=−10 dB).

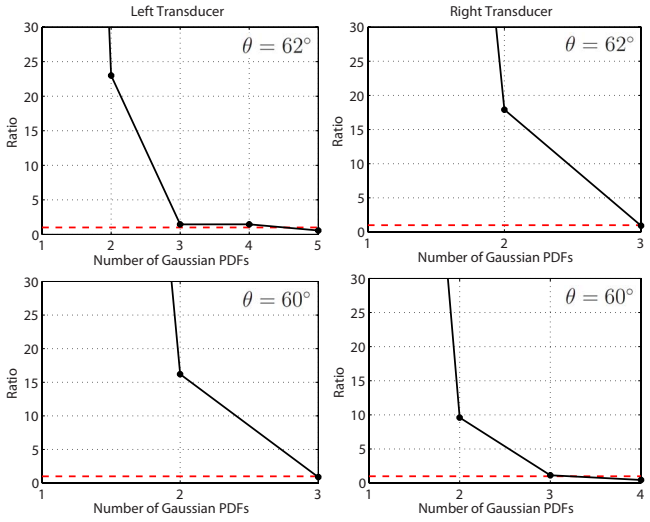


FIG. 6. (Color online) Normalized (ratio of test statistics to their corresponding critical values) chi-squared test statistic versus number r of Gaussian PDFs. Dashed lines indicate critical values which are set to one for normalized statistics; left and right transducers; data used at angles θ=62° (top) and θ=60° (bottom).

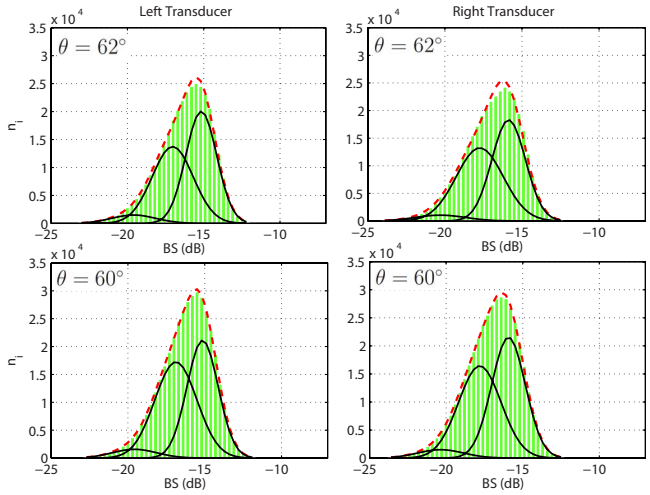


FIG. 7. (Color online) Histograms (light bar) of averaged (over small surface patches) backscatter data corrected for local slopes, its three Gaussians (solid line), and its best fit (dashed line) at angles θ=62° (top) and θ=60° (bottom) over the whole area; left and right transducers; number of Gaussians r=3.

dependent of the absolute values of the backscatter data. One may, for instance, think of the angular dependence of the backscatter data or the intrinsic difference between the backscatter data of the left and right transducers due to MBES transducer calibration effects.

To explore the full high-resolution mapping potential of the method, one may consider to use multiple beams instead of only one (Sec. III B, Step II). The ultimate goal of the acoustic classification method is to obtain a continuous map over the whole region, as for the bathymetry map. The classification map obtained from the averaged backscatter data using beam angles at θ=64°,62°,…,20° is shown in Fig. 8 in which the three sediment classes are presented by the colors red, yellow, and green. The green represents low values, the yellow represents intermediate values, and the red represents high values for the backscatter data. At a typical angle θ=60°, the acceptance regions are as follows: [−∞ to −18] dB (Class I), [−18 to −16.25] dB (Class II), and [−16.25 to +∞] dB (Class III).

In general, a correlation between the bathymetry and the classification results is observed (Figs. 4 and 8); the deeper the depth is, the larger the backscatter values are, and hence the coarser the sediments will be. That is what we would expect, and intuitively ground truth the classification results (see Sec. V C). Note, however, that there exist also shallow water areas where the sediment is coarse grained (compare Fig. 4 with Fig. 8). That is an indication for the absence of any depth-dependent artifacts or unmodeled effects in the backscatter data.

To further elaborate on the performance of the classifi-

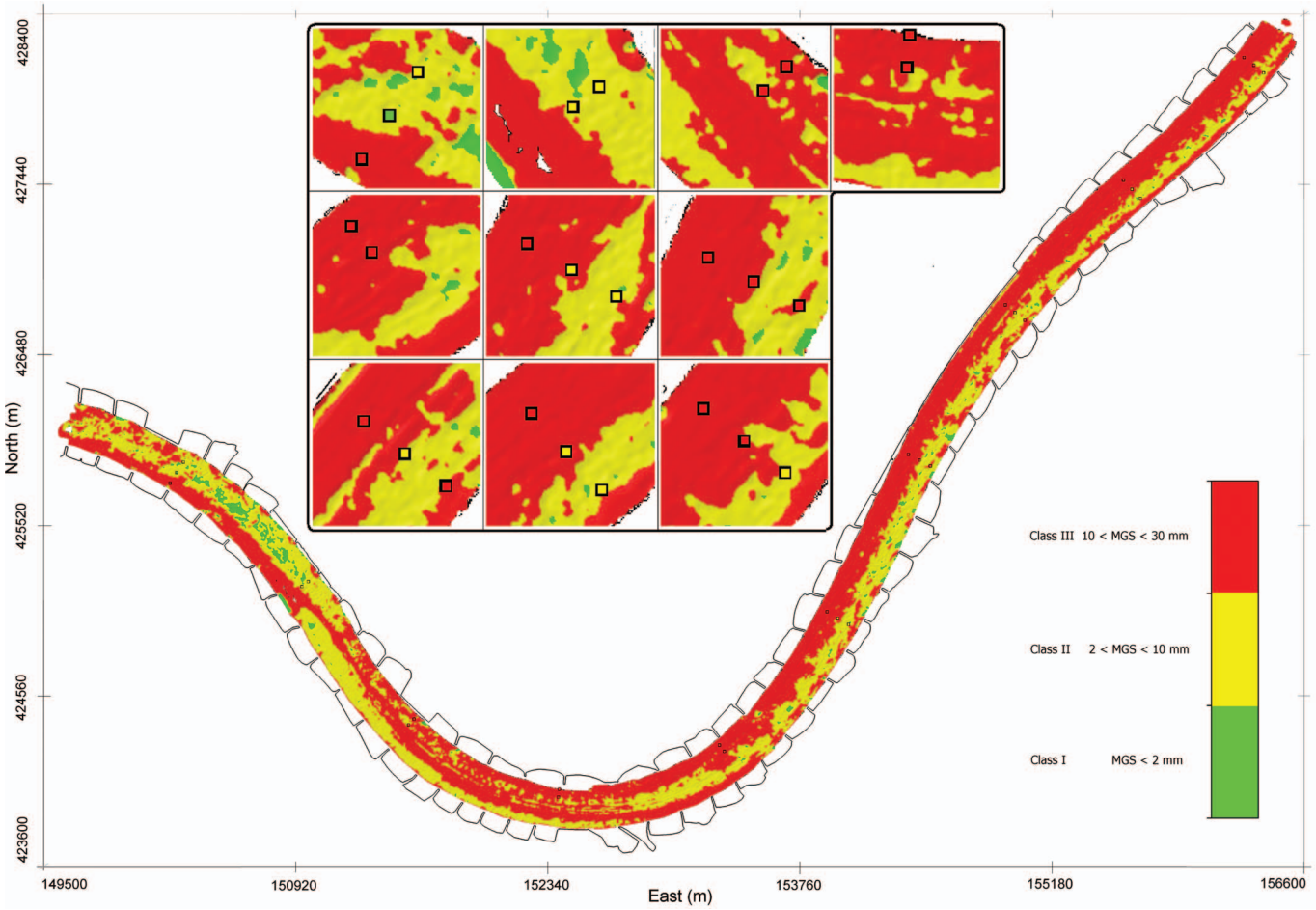


FIG. 8. Acoustic classification map of Waal river (km 920–930) obtained from backscatter at θ=64°,62°,…,20°. For each angle separate classification has been applied and results put in a single map. The frames on top indicate a zoom-in of classification results for areas where grab samples have been taken. A correlation of 0.75 is obtained between acoustic sediment classification and mean grain size from core analysis.

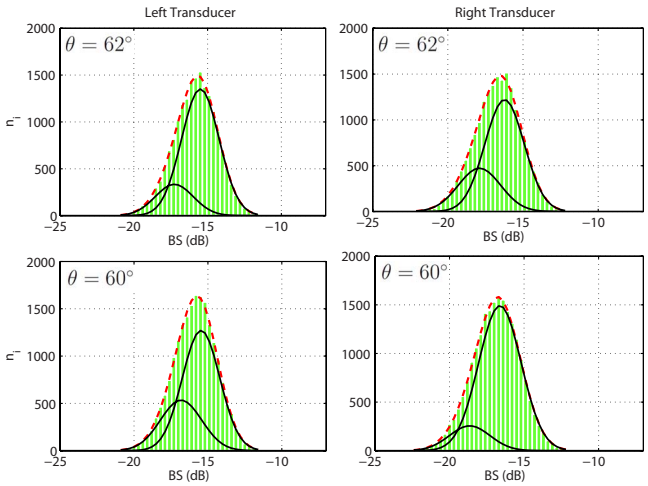


FIG. 9. (Color online) Histograms (light bar) of averaged (over small surface patches) backscatter data at θ=62° (top) and θ=60° (bottom) over the stable flat area; left and right transducers; number of Gaussians r=2.

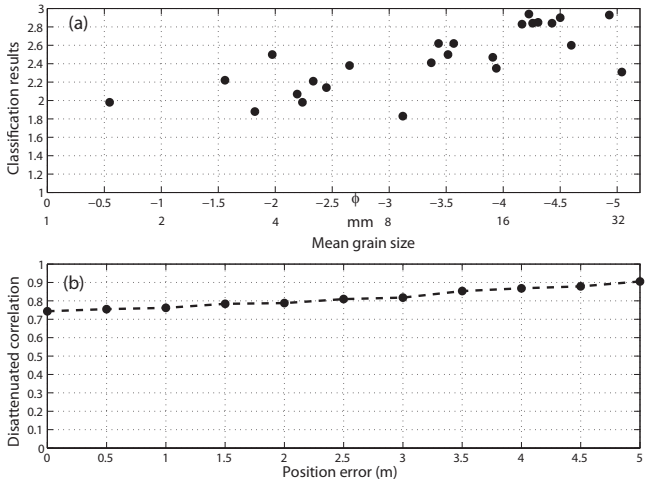


FIG. 10. Mean grain size from individual grab samples versus classification results (a). Disattenuated correlation coefficient (compensated for position error) between classification results and mean grain size of grabs (b). Note that ϕ scale has been inverted and positive correlation means that higher backscatter data are obtained over coarser sediment.

cation method, the averaged backscatter data over the southern part of a smaller area where the east component lies between 151 000 and 153 350 m—the stable flat area in Fig. 4—is now considered. The corresponding histogram along with two Gaussian PDFs is shown in Fig. 9. The left tails of the histograms are now shorter than those in Fig. 7. The two detected PDFs coincide with the results of the classification map using the three PDFs over the whole river (Fig. 8, two classes can be seen in this area).

C. Correlation with core analysis

The ultimate goal of MBES data analysis is to transform the backscatter classification results into estimates of seafloor sediment properties such as mean grain size. The goal of the sediment grab sampling and grain-size analysis is to evaluate the potential correlation between the mean grain size and the results from acoustic classification. A total number of 29 grab samples taken at the central axis of the river and at both sides (70 m apart from the central path) were collected and analyzed for grain-size distribution. There exist 25 samples which fall inside the survey area. The grab samples were washed, dried, and sieved through a series of mesh sizes ranging from 30 to 0.1 mm. The sieve sizes were converted into ϕ (phi) units[43] using the equation $\phi=-\log_2 d$, where d is diameter of grain in millimeters. Note that fine sediments have large ϕ values. Based on the comparison with the acoustic classification results, it can be concluded that the areas of high backscatter values correspond to gravel and lower backscatter values correspond to sand.

We now make a comparison between the classification results and the mean grain size of the samples. Our strategy is to use the results of the core analysis for comparison and to perform a correlation analysis afterward. The mean grain sizes were sorted from fine to coarse sediment. Considering the grab samples as an unbiased representative for the whole area, the percentages of 5%, 40%, and 55% were then applied to the 25 samples. This corresponds to 1, 10, and 14 samples, respectively for sand, gravelly sand, and sandy gravel areas. The classification results show good overall agreement with the ground truth information obtained from the core analysis [Fig. 8 zoom-in part, and Fig. 10(a)]. Most of the differences belong to the areas where the grab samples are in the boundary region of two classes. The dependence of acoustic backscatter techniques on sediment physical properties is examined using the Pearson correlation coefficient as well as a disattenuated correlation (Appendix B) between mean grain size of the samples and the classification results.

Due to the river currents' interaction with bottom sediments, the rivers are dynamic environments and hence sediment distribution can be highly heterogeneous. Ground truthing our classification results from core analysis of the sediments is prone to a few sources of uncertainty, of which we mention five. (i) Positioning error of the grab samples which is considered to be about 4–5 m: this issue is addressed in this section. (ii) The complexity inherent in ascertaining whether a single sample is representative of a larger region:[3] this originates from the heterogeneity of the river sediment distribution. (iii) A finite number of grab samples when assigning sediment types to acoustic classes, e.g., the percentage of the Class I (green) is 5% which leads to just one sample (if any) from 25 samples. (iv) Large standard deviation of backscatter data due to the shallowness of water, which leads to a small beam footprint: this has been accounted for, to a large extent, because of the averaging procedure. (v) Considering other physical properties of sediments rather than just the mean grain size. We can also use the full grain-size distribution and perform similar comparison.

We examine the potential correlation between classification results and the mean grain size. Larger grain sizes are expected to produce stronger backscatter for sandy and gravelly sediment. The Pearson correlation coefficient [Eq. (B7)] between the mean grain size and the results of the classification is 0.75. It indicates a high positive correlation (it is negatively correlated with ϕ values). The uncertainties (errors) mentioned above underestimate the correlation coefficient below the level it would have reached if the measure-

ments had been precise. Such uncertainties can be accounted for in the correlation coefficient, which gives rise to the disattenuated correlation coefficient (see Appendix B). This correlation coefficient indicates whether the correlation between two data sets is low(er) because of measurement error or because the two sets are really uncorrelated. Therefore, if the measure for the uncertainties is available (for example, if it is estimated by an independent method), one can then obtain the disattenuated correlation coefficients.

To see how significantly the aforementioned effects impact the correlation coefficients, the effect of position errors of the samples is assessed. The disattenuated correlation is given by Eq. (B9), where its disattenuation coefficient λ is obtained from Eq. (B10). Based on the data themselves, one cannot estimate the measurement errors σ_{e_x} and σ_{e_y} (variation within the sub-populations). Variation in the sediment composition (due to sediment heterogeneity) versus position can be simulated using the classification results in the following way: (i) consider randomly a large number of points with their true positions and classification results, (ii) simulate the position error and add it up with the true position, (iii) look at the classification results at the simulated point, and (iv) obtain its difference with the classification results at the true positions and hence its standard deviation σ_{e_x}. A positive trend for the disattenuated correlation is observed versus position error [Fig. 10(b)]. The correlation coefficient, when taking the 5 m position error into account, increases from 0.75 to 0.90. Note, however, that the disattenuated correlation is an indication for the presence of measurement errors and not a substitute for precise measurements. Future work can use and apply a similar correlation analysis to the grain-size distribution of the sediments.

D. *K*-distributed backscatter intensity

1. Background

The classical Rayleigh distribution theory is not applicable when the number of scatterers within the signal footprint is not large enough to apply the central limit theorem. This theory is not applicable, at least, when (1) the seafloor and hence seafloor data are rough, (2) the number of scatterers within the signal footprint is not large enough, (3) the number of scatterers is a random variable with high variance, and (4) the assumption of independent and identically distributed scatterers is violated.

Statistical analysis of backscatter intensity typically deals with fitting a set of theoretical distributions to see which one describes the data best. Non-Rayleigh distributions can better fit the skewed distributions and provide new parameters for characterization. It is widely accepted to use the *K*-distribution when the classical Rayleigh distribution is not applicable to backscatter amplitudes.[24–27,29] The *K*-distribution is

$$f_I(I) = \frac{2\left(\frac{N\nu}{\mu}\right)^{(N+\nu)/2} I^{(N+\nu-2)/2}}{\Gamma(\nu)\Gamma(N)} K_{\nu-N}\left(2\sqrt{\frac{N\nu}{\mu}I}\right), \tag{19}$$

where μ is the scale parameter, ν is the shape parameter, N is the multi-look parameter (i.e., the number of scatter pixels in the beam footprint), and $K_{\nu-N}$ is the modified Bessel function of the second kind. The *K*-distribution results from two independent Γ -distributed random variables. It has the advantage of being able to reproduce the classical distributions. For a given N, the *K*-distribution tends to the gamma distribution when the shape parameter increases to infinity ($\nu \to \infty$). In a special case when $N=1$, it then reduces to the exponential distribution.

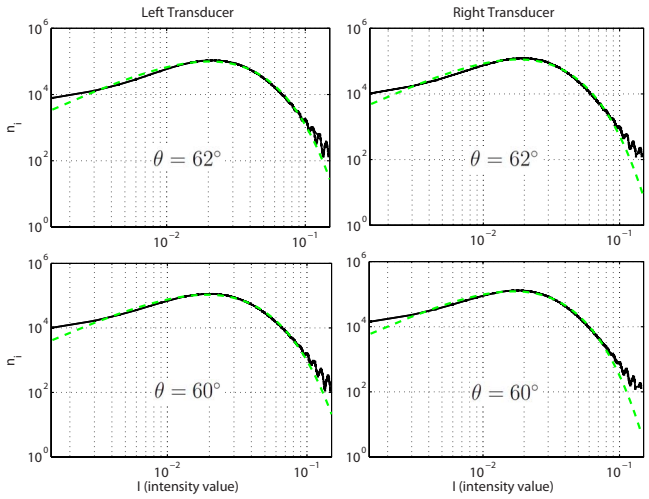


FIG. 11. (Color online) Measured histograms of backscatter intensity (solid black line) and statistical *K*-distribution (dashed light line) fitting; left and right transducers; θ=62° (top), θ=60° (bottom).

The maximum likelihood estimation method is usually applied to estimate the parameters μ and ν of the *K*-distribution.[44,45] An alternative is based on the method of moments.[46] We use a method which is based on the least-squares principle: fitting a curve to the histogram of the data in a least-squares sense (similar to Sec. III B, Step 1). Such estimates are first of all independent of the distribution of the data, second they are unbiased, and third they give the best possible precision (minimum variance) for the unknown parameters of the distribution.

The parameters μ and ν of the *K*-distribution depend on the incident angle. The *K*-distribution has proved to be a promising and useful model for the backscattering statistics in MBES and side-scan sonar data.[31,32] Also the *K*-distribution is of particular interest because its shape parameter is related to physical descriptors (e.g., spatial density of scatterers) of the seafloor.[29,30] Our application of the *K*-distribution is to further study the problem of riverbed characterization using the original backscatter intensities (without averaging). The angular evolution of the shape parameter ν is, in particular, investigated.

2. Results

The *K*-distribution is compared to the experimental PDF of the original backscatter intensities. Figure 11 shows typical graphical examples of the observed backscatter intensities along with their least-squares fit. The goodness of fit criterion [χ^2 values, similar to those in Eq. (3)] are as follows. For left transducer they are χ^2=242 and χ^2=209 at θ =62° and θ=60°, respectively. For right transducer they are χ^2=333 and χ^2=280 at θ=62° and θ=60°, respectively (see

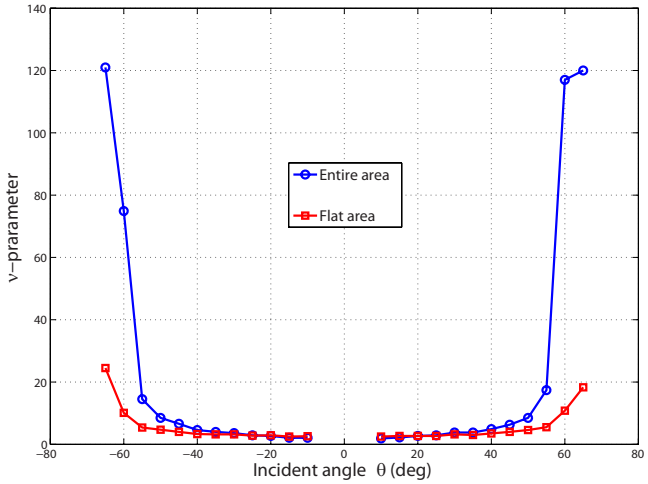


FIG. 12. (Color online) Angular evolution of shape parameter ν of K-distribution for entire area and flat area.

Fig. 6). The critical value is only 1.43. This indicates that the fit is not good enough because of large χ^2 values, which is a further indication for having more than one sediment types (this came out of the Gaussian fitting).

The angular evolution of backscattering statistics via the evolution of one parameter, the shape parameter ν in Eq. (19), of the K-distribution is now investigated. Its estimate is based on the least-squares curve fitting using the simultaneous estimation of the shape parameter ν and the scale parameter μ. The results for ν are shown in Fig. 12. The angular evolution of the shape parameter ν coincides with the findings of Refs. 31 and 32. For the intermediate incident angles, the shape-parameter values are low. At high incident (low grazing) angles, the increase is due to the extension of the beam footprint, which includes a greater number of scatterers; the central limit theorem applies, and the K-distribution tends to an exponential distribution (and correspondingly the Rayleigh distribution for amplitude).

The results given in Ref. 32 show a point where the functional behavior of the shape-parameter curves [i.e., $\nu = \nu(\theta)$] reverse for soft sediments, and that the rough seafloor does not seem to exhibit this transition angle. They indicate that the riverbed can be considered to be a rough surface (we cannot see such transition point here as the shape parameter increases with θ). We have already identified that the river sediment composition is formed primarily of coarse sand and gravel. This makes sense because the grain size of the sediment is a major contributor to the surface roughness.

We also observe that the shape parameters, in the stable flat area, are significantly smaller than their corresponding values in the entire area (Fig. 12). This can be considered as the effect of the bottom slopes, which is not seriously present in the flat area. The strong local slope variation as a normally distributed random variable will further increase the shape parameter of the K-distribution. Therefore, the K-distribution tends to an exponential distribution as in a Rayleigh reverberation process.

VI. SUMMARY AND CONCLUSIONS

Riverbed sediment classification using MBESs, backscatter data is a promising approach. The degree to which different bottom types can be discriminated using the backscatter data depends on the following. (i) Geoacoustical features of the bottom types: an obvious effect is on the mean and variance of the backscatter data. (ii) Measurement configuration such as beam grazing angle, water depth, pulse length, and number of scatterers in the signal footprint: such issues usually affect the distribution of the backscatter data. (iii) The presence of local slopes of the seafloor. In principle, the backscatter data should be corrected for bottom slopes. The differential slopes might, however, significantly affect the distribution parameters.

This contribution presented a methodology to use the (very) high-frequency MBES backscatter data for the sediment classification in very shallow water applications (depth of 2–10 m). The MBES used is an EM3002, typically working at 300 kHz with the maximum 254 number of beams. However, there is no restriction regarding the capability of the method for other water depths and frequencies. This method employs the MBES backscatter data to obtain the number of classes and to discriminate between them by applying the Bayes decision rule for multiple hypotheses. This is achieved by fitting a series of Gaussian PDFs to the backscatter strength histogram. Since the classification is done per beam, the method is independent of the possible incorrect calibration effects and the angular behavior of the backscatter data.

The performance of the method was tested by using the backscatter data acquired in the river Waal, The Netherlands. Extensive sediment grab samples analyzed for the grain-size distribution were used to evaluate the performance of the classification results. The following aspects of the research are highlighted.

- Shallow water depths result in small beam footprints and hence a small number of scatter pixels per beam. That makes the backscatter data highly variable and consequently the classification method becomes less efficient. To increase the discriminating power of the classification results, we used an averaging procedure over small surface patches 0.5×0.5 m^2. The high resolution bathymetry data provide precise bottom slope corrections to convert the arrival angle of the signal into the true incident angle, and the high resolution backscatter data allow one to reduce the statistical fluctuation in backscatter strength.
- We performed a correlation analysis. The dependence of acoustic backscatter classification on sediment physical properties was verified by using Pearson (0.75) and disattenuated (0.90) correlation coefficients between the classification results and sediment mean grain size. The disattenuated correlation gives an indication for the effects of measurement errors that attenuate the correlation below the level it would have reached had the measurements been precise.
- We considered the backscattered intensity statistics using the K-distribution to further study the riverbed characterization. Angular evolution of the K-distribution shape parameter indicated that the Waal riverbed is indeed a rough surface. It is in agreement with the ground truth information from the core analysis.

ACKNOWLEDGMENTS

The research is financially supported by the Dutch Ministry of Transportation and Water Management, Rijkswaterstaat. We acknowledge, in particular, Dr. Arjan Sieben, Ir. Ben Dierikx, and Adri Wagener for providing the Waal data and for their support during this project. We would also like to acknowledge the associate editor and two anonymous reviewers for their constructive comments which improved the presentation and quality of the paper.

APPENDIX A: COORDINATE TRANSFORMATION

The linear model of observation equations in the vessel coordinate system x-y is

$$E(z) = Aa, \tag{A1}$$

where z is the bathymetry data, A is the $m \times 6$ design matrix of the form

$$A = [1\ X\ Y\ X^2\ Y^2\ XY], \tag{A2}$$

or, written out as

$$A = \begin{bmatrix} 1 & x_1 & y_1 & x_1^2 & y_1^2 & x_1 y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & y_m & x_m^2 & y_m^2 & x_m y_m \end{bmatrix}, \tag{A3}$$

and a is the unknown coefficients of the bi-quadratic polynomial. The least-squares estimate of a is

$$\hat{a} = (A^T Q_z^{-1} A)^{-1} A^T Q_z^{-1} z. \tag{A4}$$

Another parametrization of the above system of observation equations is based on the UTM coordinate system (north N and east E)

$$E(z) = A' a', \tag{A5}$$

where

$$A' = [1\ N\ E\ N^2\ E^2\ NE]. \tag{A6}$$

The prime $'$ indicates that the terms are now defined in the new coordinate system. The least-squares estimate for a' is

$$\hat{a}' = (A'^T Q_z^{-1} A')^{-1} A'^T Q_z^{-1} z. \tag{A7}$$

One can simply show that the design matrices A and A' are related using the transformation $A = A'T$, where T is a 6×6 regular matrix of the form

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha & 0 & 0 & 0 \\ 0 & \sin\alpha & \cos\alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos^2\alpha & \sin^2\alpha & \frac{-1}{2}\sin 2\alpha \\ 0 & 0 & 0 & \sin^2\alpha & \cos^2\alpha & \frac{1}{2}\sin 2\alpha \\ 0 & 0 & 0 & \sin 2\alpha & -\sin 2\alpha & \cos 2\alpha \end{bmatrix}, \tag{A8}$$

with the inverse

$$T^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 & 0 & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos^2\alpha & \sin^2\alpha & \frac{1}{2}\sin 2\alpha \\ 0 & 0 & 0 & \sin^2\alpha & \cos^2\alpha & \frac{-1}{2}\sin 2\alpha \\ 0 & 0 & 0 & -\sin 2\alpha & \sin 2\alpha & \cos 2\alpha \end{bmatrix}, \tag{A9}$$

where α is the heading angle of the vessel. Substituting $A = A'T$ in Eq. (A4) gives the relation between the least-squares estimates $\hat{a}$ and $\hat{a}'$ as

$$\hat{a} = T^{-1}(A'^T Q_z^{-1} A')^{-1} A'^T Q_z^{-1} z = T^{-1}\hat{a}' \tag{A10}$$

or

$$\begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \end{bmatrix} = \begin{bmatrix} \hat{a}'_0 \\ \hat{a}'_1 \cos\alpha + \hat{a}'_2 \sin\alpha \\ \hat{a}'_2 \cos\alpha - \hat{a}'_1 \sin\alpha \\ \hat{a}'_3 \cos^2\alpha + \hat{a}'_4 \sin^2\alpha + \frac{1}{2}\hat{a}'_5 \sin 2\alpha \\ \hat{a}'_3 \sin^2\alpha + \hat{a}'_4 \cos^2\alpha - \frac{1}{2}\hat{a}'_5 \sin 2\alpha \\ \hat{a}'_4 \sin 2\alpha - \hat{a}'_3 \sin 2\alpha + \hat{a}'_5 \cos 2\alpha \end{bmatrix}. \tag{A11}$$

Equations $\hat{a}_x = \hat{a}_1 + 2\hat{a}_3 x + \hat{a}_5 y$ and $\hat{a}_y = \hat{a}_2 + 2\hat{a}_4 y + \hat{a}_5 x$ with $x = N\cos\alpha + E\sin\alpha$, $y = E\cos\alpha - N\sin\alpha$, and Eq. (A11) simplify to

$$\hat{a}_x = (\hat{a}'_1 + 2\hat{a}'_3 N + \hat{a}'_5 E)\cos\alpha + (\hat{a}'_2 + 2\hat{a}'_4 E + \hat{a}'_5 N)\sin\alpha, \tag{A12}$$

$$\hat{a}_y = (\hat{a}'_2 + 2\hat{a}'_4 E + \hat{a}'_5 N)\cos\alpha - (\hat{a}'_1 + 2\hat{a}'_3 N + \hat{a}'_5 E)\sin\alpha, \tag{A13}$$

or simply

$$\hat{a}_x = \hat{a}_N \cos\alpha + \hat{a}_E \sin\alpha, \tag{A14}$$

$$\hat{a}_y = \hat{a}_E \cos\alpha - \hat{a}_N \sin\alpha, \tag{A15}$$

where $\hat{a}_N = \hat{a}'_1 + 2\hat{a}'_3 N + \hat{a}'_5 E$ and $\hat{a}_E = \hat{a}'_2 + 2\hat{a}'_4 E + \hat{a}'_5 N$ are the estimated slopes in the north and east directions, respectively.

APPENDIX B: DISATTENUATED CORRELATION COEFFICIENT

Consider the data set $x = [x_1, \ldots, x_m]^T$ and $y = [y_1, \ldots, y_m]^T$, as a realization of the following random variables: $\underline{x}_i = \mu_x + \epsilon_{x_i} + e_{x_i}$ and $\underline{y}_i = \mu_y + \epsilon_{x_i} + e_{y_i}$, where the μ's represent the mean values, the ϵ's represent variation *between* sub-population, and the e's represent variation *within* the sub-populations (measurement error). The underline indicates randomness. The variances of the components ϵ_{x_i}, ϵ_{y_i}, e_{x_i}, and e_{y_i} are assumed to be $\sigma^2_{\epsilon_x}$, $\sigma^2_{\epsilon_y}$, $\sigma^2_{e_x}$, and $\sigma^2_{e_y}$, respectively. The ϵ_x and ϵ_y are assumed to be correlated ($\sigma_{\epsilon_x \epsilon_y} \neq 0$), but the measurement errors e_x and e_y are assumed to be uncorrelated ($\sigma_{e_x e_y} = 0$). The above formulas can be written in a matrix notation as

$$\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \epsilon_x \\ \epsilon_y \\ e_x \\ e_y \end{bmatrix}, \tag{B1}$$

with the covariance matrix of

$$D\left\{\begin{bmatrix} \epsilon_x \\ \epsilon_y \\ e_x \\ e_y \end{bmatrix}\right\} = \begin{bmatrix} \sigma_{\epsilon_x}^2 & \sigma_{\epsilon_x\epsilon_y} & 0 & 0 \\ \sigma_{\epsilon_x\epsilon_y} & \sigma_{\epsilon_y}^2 & 0 & 0 \\ 0 & 0 & \sigma_{e_x}^2 & 0 \\ 0 & 0 & 0 & \sigma_{e_y}^2 \end{bmatrix}. \tag{B2}$$

Application of the variance propagation law[47] gives the covariance matrix of $[\underline{x},\underline{y}]^T$ as

$$D\left\{\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\right\} = \begin{bmatrix} \sigma_{\epsilon_x}^2 + \sigma_{e_x}^2 & \sigma_{\epsilon_x\epsilon_y} \\ \sigma_{\epsilon_x\epsilon_y} & \sigma_{\epsilon_y}^2 + \sigma_{e_y}^2 \end{bmatrix}. \tag{B3}$$

If we have a series of m samples of random variables x and y, written as x_i and y_i, where $i=1,\ldots,m$, then the variances $\sigma_x^2=\sigma_{\epsilon_x}^2+\sigma_{e_x}^2$ and $\sigma_y^2=\sigma_{\epsilon_y}^2+\sigma_{e_y}^2$ and the covariance $\sigma_{xy}=\sigma_{\epsilon_x\epsilon_y}$ can be estimated using the following linear model of observation equations:

$$E\left\{\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\right\} = \begin{bmatrix} u & 0 \\ 0 & u \end{bmatrix} \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \tag{B4}$$

with the covariance matrix

$$D\left\{\begin{bmatrix} \underline{x} \\ \underline{y} \end{bmatrix}\right\} = \begin{bmatrix} \sigma_x^2 I & \sigma_{xy} I \\ \sigma_{xy} I & \sigma_y^2 I \end{bmatrix}, \tag{B5}$$

where $u=[1,1,\ldots,1]^T$ is a summation vector. The least-squares estimates of σ_x^2, σ_y^2, and σ_{xy} are then[40,48]

$$\hat{\sigma}_x^2 = (x-\bar{x})^T(x-\bar{x})/(m-1),$$

$$\hat{\sigma}_y^2 = (y-\bar{y})^T(y-\bar{y})/(m-1),$$

$$\hat{\sigma}_{xy} = (x-\bar{x})^T(y-\bar{y})/(m-1), \tag{B6}$$

where $\bar{x}=(1/m)\Sigma_{i=1}^m x_i$ and $\bar{y}=(1/m)\Sigma_{i=1}^m y_i$ are the sample means. The Pearson correlation coefficient (sample correlation) can then be used to estimate the correlation between x and y:[49]

$$\rho_{xy} = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2\hat{\sigma}_y^2}} = \frac{\hat{\sigma}_{\epsilon_x\epsilon_y}}{\sqrt{(\hat{\sigma}_{\epsilon_x}^2+\hat{\sigma}_{e_x}^2)(\hat{\sigma}_{\epsilon_y}^2+\hat{\sigma}_{e_y}^2)}}. \tag{B7}$$

When two data sets x and y are correlated, measurement errors underestimate the correlation coefficient. Measurement error can be accounted for in a correlation coefficient, which gives rise to the correlation coefficient disattenuated of measurement error

$$\rho_{xy}^d = \frac{\hat{\sigma}_{\epsilon_x\epsilon_y}}{\sqrt{\hat{\sigma}_{\epsilon_x}^2\hat{\sigma}_{\epsilon_y}^2}}. \tag{B8}$$

Disattenuated correlation coefficient indicates whether the correlation between two data sets is low because of measurement error or because the two sets are really uncorrelated. Note that Eq. (B6) gives $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$, and not separately $\hat{\sigma}_{\epsilon_x}^2$, $\hat{\sigma}_{e_x}^2$, $\hat{\sigma}_{\epsilon_y}^2$, and $\hat{\sigma}_{e_y}^2$. If the measurement errors are available (for example, if they are estimated by an independent tool such as simulation), one can account for them and obtain the disattenuated correlation coefficients. The relation between Pearson correlation coefficient and its disattenuated one is

$$\rho_{xy} = \lambda\rho_{xy}^d, \tag{B9}$$

where λ is the attenuation coefficient

$$\lambda = \frac{\sigma_{\epsilon_x}\sigma_{\epsilon_y}}{\sqrt{(\sigma_{\epsilon_x}^2+\sigma_{e_x}^2)(\sigma_{\epsilon_y}^2+\sigma_{e_y}^2)}}. \tag{B10}$$

How well the variables are measured affects the correlation of x and y. The correction for attenuation shows the correlation as if one measures x and y without errors.

[1] X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications* (Springer, New York, 2002), pp. 90–91.
[2] D. G. Simons and M. Snellen, "A Bayesian approach to seafloor classification using multi-beam echo-sounder backscatter data," Appl. Acoust. **70**, 1258–1268 (2009).
[3] J. A. Goff, H. C. Olson, and C. S. Duncan, "Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin," Geo-Mar. Lett. **20**, 43–49 (2000).
[4] J. S. Collier and C. J. Brown, "Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments," Mar. Geol. **214**, 431–449 (2005).
[5] J. C. Borgeld, J. E. Hughes Clarke, J. A. Goff, L. A. Mayer, and J. A. Curtis, "Acoustic backscatter of the 1995 flood deposit on the Eel shelf," Mar. Geol. **154**, 197–210 (1999).
[6] R. Urgeles, J. Locat, T. Schmitt, and J. E. Hughes Clarke, "The July 1996 flood deposit in the Saguenay Fjord, Quebec, Canada: Implications for sources of spatial and temporal backscatter variations," Mar. Geol. **184**, 41–60 (2002).
[7] V. L. Ferrini and R. D. Flood, "The effects of fine-scale surface roughness and grain size on 300 kHz multibeam backscatter intensity in sandy marine sedimentary environments," Mar. Geol. **228**, 153–172 (2006).
[8] D. G. Simons, M. Snellen, and M. Ainslie, "A multivariate correlation analysis of high-frequency bottom backscattering strength measurements with geotechnical parameters," IEEE J. Ocean. Eng. **32**, 640–650 (2007).
[9] T. Medialdea, L. Somoza, R. León, M. Farrán, G. Ercilla, A. Maestro, D. Casas, E. Llave, F. J. Hernández-Molina, M. C. Fernández-Puga, and B. Alonso, "Multibeam backscatter as a tool for sea-floor characterization and identification of oil spills in the Galicia Bank," Mar. Geol. **249**, 93–107 (2008).
[10] D. Alexandrou and D. Pantzartzis, "A methodology for acoustic seafloor classification," IEEE J. Ocean. Eng. **18**, 81–86 (1993).
[11] Z. H. Michalopoulou, D. Alexandrou, and C. de Moustier, "Application of neural and statistical classifiers to the problem of seafloor characterization," IEEE J. Ocean. Eng. **20**, 190–197 (1995).
[12] I. Marsh and C. Brown, "Neural network classification of multibeam backscatter and bathymetry data from Stanton Bank (Area IV)," Appl. Acoust. **70**, 1269–1276 (2009).
[13] L. Fonseca and L. Mayer, "Remote estimation of surficial seafloor properties through the application angular range analysis to multibeam sonar data," Mar. Geophys. Res. **28**, 119–126 (2007).
[14] L. Fonseca, C. Brown, B. Calder, L. Mayer, and Y. Rzhanov, "Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures," Appl. Acoust. **70**, 1298–1304 (2009).
[15] X. Zhou and Y. Chen, "Seafloor classification of multibeam sonar data using neural network approach," Mar. Geodesy **28**, 201–206 (2005).
[16] J. Hughes Clarke, "Toward remote seafloor classification using the angular response of acoustic backscattering: A case study from multiple overlapping GLORIA data," IEEE J. Ocean. Eng. **19**, 112–127 (1994).
[17] I. M. Parnum, A. N. Gavrilov, P. J. W. Siwabessy, and A. J. Duncan, "Analysis of high-frequency multibeam backscatter statistics from different seafloor habitats," in *Proceedings of the Eighth European Conference*

on Underwater Acoustics, Carvoeiro, Portugal (2006), pp. 1020–1031.
[18]G. S. Brown, "Backscattering from Gaussian-distributed perfectly conducting rough surface," IEEE Trans. Antennas Propag. **AP-26**, 472–482 (1978).
[19]F. T. Ulaby, T. F. Haddock, and R. T. Austin, "Fluctuation statistics of millimeter-wave scattering from distributed targets," IEEE Trans. Geosci. Remote Sens. **26**, 268–281 (1988).
[20]M. Gensane, "A statistical study of acoustic signals backscattered from the sea bottom," IEEE J. Ocean. Eng. **14**, 84–93 (1989).
[21]E. Jakeman and P. N. Pusey, "A model for non-Rayleigh sea echo," IEEE Trans. Antennas Propag. **AP-24**, 806–814 (1976).
[22]E. Jakeman and P. N. Pusey, "Significance of K-distributions in scattering experiments," Phys. Rev. Lett. **40**, 546–550 (1978).
[23]C. J. Oliver, "A model for non-Rayleigh scattering statistics," Opt. Acta **31**, 701–722 (1984).
[24]E. Jakeman, "Non-Gaussian models for the statistics of the scattered waves," Adv. Phys. **37**, 471–529 (1988).
[25]C. J. Oliver, "Correlated K-distributed clutter models," Opt. Acta **32**, 1515–1547 (1985).
[26]J. Dunlop, "Statistical modeling of sidescan sonar images," in *Oceans Conference Record (IEEE)*, Piscataway, NJ (1997), Vol. **1**, pp. 33–37.
[27]A. P. Lyons and D. A. Abraham, "Statistical characterization of high-frequency shallow-water seafloor backscatter," J. Acoust. Soc. Am. **106**, 1307–1315 (1999).
[28]F. P. A. Lam, N. J. Konijnendijk, J. Groen, and D. G. Simons, "Non-Rayleigh wideband sonar reverberation modeling including hybrid multipaths," in *Proceedings of the MTS/IEEE Oceans'06*, Boston, MA, edited by M. Prior and D. Abraham (IEEE Operations Center, Piscataway, NJ, 2006).
[29]D. A. Abrahams and A. P. Lyons, "Novel physical interpretations of K-distributed reverberation," IEEE J. Ocean. Eng. **27**, 800–813 (2002).
[30]M. K. Prior, "Estimation of K-distribution shape parameter from sonar data: Sample size limitations," IEEE J. Ocean. Eng. **34**, 45–50 (2009).
[31]L. Hellequin and J. M. Boucher, "Processing of high-frequency multibeam echo sounder data for seafloor characterization," IEEE J. Ocean. Eng. **28**, 78–89 (2003).
[32]G. Le Chenadec and J. M. Boucher, "Angular dependence of *K*-distributed sonar data," IEEE Trans. Geosci. Remote Sens. **45**, 1224–1235 (2007).
[33]E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
[34]R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest and smallest member of a sample," Proc. Cambridge Philos. Soc. **24**, 180–190 (1928).
[35]T. F. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," SIAM J. Optim. **6**, 418–445 (1996).
[36]C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems* (Prentice-Hall, Englewood Cliffs, NJ, 1974).
[37]J. A. Goff, B. J. Kraft, L. A. Mayer, S. G. Schock, C. K. Sommerfield, H. C. Olson, S. P. S. Gulick, and S. Nordfjord, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," Mar. Geol. **209**, 147–172 (2004).
[38]FLEDERMAUS software: Interactive visualization systems (IVS 3D), a 3D visualization and analysis software, http://www.ivs3d.com/ (Last viewed 3/31/09).
[39]P. J. G. Teunissen, *Testing Theory: An Introduction* (Delft University Press, Delft, The Netherlands, 2000), pp. 132–136, http://www.vssd.nl/hlf/landmeet.html (Last viewed 7/12/09).
[40]A. R. Amiri-Simkooei, "Least-squares variance component estimation: Theory and GPS applications," Ph.D. thesis, Delft University of Technology, Delft, The Netherlands (2007).
[41]P. J. G. Teunissen, *Adjustment Theory: An Introduction* (Delft University Press, Delft, The Netherlands, 2000), pp. 178–191, http://www.vssd.nl/hlf/landmeet.html (Last viewed 7/12/09).
[42]International Committee of the hydrology of the Rhine basin (CHR), basin of the Rhine, Hydrological monograph, CHR/KHR, The Hague, The Netherlands, 1977.
[43]C. K. Wentworth, "A scale of grade and class terms for clastic sediments," J. Geol. **30**, 377–392 (1922).
[44]C. J. Oliver, "Optimum texture estimators for SAR clutter," J. Phys. D: Appl. Phys. **26**, 1824–1835 (1993).
[45]I. R. Joughin, D. B. Percival, and D. P. Winebrenner, "Maximum likelihood estimation of K distribution parameters for SAR data," IEEE Trans. Geosci. Remote Sens. **31**, 989–999 (1993).
[46]C. J. Oliver, "Information from SAR images," J. Phys. D **24**, 1493–1514 (1991).
[47]P. J. G. Teunissen, D. G. Simons, and C. C. J. M. Tiberius, *Probability and Observation Theory* (Delft University, Delft, The Netherlands, 2005).
[48]P. J. G. Teunissen and A. R. Amiri-Simkooei, "Least-squares variance component estimation," J. Geodesy, Berlin **82**, 65–82 (2008).
[49]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, New York, 1992).

On the consideration of motion effects in the computation of impulse response for underwater acoustics inversion

Nicolas F. Josso,[a] Cornel Ioana, and Jérôme I. Mars
GIPSA-lab, Grenoble Institute of Technology (GIT), 961 rue de la Houille Blanche, 38402 St Martin d'Hères, France

Cédric Gervaise
E3I2, EA3876, ENSIETA, Université Européenne de Bretagne, 2 rue Francois Verny, 29806 Brest Cedex, France

Yann Stéphan
SHOM, CS52817, 13 rue du Chatellier 29228 Brest Cedex 2, France

(Received 14 January 2009; revised 7 July 2009; accepted 9 July 2009)

The estimation of the impulse response (IR) of a propagation channel may be of great interest for a large number of underwater applications: underwater communications, sonar detection and localization, marine mammal monitoring, etc. It quantifies the distortions of the transmitted signal in the underwater channel and enables geoacoustic inversion. The propagating signal is usually subject to additional and undesirable distortions due to the motion of the transmitter-channel-receiver configuration. This paper shows the effects of the motion while estimating the IR by matched filtering between the transmitted and the received signals. A methodology to compare IR estimation with and without motion is presented. Based on this comparison, a method for motion effect compensation is proposed in order to reduce motion-induced distortions. The proposed methodology is applied to real data sets collected in 2007 by the Service Hydrographique et Océanographique de la Marine in a shallow water environment, proving its interest for motion effect analysis. Motion compensated estimation of IRs is computed from sources transmitting broadband linear frequency modulations moving at up to 12 knots in the shallow water environment of the Malta plateau, South of Sicilia.
 [DOI: 10.1121/1.3203308]

PACS number(s): 43.30.Pc, 43.60.Mn, 43.60.Pt, 43.30.Cq [AIT] Pages: 1739–1751

I. INTRODUCTION

The knowledge of the impulse response (IR) of propagation channels is potentially interesting for a large number of underwater acoustics applications such as underwater communication, sonar detection and localization, marine mammal monitoring, etc. The IR estimate is also central for geoacoustic inversion using matched impulse response (MIR) techniques.[1] The most popular method to estimate IR is the so-called matched filtering,[2] where the received signal is correlated with the transmitted one.

Ideally, with an additive white-noise background, the matched filtering operation processing correlates the received signal with time-delayed versions of the transmitted signal. When the motion of the transmitter and receiver is well monitored, as in the case of active ocean acoustic tomography, methods such as matched-field-processing.[3] or MIR can take into account the motion effects, even if a highly computational cost may be required for broadband signals. When the motion of the transmitter and receiver is unknown (as it is the case for passive ocean acoustic tomography[4,5]), these methods cannot be applied any longer. In this paper, we propose a new method to estimate and compensate the motion effects. This work is a contribution to the development of a passive tomography system using transient signals.

When the motion of the transmitter-channel-receiver configuration is not known, the received signal could be correlated against a family of reference signals that represent as well all possible receptions. The set of reference signals would account for all the possible velocities of the configuration and the multipath propagation effects of the environment. For example, Qian and Chen[6] and Mallat and Zhang[7] proposed a matching pursuit algorithm on transitory signals, which adaptively decomposes any signal into a linear combination of best-matched basis functions that are selected from a dictionary of Gabor atoms. Zou *et al.*[8] extended some earlier results on steady-motion based Dopplerlet transform and introduced the application of Dopplerlet transform to the estimation of range and speed of a moving source.

Not many works have been reported on the problem of solving the resulting wave equation for a moving source in an acoustic waveguide. Guthrie *et al.*,[9] Hawken,[10] and more recently Lim and Ozard[11] considered sources moving radially or horizontally and obtained expressions for the acoustics field using normal theory. Flanagan *et al.*[12] and Clark *et al.*[13] formulated the moving source problem in terms of ray theory, where each raypath has different Doppler shift according to its angle of emission. Most solutions are given in

[a]Author to whom correspondence should be addressed. Electronic mail: nicolas.josso@gipsa-lab.inpg.fr

terms of contemporary time, i.e., the time at which the sound reaches the receiver but Lim and Ozard expressed their solution in terms of retarded time, i.e., the time at which the sound was transmitted by the source.

This paper investigates the effects of the motion that often exists in an operating transmitter-channel-receiver configuration in geoacoustic inversion. Doppler effect consequences on the estimation of the IR are shown and explained for shallow water environments with matched filtering between the transmitted and received signals. The studied signals have very low central frequencies (around 1300 Hz) and high bandwidth (around 2000 Hz). For feasibility purposes, it is considered that the relative motion existing between the transmitter and the receiver is horizontal with a constant speed during transmissions.

The performance of the correlation receiver in delay and Doppler can be described with the ambiguity function. If the signal is narrowband, then the conventional formulation of the ambiguity function is appropriate. In this case, the effects of motion, which are a compression in time for approaching sources and an expansion for receding sources, are approximated as simple carrier-frequency shifts of the transmitted waveform. For the narrowband case, the correlation receiver has a reference set of signals composed by time-delayed and carrier-frequency-shifted versions of the transmitted signal. However, this is no longer valid when the ratio bandwidth under central frequency increases. In our study, a multipath wideband ambiguity function is introduced in order to account for a different broadband Doppler effect for each path. In the wideband case, the correlation receiver has reference signals that are time-delayed and time-scaled versions of the transmitted one. Hermand and Roderick[14] fully described and formulated the narrowband and wideband ambiguity functions for active sonar systems. The interferences that can occur in the cases of reflection on multiple moving targets are also analyzed. The purpose of this paper is to analyze the broadband ambiguity function with the same approach as in Ref. 14 and apply it to the estimation of the IR of an underwater acoustic propagation channel when motion exists between the source and the receiver. We will show that when high bandwidth and very low central frequency signals are transmitted, the wideband ambiguity plane enables estimating and compensating the Doppler effects which modify the underwater acoustic propagation channel. Doppler effects will be shown to differ for each propagation path on simulated and real data.

This paper is organized as follows. Section II presents a transient signal modeling for a multipath environment with rectilinear, constant speed motion. Then Sec. III describes the motion effects both on the estimation of an IR computed with a correlation receiver process and on the narrowband and wideband ambiguity functions in a multipath environment. A Doppler effect estimation and removal technique on the IR and its applications on simulated data are presented in Sec. IV. The results on a real data set are presented in Sec. V. We close in Sec. VI with conclusions.

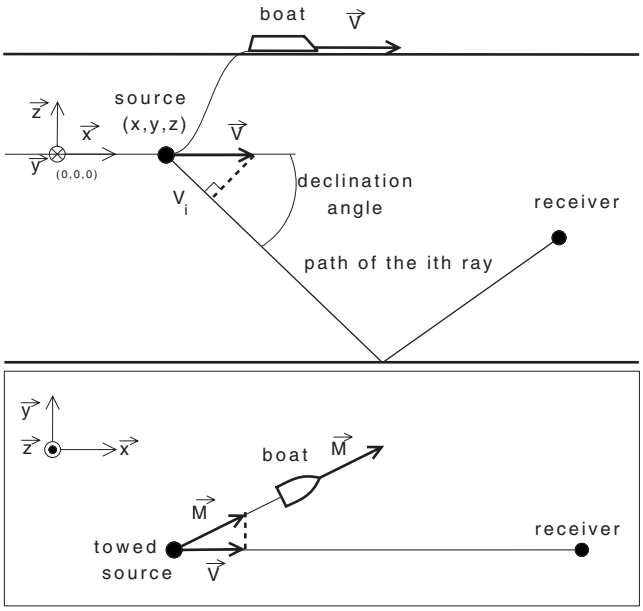


FIG. 1. Schematic illustration of how v_i is computed from the motion vectors $\vec{M}$ and $\vec{v}$. The top panel represents a side view of the source-receiver configuration while the bottom panel is a top view. For simplicity, it is assumed here that the speed is constant.

II. MODELING WAVE PROPAGATION AND MOVING TRANSIENT EMISSION

The wideband Doppler effect in a multipath environment is presented in this section using contemporary time, i.e., the time at which the sound reaches the receiver, and retarded time, i.e., the time at which the sound was transmitted by the source considering a constant speed of propagation in the medium.

A. Signal received for one ray

In this section, we consider a fixed receiver and a source with constant speed motion $\vec{M}$. The source emits a signal for T s while moving at a constant speed v along the x axis, as illustrated in Fig. 1. For one emission, the source position along the x axis is

$$x_0 - \frac{vT}{2} \leq x \leq x_0 + \frac{vT}{2}, \tag{1}$$

where x_0 is the position of the source after $T/2$ s. We define a time axis u for the emission, which is the retarded time, and t for the reception, which is the contemporary time, following a relation of the form

$$u + T_i(u) = t, \tag{2}$$

where $T_i(u)$ refers to the time delay of the ith ray and u verifies

$$-\frac{T}{2} \leq u \leq \frac{T}{2}. \tag{3}$$

Relation (2) means that a signal transmitted at the delayed time u is received on the ith ray at the contemporary time t equal to u plus the propagation time along the ith ray. If $L_i(u)$

is defined as the path length of the ith ray in meters, formula (2) can be rewritten as

$$u + \frac{L_i(u)}{c} = t, \tag{4}$$

where c is the speed of the sound in the medium. Assuming that the latter is constant, the path length of the ith ray is defined as

$$L_i(u) = \sqrt{x_i(u)^2 + z_i^2}, \tag{5}$$

where $(x_i(u), z_i)$ represents the position of the virtual source from which the signal propagating along the ith ray seems to have been radiated (the expressions of z_i are given in Appendix A for all rays). It is worthy noting that each of the virtual sources seems to move at a different apparent speed v_i. The term $x_i(u)$ is the distance existing between the virtual source and the hydrophone along the x axis

$$x_i(u) = x_0 - v_i u, \tag{6}$$

where v_i is considered positive for approaching sources and negative for receding sources. Expression (2) can be rewritten as

$$u + \frac{\sqrt{(x_0 - v_i u)^2 + z_i^2}}{c} = t. \tag{7}$$

In order to obtain the time of emission as a function of the time of reception, some hypotheses are necessary. The first one considers that the depth of the propagation channel can be neglected compared with its length. After calculation (given in Appendix B) and under our first hypothesis, expression (7) can be expressed by

$$u + \frac{x_0 - v_i u}{c} + \frac{z_i^2}{2c(x_0 - v_i u)} = t. \tag{8}$$

The second hypothesis states that the distance covered by the moving source during one transmission can be neglected compared with the source-hydrophone separation. Some calculation (given in Appendix B) yields an approximation of the expression of the time of emission as a function of the time of reception:

$$u = \frac{t - \tau_i}{1 - v_i\left(\frac{1}{c} - \frac{z_i}{2cx_0^2}\right)}, \tag{9}$$

where τ_i corresponds to the time-delay associated with the ith path for a fixed source located at $x=x_0$,

$$\tau_i = \frac{x_0}{c} + \frac{z_i^2}{2cx_0}. \tag{10}$$

Relation (9) means that the signal received for the ith ray is a time-delayed and time compressed (or expanded) version of the transmitted one. Expression (10) illustrates what appears logical: if one computes an IR with a moving source and wants to compare it with the motionless case, it should be done with a source located in the middle of the motion.

B. Signal received in multipath configurations

In Eq. (6), it has been assumed that the speed of the source appears to be different for each ray. The projection of the source's speed on the sight line between the transmitter and the receiver is called $\vec{v}$. It is assumed that $\vec{v}$ can vary with time as the line defined by the source and the receiver changes. As shown in Fig. 1, the motion vector $\vec{v}$ is then projected on the path of the ith ray with the declination angle θ_i which leads to

$$v_i = \|\vec{v}\|\cos(\theta_i). \tag{11}$$

A solution of the wave equation for the sound field in an iso-speed ocean channel overlying a homogeneous fluid half-space was developed and published over half a century ago in a classic paper by Pekeris.[15] For Pekeris waveguides[16] the departure angle of one path, θ_i, and its angle of arrival differ in sign for one ray out of two but they have the same cosine. This basic property can be used to validate the assumption of using a Pekeris waveguide. We consider that the received signal is distorted by the combined effects of propagation and source motion (i.e., time-delayed, amplitude attenuated, and Doppler transformed). Using Eq. (9), and adding a change in amplitude to conserve energy yields the expression of the signal received at time t for the ith ray

$$s_i(t) = a_i \eta_i^{1/2} e((t - \tau_i) \cdot \eta_i), \tag{12}$$

where $e(t)$ is the transmitted signal, a_i represents the amplitude attenuation due to propagation losses, and η_i is the scale factor due to the broadband Doppler effect satisfying

$$\eta_i = \frac{1}{1 - \|\vec{v}\|\cos(\theta_i)\left(\frac{1}{c} - \frac{z_i}{2c \cdot x_0^2}\right)}. \tag{13}$$

The received signal $s(t)$ is the sum of all the $s_i(t)$ received from each ray which leads to the following expression:

$$s(t) = \sum_i s_i(t). \tag{14}$$

We made the hypothesis that the distance covered by the source during the transmission can be neglected compared with the source-receiver separation to obtain Eq. (8). From now on, we can consider the propagation time, $T_i(u)$, and the projection of the motion vector $\vec{v}$ along the path of the ith ray, v_i, to be constant during one emission. By using the complete formulation of the signal received for each ray (12) and of the compression factor η_i (13), Eq. (14) can be rewritten as

$$s(t) = \sum_i a_i \eta_i^{1/2} e((t - \tau_i)\eta_i). \tag{15}$$

This expression illustrates that the signal received from a moving source with a multipath propagation is a weighted sum of amplitude attenuated, time-delayed, and Doppler-transformed versions of the transmitted signal. The compression factor η_i depends on the velocity of the source $\vec{v}$, on the angle of emission of ray i, and on the position of its corresponding virtual source, as shown in Eq. (13).

III. EFFECTS OF SOURCE MOTION

In Sec. II, the received multipath signal has been characterized. The effects of source motion on the estimation of an IR with matched filtering are formulated and analyzed in this section.

A. Effects of source motion on the matched-filter output

Considering that the received signal is defined as the sum of amplitude-attenuated, time-delayed, and Doppler-transformed versions of the transmitted signal, the Doppler-transformation stands for a Doppler scaling which is not approximated as a simple frequency shifting. Assuming that the transmitted signal is known, the propagation time and the velocity associated with each ray can be estimated by cross correlating the received signal with a set of reference signals. The set of reference signals is composed of time-delayed and Doppler-transformed versions of the emitted signal for the range of time delays and speeds expected.[14] For each reference signal, the cross correlation depends on the speed v because of the η dependency and is computed by

$$R(\tau,v) = \int_{-\infty}^{\infty} s(t+\tau)\,\eta^{1/2} e^{T}(\eta t)dt, \tag{16}$$

where ^{T} denotes the complex conjugation, $s(t)$ is the received signal, $e(t)$ is the transmission, and η is the compression factor due to the Doppler effect. Using the expression of the received signal in a multipath environment (15) in Eq. (16) yields

$$R(\tau,v) = \sum_i a_i(\eta\eta_i)^{1/2} \int_{-\infty}^{\infty} e(\eta_i(t+\tau-\tau_i))e^{T}(\eta t)dt. \tag{17}$$

Local maxima of this correlation function are reached for each ray. For the ith ray, the maximum is reached when the reference and the propagated signal are exactly aligned in time delay and Doppler. It is assumed that the smallest time difference between two consecutive arrivals is larger than the inverse of the time-bandwidth product of the transmitted signal so rays are well separated for the motionless case, and each peak of the correlation can be detected. The interferences that could occur between local maxima are studied in Ref. 14 and are not the purpose of this paper. The time of propagation and the apparent speed of the ith ray can be estimated once a local maximum is detected.

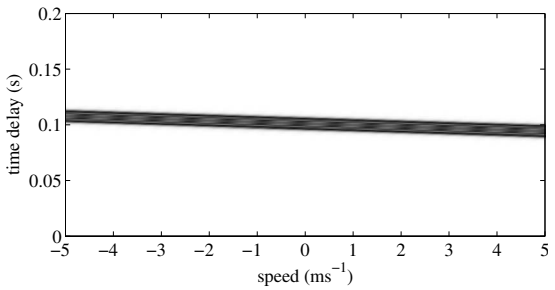


FIG. 2. Representation of a LFM in the narrowband ambiguity plane without multipath with a central frequency of 1300 Hz, a bandwidth of 2000 Hz, and a time duration of 4 s. The simulated speed is 2.5 ms^{-1} and the time delay is 0.1 s.

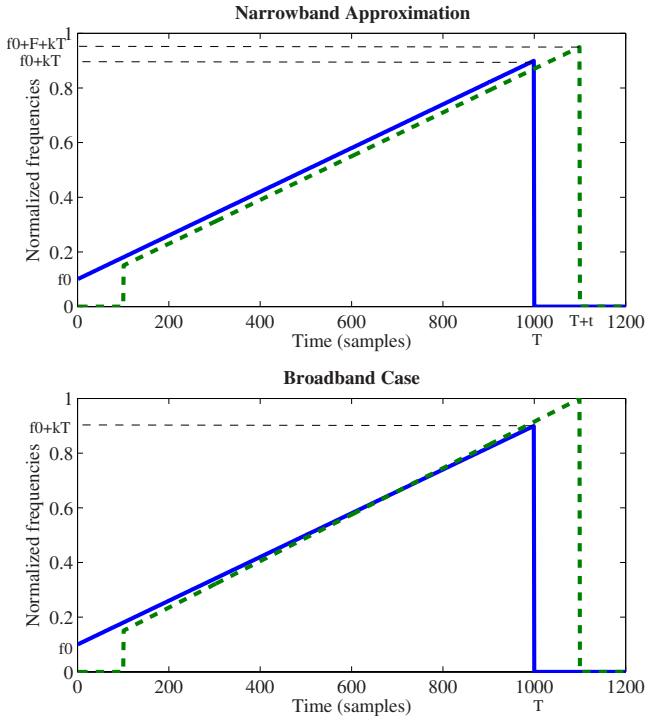


FIG. 3. (Color online) Time frequency ideal representation illustrating the mismatch existing between the transmitted LFM signal (solid line) and the Doppler-transformed signal (dashed line) for the narrowband approximation and the broadband case. $f0$ is the beginning frequency of the LFM.

B. Effects of source motion on the ambiguity plane

The wideband ambiguity plane introduced here is the squared magnitude of the result of the correlation equation (17). The propagation time and the velocity associated with the received signal are estimated by cross correlating the received signal with a set of reference signals. The set of reference signals is composed of time-delayed and Doppler-transformed versions of the transmitted signal which is assumed to be known.

For geoacoustic inversion applications, the emitted signals are wideband signals, and the motion effect cannot be approximated by a frequency shifting. It is well known that the representation of a linear frequency modulation (LFM) signal in the narrowband ambiguity plane is ambiguous, as shown in Fig. 2. A LFM frequency shifted by the narrowband approximation of the Doppler effect is really close to one another which is just time-delayed, as illustrated in Fig. 3. Figure 3 shows the ideal time frequency representation of a LFM signal compared with the Doppler-transformed versions of the signal under narrowband and broadband approximations of the Doppler effect. That is why there is an ambiguity and one cannot find accurately the right coordinate of the maximum in the ambiguity plane. There is no absolute maximum in the narrowband ambiguity plane, and possible solutions are represented by all the points of the

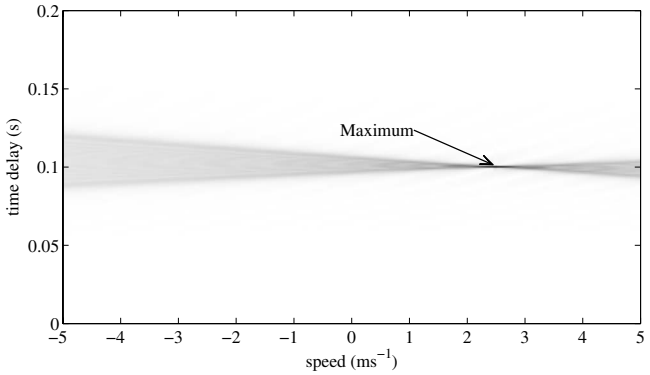


FIG. 4. Representation in the wideband ambiguity plane of a LFM without multipath with a central frequency of 1300 Hz, a bandwidth of 2000 Hz, and a time duration of 4 s. The simulated speed is 2.5 ms^{-1} and the time delay is 0.1 s.

straight line containing the energy. Hence neither the propagation time nor the source speed can be estimated accurately when a broadband LFM is transmitted.

As shown in Fig. 4, the representation of a time-delayed and Doppler-transformed LFM provides a finite resolution in the (τ,v) domain but remains ambiguous in the wideband ambiguity plane, although there is one absolute maximum that can be detected far more accurately than in the narrowband ambiguity function. That is why the wideband ambiguity plane is well adapted to geoacoustic inversion applications. The maximum is reached when the parameters of the reference signal match exactly with the parameters of the estimate. The amplitude of the correlation stays high for time delays close to the simulated one, and the correlation broadens farther from the simulated speed.

For multipath propagation, paths have different apparent speeds and different time delays, as illustrated in Fig. 5. Each path is seen as a sweep-like shape which broadens with the distance between the reference and the simulated speed. Figure 5 illustrates the wideband ambiguity plane for a simulated multipath propagation centered on the six first arrivals.

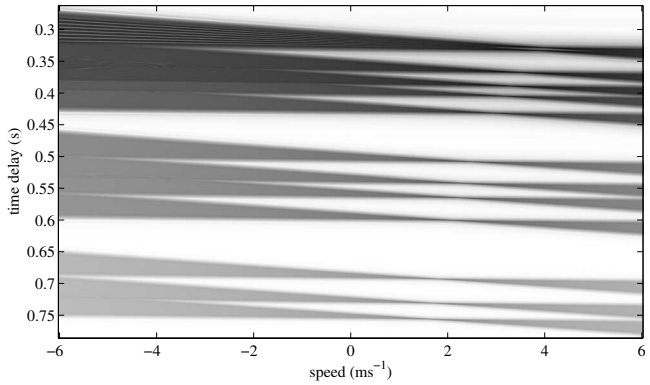


FIG. 5. Ambiguity plane of a simulated multipath propagation. The relative speed simulated is almost 8 knots and the source-receiver separation is 500 m. The signal transmitted is a LFM with a central frequency of 1300 Hz, a bandwidth of 2000 Hz, and a time duration of 4 s. The propagation channel is 165 m deep with a constant sound speed of 1500 ms^{-1}. The bottom is a half space with a sound speed of 1800 ms^{-1} and a density of 1800 kg m^{-3}.

C. The LFM case

LFMs are classically used for geoacoustic inversions and tomography because their large time-bandwidth product provides a good resolution and because of electro-acoustic transduction technological constraints. From now on, the transmitted signal is assumed to be a LFM signal with known parameters

$$e(t) = \frac{\mathrm{rect}\left(\frac{t}{T}\right)}{\sqrt{(T)}} \exp\left(j2\pi\left(f_c t + \frac{k}{2}t^2\right)\right), \tag{18}$$

where f_c is the central frequency of the sweep, k is the chirp rate or sweep rate, T is the duration of the signal, and the rect function is defined by

$$\mathrm{rect}(t) = \begin{cases} 1 & \text{if } |t| \le \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Some algebraic manipulations[14,17,18] detailed in Appendix C with Eqs. (17) and (18) lead to the analytic expression of the cross correlation defined previously if v is different from v_i,

$$R(\tau,v) = \sum_i \frac{C_i D_i E_i}{2\sqrt{|\beta_i|}} \int_{X_i}^{Y_i} \exp\left(\gamma\frac{\pi}{2}jt^2\right)dt, \tag{20a}$$

$$\text{where} \quad C_i = \frac{\eta_i^{1/2}\eta^{1/2}a_i}{T}\exp(j\pi(f_c\Delta\tau_i(\eta+\eta_i))), \tag{20b}$$

$$D_i = \exp\left(\pi\frac{k\Delta\tau_i^2}{4}j(\eta_i^2-\eta^2)\right), \tag{20c}$$

$$E_i = \exp\left(-\gamma 2j\pi\left(\frac{\alpha_i}{2\sqrt{|\beta_i|}}\right)^2\right), \tag{20d}$$

$$\beta_i = \frac{k}{2}(\eta_i^2-\eta^2), \tag{20e}$$

$$\alpha_i = f_c(\eta_i-\eta) + k\Delta\tau_i(\eta^2-\eta_i^2), \tag{20f}$$

$$\Delta\tau_i = \tau - \tau_i, \tag{20g}$$

$$X_i = \frac{\gamma\alpha_i}{\sqrt{|\beta_i|}} + 2t_1\sqrt{|\beta_i|}, \tag{20h}$$

$$Y_i = \frac{\gamma\alpha_i}{\sqrt{|\beta_i|}} + 2t_2\sqrt{|\beta_i|}, \tag{20i}$$

$$\gamma = \mathrm{sgn}(k(\eta_i-\eta)). \tag{20j}$$

The bounds of integration of Eq. (20) depend on t_1 and t_2 which are given in Appendix D. Finally, the result of Eq. (20) can be expressed and simplified with a complex form of the Fresnel integrals if v is different from v_i,

$$R(\tau,v) = \sum_i \frac{C_i D_i E_i}{2\sqrt{|\beta_i|}}(F(Y_i)-F(X_i)), \tag{21}$$

$$F(u) = C(u) + j\gamma S(u), \tag{22}$$

$$C(u) = \int_0^u \cos\left(\frac{ut^2}{2}\right)dt, \tag{23}$$

$$S(u) = \int_0^u \sin\left(\frac{\pi t^2}{2}\right)dt. \tag{24}$$

When the Doppler transformation of the reference signal matches exactly the Doppler transformation of the *i*th path, expression (21) is no longer valid, and the *i*th term of the sum, $r_i(\tau, v_i)$, becomes

$$r_i(\tau, v_i) = C_i\left(|\Delta\tau_i| - \frac{T}{\eta_i}\right)\frac{\sin(\xi_i)}{\xi_i}, \tag{25a}$$

$$\xi_i = \pi k \Delta\tau_i(\eta_i|\Delta\tau_i| - T). \tag{25b}$$

When the reference signal matches exactly the *i*th path, expression (25) reaches its maximum as expected and equals a_i meaning that the amplitude associated with each ray can be recovered. It is worthy noting that Eq. (25) is a sine cardinal multiplied by a constant that can be compared with the classical LFM ambiguity function. According to the asymptotic evaluation of Harris and Kramer[19] and Kramer,[17] the Doppler tolerance, i.e., half-power contour, is given by

$$V_{-3\ \mathrm{dB}} = \pm\frac{2610}{TW}\ \text{knots}, \tag{26}$$

where T is the duration of the LFM and W the signal bandwidth. As an example, we consider a large TW-product and wideband LFM signal with known parameters:

$$f_c = 1300\ \text{Hz}, \quad W = 2000\ \text{Hz}, \quad T = 4\ \text{s}, \tag{27}$$

where f_c is the central frequency of the LFM. From Eq. (26) the Doppler tolerance of this signal in the wideband ambiguity plane is $V_{-3\ \mathrm{dB}} = \pm 0.32$ knots, whereas its classical Doppler tolerance in the narrowband ambiguity plane would be[7]

$$V_{-3\ \mathrm{dB}} = \pm\frac{450W}{f_c} = \pm 692.3\ \text{knots}. \tag{28}$$

This example confirms the results obtained with our wave propagation modeling, the wideband ambiguity function is well adapted for the study of Doppler scenarios in geoacoustic inversion applications, and the narrowband approximation is not valid. For the narrowband case, the correlation receiver has a reference set of signals composed of time-delayed and carrier-frequency-shifted versions of the transmitted signal. As can be seen in Fig. 3, a LFM frequency-shifted by the effects of motion is not very different from one another which is only time-delayed; that is why LFM signals have a poor Doppler tolerance and are ambiguous in the narrowband ambiguity plane. The inclusion of the motion effect for broadband signals, i.e., time compression (or expansion), in the computation of the correlation receiver clearly enhances the Doppler tolerance of wideband signals so that a LFM is no longer ambiguous in the wideband ambiguity plane. Different Doppler removal techniques derived from the wideband ambiguity plane are introduced and studied in Sec. IV.

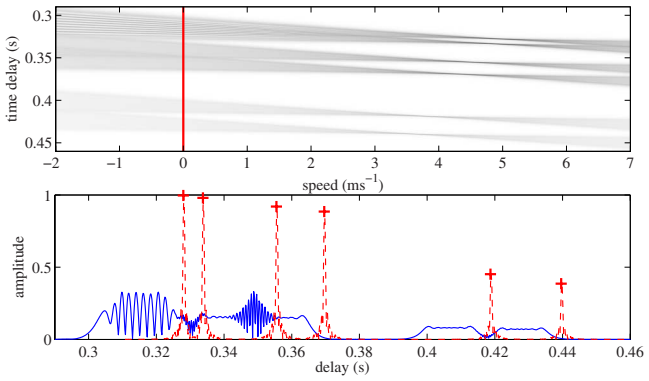


FIG. 6. (Color online) Estimation of the IR of a propagation channel computed by keeping a column at zero speed in the ambiguity plane represented by solid lines. The dotted line represents the IR estimated with a motionless source. The relative speed simulated is almost 10 knots and the source-receiver separation is 500 m.

IV. DOPPLER REMOVAL TECHNIQUES

A computational method has been presented in Sec. III to study the multipath propagation when a LFM is transmitted. The wideband ambiguity plane was introduced as an adapted representation of the multipath propagation for geoacoustic applications whenever the source is moving or not.

A. The motionless source hypothesis

A well known method of estimating the IR of a propagation channel is to compute the cross correlation between the time-delayed transmitted signal and the received signal as

$$\mathrm{IR}(\tau) = \sum_i a_i\eta_i\int_{-\infty}^{\infty} e(\eta_i(t+\tau-\tau_i))e^T(t)dt. \tag{29}$$

This is equivalent to computing expression (17) with a zero speed, meaning $\eta = 1$, for all references. The result of Eq. (29) is a sub-part of the broadband ambiguity plane and can be obtained by keeping the column at zero speed in this plane. From Eq. (29) it can be seen that the motion effects are not considered and the computed IR will be biased. Both the estimation of the time of propagation (τ_i) and the amplitude of the *i*th ray (a_i) will be incorrect. An example of the zero speed correlation is illustrated with the solid line in Fig. 6. The parameters used for this simulation are the same as the one used for the simulation presented in Fig. 5 except for the source speed. The first two paths are not resolved because the transmitted signal is transformed by the Doppler effect which is not taken into account during the processing. The amplitude of each peak is lowered and the time delays are not correctly estimated. The exact values of amplitude and time delay of the *i*th path are located at the maximum value of the shape associated with the *i*th path on the wideband ambiguity plane. The effects of Doppler transformations on wideband LFM signals have been verified for a large set of simulated data, confirming the necessity of adding a

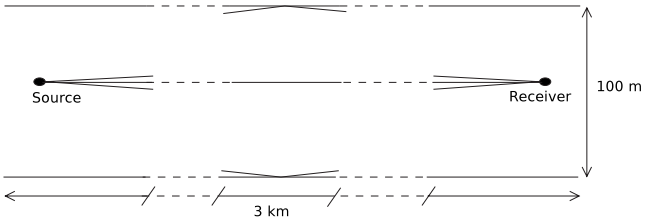


FIG. 7. Schematic illustration of the hypothesis necessary to obtain relation (30). Here the source is 30 times farther than the channel depth and only the first three paths are represented for simplicity.

speed parameter to the matched filtering processing so the motion of the source can be estimated and then compensated.

B. The uniform speed compensation

The cross correlation between the time-delayed transmitted signal and the received signal has been shown to have poor performances in estimating the IR when the motion is not taken into account. We propose here a new method to compensate the motion in the estimation of an IR of the propagation channel. This aims at reconstructing a motionless IR from an acoustic observation with a moving source in order to enable the use of classical acoustic inversion processes. The declination angle of the direct path is very low, and expression (11) shows that the apparent speed of this path will be the projection of the motion vector along the receiver-source line. If the amplitude of the direct path is considered higher than any other, then the amplitude of its correlation with the reference signals is also higher than any other. The speed of the source v is estimated as the coordinates of the absolute maximum in the ambiguity plane. If the source is far away from the receiver, as illustrated in Fig. 7, the received paths will have low declination angles, meaning an apparent speed close to v and Eq. (11) becomes

$$v_i = \|\vec{v}\|. \tag{30}$$

The uniform speed compensation is defined as keeping the column at constant speed v in the wideband ambiguity plane and is presented in Fig. 8.

We developed a software simulating all the propagation process with ray theory for signals transmitted from a moving source in order to test our uniform motion compensation method. The simulation is presented in Fig. 8 and was computed with a source moving at a constant speed of 5 ms^{-1}, at a depth of 24 m, 4 km from the receiver at a depth of 90 m on a 165 m deep channel with a constant sound speed of 1500 ms^{-1}. The signal transmitted is a LFM with a central frequency of 1300 Hz, a bandwidth of 2000 Hz, and a duration of 4 s. In Fig. 8, the star shows the absolute maximum detected in the wideband ambiguity plane, giving an estimated speed of 5 ± 0.16 ms^{-1} as expected. In this case, the source is far away from the receiver at a distance much greater than the depth of the propagation channel so the conditions necessary to apply uniform speed compensation are met. The compensation of the motion on the estimated IR is made by keeping the column at the estimated speed which is represented with solid line on the top of Fig. 8. The panel on the bottom of Fig. 8 represents estimates of the IR where the

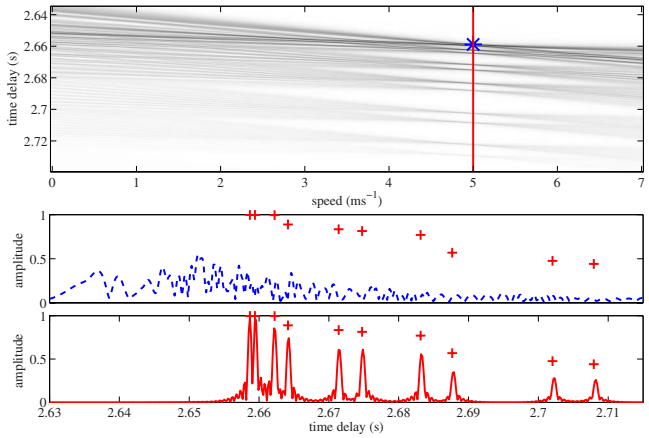


FIG. 8. (Color online) The top panel shows the ambiguity plane for a 4000 m multipath propagation. The star represents the detected absolute maximum. The bottom panel shows the IR estimated with zero speed compensation in dashed lines (on the top); the IR estimated with uniform speed compensation is represented with solid line (on the bottom) and the crosses stand for the ideal simulated IR.

dashed line stands for the IR estimated with zero speed compensation (no speed compensation). The solid line stands for the estimation of the motion compensated IR while the crosses represent the simulated IR. This compensation method is effective and improves the estimated IR. Both amplitudes, time delays, and peak detections have been improved. It is not possible to distinguish any path without compensation while they are clearly detectable with compensation. The general shape of the IR is well recovered even if amplitudes are not the theoretical ones.

The estimated amplitudes of the motion compensated IR are biased because the hypothesis made in relation (30) is not valid for all rays, and the column at constant speed v does not cut each chirp-like shape around its maximum in the ambiguity plane. The interferences existing between paths are clearly illustrated on the broadband ambiguity plane of Fig. 8. As expected, the Doppler effect lowers amplitudes, shifts time delays, and leads to the appearance of interferences between peaks which are not detectable if the motion is not compensated. The time delay shifting is clear on the estimated IR represented in Fig. 8 and can be explained by the sweep-like shape of each path on the ambiguity plane.

Figure 9 illustrates another example of uniform speed compensation where the source is moving at a speed of 4 ms^{-1} at a depth of 32 m and a distance of 500 m from the receiver. The recorder is at a depth of 90 m and the propagation channel is the same as before. The star represented on the wideband ambiguity plane of Fig. 9 shows the absolute maximum detected in the wideband ambiguity, giving an estimated speed of 3.99 ± 0.16 ms^{-1} which is valid. For this simulation, the depth of the channel cannot be neglected compared with the source-hydrophone separation so the conditions required to apply the uniform speed compensation are not met. We consider that the uniform speed compensation can be applied when the source-hydrophone separation is at least ten times larger than the channel depth. The motion compensation for the IR estimate is made by keeping the column at the estimated speed which is represented with

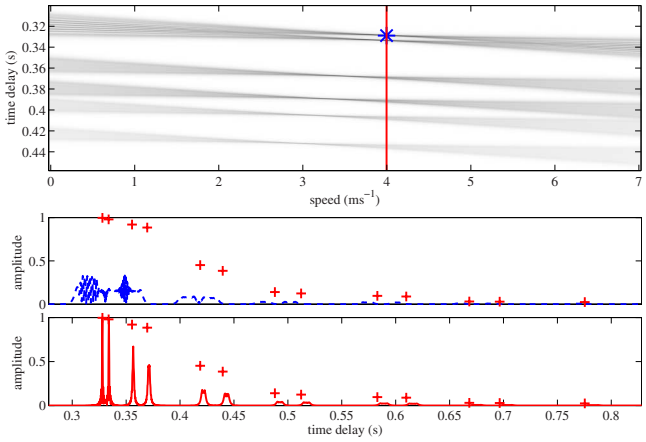


FIG. 9. (Color online) The top panel shows the ambiguity plane for 500 m multipath propagation. The star represents the detected absolute maximum. The panel on the bottom shows the IR estimated with zero speed compensation in dashed lines (on the top); the IR estimated with uniform speed compensation is represented with solid line (on the bottom) and the crosses stand for the ideal simulated IR.

solid line in Fig. 9. Contrary to the previous simulation, it can be seen that the line along which the compensation is made on the ambiguity plane is no longer cutting each sweep-like shape around its maximum. This means that the apparent speed, amplitude, and time delay of each ray will not be estimated properly. In Fig. 9, it is worthy noting that the distance between local maxima and the line along which the compensation is made increases with the time delay, which explains the degradation of estimations for the last rays. These drawbacks can also be seen on the estimation of the motion compensated IR which is represented by a solid line in the bottom of Fig. 9 and can be compared with the dashed line representing the zero speed compensation and the crosses representing the simulated IR. The first two rays of the motion compensated IR are well estimated and their time-delays and amplitudes are well recovered. However, the estimation error of the motion compensated IR increases with the time-delays.

The uniform speed compensation is a good way to compensate the motion of a source while estimating an IR. Though the hypothesis is necessary, its validity cannot always be verified. A compensation method that considers the speed of each ray individually should lead to a more general result and is one of our short term prospects.[20]

V. APPLICATION TO REAL DATA

In this section, the previously described methods are tested on a dedicated real data set recorded in a shallow water environment.

A. Experiment description

The BASE'07 experiment was jointly conducted by the NATO Undersea Research Center, the Forschungsanstalt der Bundeswehr für Wasserschall und Geophysik, the Applied Research Laboratory, and the Service Hydrographique et Ocanographique de la Marine (SHOM). The main objective of the experiment is to investigate broadband adaptive sonar techniques in the shallow water. Two additional days of measurement were also conducted by SHOM for geoacoustic inversion. Some results can be seen in Refs. 21–23. The first day was dedicated to a general survey and the second day was dedicated to motion effect compensation. The real data of this second day are used in the following. The campaign took place on the Malta Plateau in shallow water (130 m depth). Underwater LFM, as illustrated in Fig. 10, was transmitted by a source moving rectilinearly at constant speed from 2 to 12 knots and different depths. The transmitted LFM has a bandwidth of 2000 Hz, a central frequency of 1300 Hz, and a duration of 4 s, and the effects of the multipath propagation can be seen in Fig. 10. The transmitted signal's bandwidth is very large compared with the signal's central frequency, which ensures a large propagation distance and good auto-correlation properties for geoacoustic inversion. However, the transmitted signals are Doppler sensitive. The source-receiver separation varied from 500 to 25 000 m, and the transmitted signals were recorded by an array of six hydrophones located at different depths (from 9 to 94 m). As shown in Fig. 11, the array of hydrophones had its own global positioning system (GPS) and clock for localization and was not anchored so it can move freely with currents and avoid additional flow noise. The boat had a GPS which was used to derive the position of the towed source. Both position and speed of the source and the hydrophone array are known at any moment so the results can be compared and analyzed.

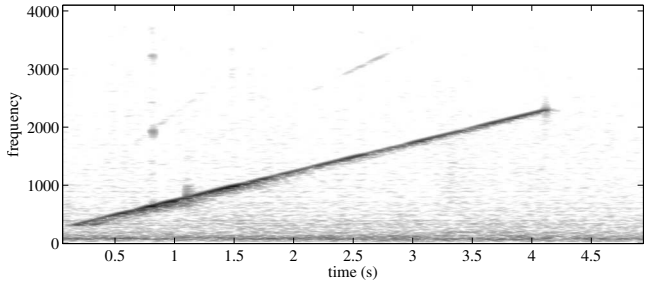


FIG. 10. Spectrogram of the LFM signal transmitted by the towed source. The effects of the multipath underwater propagation such as the apparition of time-delayed echoes can be seen on this time frequency representation.

B. Results

The wideband ambiguity plane has been studied for more than 100 different scenarios on each of the six hydrophones of the array. The motion existing between the source and the receivers was clearly seen on all the wideband ambiguity planes. The uniform speed compensation method was then automatically carried out to estimate the source-receiver relative speed and the motionless IR. All the broadband ambiguity planes analyzed from real data give realistic results. They are close to the simulated data and the chirp-like shapes, representing that each path of the acoustic waves is easily seen. The apparent speed of the first path is estimated in the ambiguity plane by the detection of the absolute maximum. It corresponds to the projection of the speed vector on the sight line existing between the source and the hydrophones multiplied by the cosine of the declination angle. The accuracy of the speed estimate using Eq. (26) is 0.16 ms^{-1} or 0.32 knots. An example of the wideband ambiguity plane

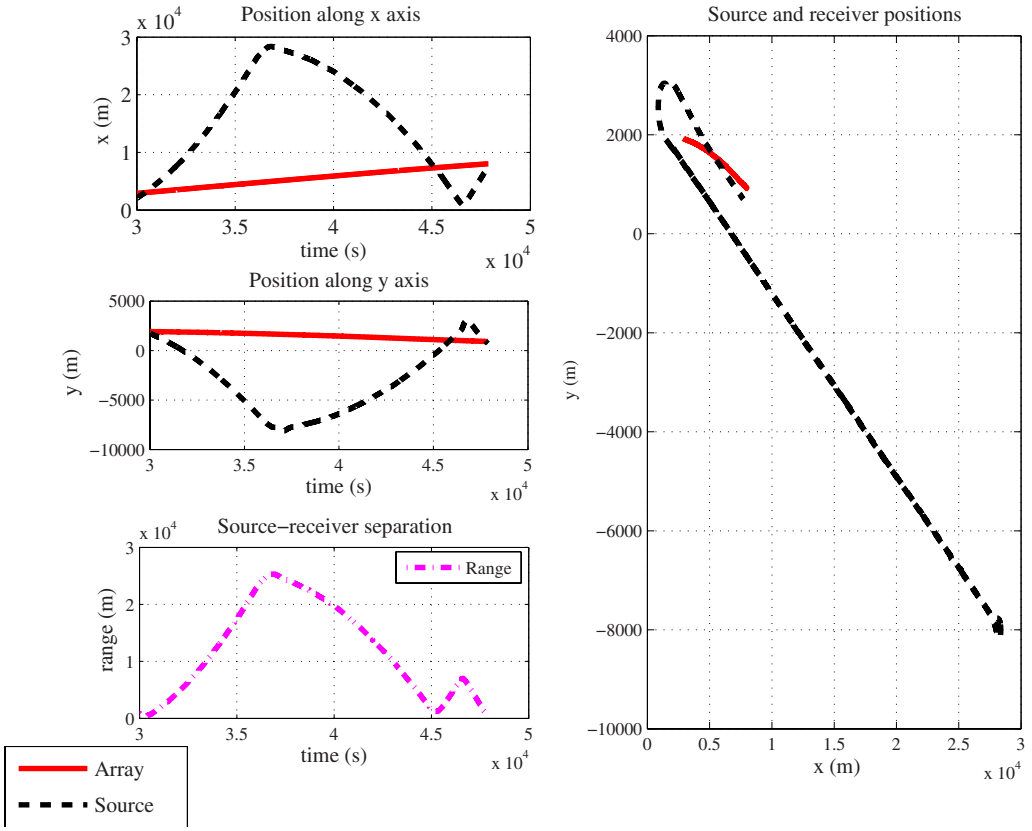


FIG. 11. (Color online) Positions of the towed source and of the hydrophones along time and evolution of source-receiver separation along time.

with its associated motion-compensated IR is presented in Fig. 12. The scenario presented here was recorded with a projected speed of 11 knots, a source at 24 m depth, and a 4015 m propagation channel. The speed estimated by the automatic process is 10.8 knots which stays within the bounds of accuracy. The ambiguity plane illustrated on the top of Fig. 12 is close to that obtained during simulations. The IR estimated with the uniform speed compensation method is shown as a solid line on the bottom panel of Fig. 12. The amplitude of each peak seems to be corrected, and the time delays are shifted compared with those of the IR estimated with classical matched filtering represented with a dashed line. It is worthy noting that the source perpetually moves so the propagation channel is different for each emission. This shows that it is not possible to improve the IR by computing means as is usually done with real data and motionless sources. A simulated IR obtained from a Pekeris waveguide with a flat bottom of sandy mud having a sound speed of 1550 ms^{-1} and a density of 1700 $kg\,m^{-3}$, which is close to the data recorded *in situ*, is represented with crosses on the bottom panel of Fig. 12. The propagation speed of the simulated canal is 1500 ms^{-1} and its depth is 125 m. Even if the simulated IR is not the real one, this provides a reference to compare the IR estimated with classical matched filtering and the motion-compensated IR. It can be seen that the uniform speed compensation both corrects the general shape of the IR and shifts the time delays which was the case for simulations. Finally the motion compensated IR is closer to the expected IR than the IR computed with zero speed com-

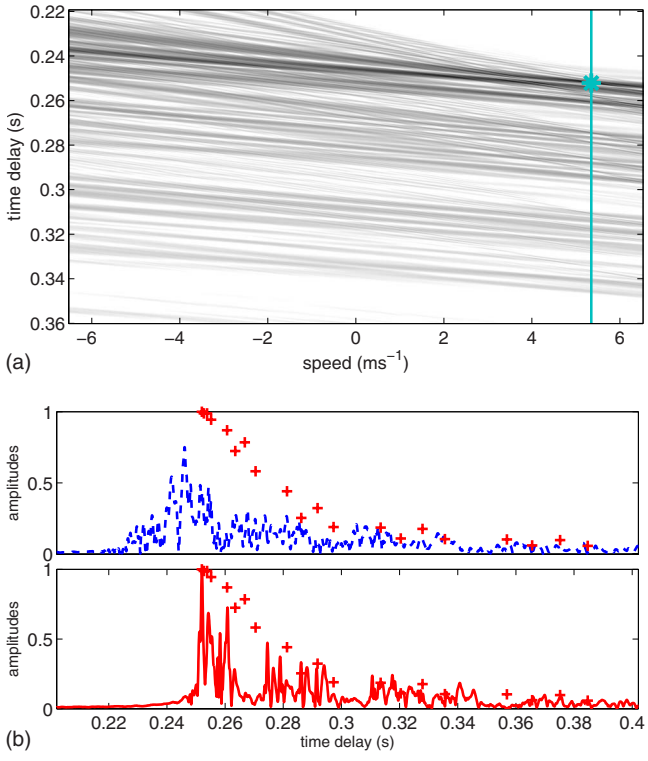


FIG. 12. (Color online) The top panel shows the wideband ambiguity plane for real data from BASE'07 campaign. The two lower panels represent the IR estimated with classical matched filtering (dashed line) and the motion-compensated IR (solid line). The crosses represent the IR simulated with a Pekeris waveguide having parameters close to the data recorded *in situ*.

TABLE I. The mean and standard deviation (std) of the difference between the estimated speed and the projection of the real speed normalized by the projection of the real speed expressed in m/s. 70 transmissions have been studied with a range source-receiver varying from 1500 to 6000 m and a source's speed varying from −4 to 6 ms^{-1}.

	H1	H2	H3	H4	H5	H6
Mean	0.11	0.09	0.15	0.20	0.03	0.12
std	0.36	0.33	0.47	0.45	0.21	0.34

pensation, showing the improvements of our compensation method on real data.

Most of the estimated speeds were correct and quite accurate, leading to good compensation of the motion, but some were poor. Results obtained on the estimated speeds from 70 transmissions are presented and analyzed for each hydrophone in Table I. In most cases, poor speed estimates occurred when the absolute maximum of the ambiguity plane corresponded to constructive interference between two paths arriving almost simultaneously or because the distance between the source and the receiver exceeds 15 000 m. Finally, results are consistent from one hydrophone to another even if the signal to noise ratio (SNR) varies.

As illustrated in Fig. 13, the motion-compensated IR obtained from several hydrophones at different depths has been analyzed and compared with classical matched filtering IR. The IRs estimated from the data recorded on hydrophone number one (H1) match well with the physics of the propagation channel. This hydrophone is located close to the sea surface in the mixed layer where the sound speed is almost constant. The upper left of Fig. 13 clearly shows a family of rays arriving first with low amplitude which were trapped close to the sea level and were reflected at the sea surface. For this application to real data, the motion compensation improves the shape of the IR and allows recovery of the

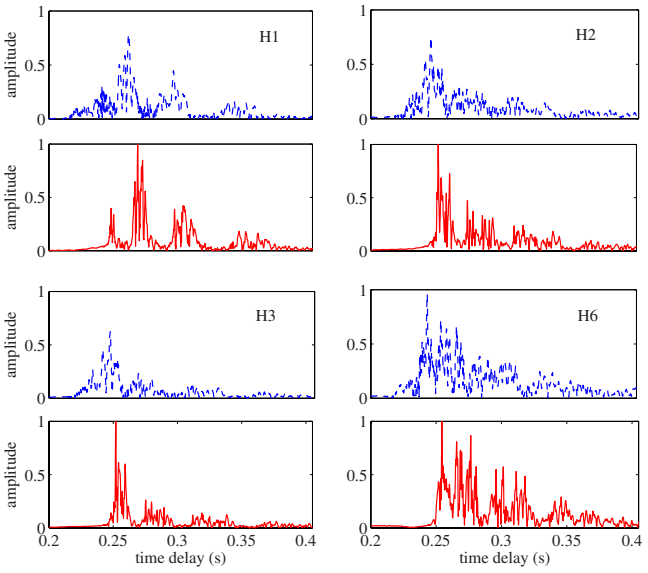


FIG. 13. (Color online) Time series of the IR estimated from four different hydrophones at different depths. The dashed lines represent the IR estimated with classical matched filtering and the solid lines stand for the motion-compensated IR. The depths of the hydrophones are H1 at −9 m, H2 at −82.5 m, H3 at −93.5 m, and H6 at −51 m.

arrival time of each group of rays. Hydrophones number two and three (H2 and H3) are located around sound speed minimum where the records have a high SNR. From H2 and H3 in Fig. 13, it can be seen that the motion compensated IRs have been corrected in time delays and amplitudes and present thinner shapes for each group of rays. Finally, hydrophone number six (H6) is located in the middle of the water column and its records have a low SNR, which leads to poor estimates of the IR for both the classical matched filtering and motion compensation methods.

VI. CONCLUSIONS

The effects of motion while estimating the IR for shallow water environments with signals having low central frequencies and high bandwidth must be taken into account. In this case, Doppler effects that cannot be modeled by a carrier frequency shift usually used to represent the narrowband cases, and wideband Doppler effects should be considered instead, modeled as a compression or expansion in time. The wideband ambiguity plane is presented here as a convenient way of representing multipath environments in a transmitter-receiver motion scenario. The uniform compensation method for motion effect compensation is proposed in the wideband ambiguity plane in order to reduce the distortions due to motion when the transmitted signal is known. This compensation method was tested on real set of data from BASE'07 campaign (SHOM, South of Sicilia, 2007) leading to realistic results.

ACKNOWLEDGMENT

This work was supported by Délégation Générale pour l'Armement (DGA) under SHOM research Grant No. N07CR0001.

APPENDIX A: EXPRESSIONS OF THE VIRTUAL SOURCE HEIGHT FOR A PROPAGATION CHANNEL WITH A CONSTANT SPEED

This appendix gives the expressions of z_i introduced in formula (5) for all rays considering that the propagation speed is constant. There are four possible expressions for z_i depending whether the number of reflections is even or odd, and on the first reflection. The first family of rays is the $2p+1$ rays which has an odd number of reflections and begins with a reflection at the surface of the sea. The expression for the height of the corresponding virtual source is

$$z_i^{2p+1} = 2(p+1)z_s - 2pz_b - z_h, \tag{A1}$$

where z_s, z_b, and z_h are the heights of the surface, of the bottom, and of the hydrophone compared to the position of the real source, respectively. p is an integer and $2p+1$ is the number of reflections. The second family of ray is $2p$ rays which has an even number of reflections and begins with a reflection at the surface of the sea. The expression for the height of the corresponding virtual source becomes

$$z_i^{2p} = 2pz_b - 2pz_s - z_h, \tag{A2}$$

where p is an integer and $2p$ is the number of reflections. The third family of rays is the $-(2p+1)$ rays which has an odd

number of reflections and begins with a reflection at the bottom of the propagation channel. The height of the corresponding virtual source is given by

$$z_i^{-(2p+1)} = 2(p+1)z_b - 2pz_s - z_h, \tag{A3}$$

where p is an integer and $2p+1$ is the number of reflections. Finally, the last family of rays, the $-2p$ rays, is made of rays having an even number of reflections beginning with a reflection at the bottom of the sea. The height of the corresponding virtual source follows,

$$z_i^{-2p} = 2pz_s - 2pz_f - z_h, \tag{A4}$$

where p is an integer and $2p$ is the number of reflections.

APPENDIX B: CHARACTERIZATION OF THE MULTIPATH RECEIVED SIGNAL

This appendix aims at explaining the calculation necessary to obtain expressions (8) and (9). The notations used in this part are the same as in Sec. II. The hypothesis that channel depth can be neglected compared with its length can be summarized as

$$(x_0 - v_i u)^2 \gg z_i^2. \tag{B1}$$

This hypothesis is used to approximate $L_i(u)$ defined in Eq. (5) and get a linear expression of u as a function of t. Expression (5) can be reformulated as

$$L_i(u) = (x_0 - v_i u)\sqrt{1 + \frac{z_i^2}{2(x_0 - v_i u)^2}}. \tag{B2}$$

Using hypothesis (B1) in Eq. (B2) yields

$$L_i(u) \approx x_0 - v_i u + \frac{z_i^2}{x_0 - v_i u}. \tag{B3}$$

The expression of $L_i(u)$ obtained in Eq. (B3) is then injected in Eq. (7) which leads directly to expressions (8) and (B4),

$$u + \frac{x_0 - v_i u}{c} + \frac{z_i^2}{2c(x_0 - v_i u)} = t. \tag{B4}$$

The first hypothesis (B1) allows expressing u as a function of t, but this expression is not yet linear. The second hypothesis made in Sec. II assumes that the distance the source moves during one transmission can be neglected compared with the source-hydrophone separation, which can be summarized as

$$x_0 \gg v_i u. \tag{B5}$$

Hypothesis (B5) approximates the part of Eq. (B4) containing the inverse of the time of emission u as

$$\frac{z_i^2}{2c(x_0 - v_i u)} \approx \frac{z_i^2}{2cx_0} + \frac{z_i^2}{2cx_0^2}. \tag{B6}$$

Some manipulations with Eqs. (8) and (B6) finally lead to the expression of u as a linear function of t:

$$u = \frac{t - \left(\frac{x_0}{c} + \frac{z_i^2}{2cx_0}\right)}{1 - v_i\left(\frac{1}{c} - \frac{z_i}{2cx_0^2}\right)}. \tag{B7}$$

APPENDIX C: THE MATCHED-FILTER OUTPUT FOR THE LFM CASE

Calculations necessary to obtain Eqs. (20) and (25) are explained in this appendix. First, recall expression (17) defining the result of the matched-filter output for any signal:

$$R(\tau,v) = \sum_i a_i(\eta\eta_i)^{1/2}\int_{-\infty}^{\infty} e(\eta_i(t+\tau-\tau_i))e^T(\eta t)dt, \tag{C1}$$

where $e(t)$ is a LFM signal with known parameters defined in Sec. III as

$$e(t) = \mathrm{rect}\left(\frac{t}{T}\right)\frac{1}{\sqrt{T}}\exp\left(j2\pi\left(f_c t + \frac{k}{2}t^2\right)\right). \tag{C2}$$

We introduce the substitution

$$t' = t + \frac{\Delta\tau_i}{2} \tag{C3}$$

to get a symmetric expression in $\Delta\tau_i$ which avoids the need to consider the two cases $\Delta\tau_i$ positive and $\Delta\tau_i$ negative. Relation (C1) becomes

$$R(\tau,v) = \sum_i a_i(\eta_i\eta)^{1/2}\int_{-\infty}^{\infty} e\left(\eta_i\left(t+\frac{\Delta\tau_i}{2}\right)\right)e^T\left(\eta\left(t-\frac{\Delta\tau_i}{2}\right)\right)dt. \tag{C4}$$

After some manipulations with Eqs. (C2) and (C4) we obtain

$$R(\tau,v) = \sum_i C_i D_i E_i\int_{-\infty}^{\infty} R_1 R_2 \exp\left(\frac{j\pi\gamma}{2}\left(\frac{\gamma\alpha_i}{\sqrt{|\beta_i|}} + 2t\sqrt{|\beta_i|}\right)^2\right)dt, \tag{C5}$$

where

$$R_1 = \mathrm{rect}\left(\frac{\eta_i\left(t+\frac{\Delta\tau_i}{2}\right)}{T}\right), \tag{C6}$$

$$R_2 = \mathrm{rect}\left(\frac{\eta\left(t-\frac{\Delta\tau_i}{2}\right)}{T}\right), \tag{C7}$$

The variables C_i, D_i, γ_i, α_i, and β_i are introduced in Sec. III by relations (20). The bounds of integration of Eq. (C5) depend on R_1 and R_2. They are called t_1 and t_2, and Appendix D explains how they are obtained. Equation (C5) is reformulated with the following change in variables:

$$X = \frac{\gamma\alpha_i}{\sqrt{|\beta_i|} + 2t\sqrt{|\beta_i|}}, \tag{C8}$$

leading to

$$R(\tau,v) = \sum_i \frac{C_i D_i E_i}{2\sqrt{|\beta_i|}}\int_{X_i}^{Y_i} \exp\left(\frac{j\gamma\pi}{2}t^2\right), \tag{C9}$$

where X_i and Y_i can be expressed as function of t_1 and t_2 according to the change in variables defined in formula (C8),

$$X_i = \frac{\gamma\alpha_i}{\sqrt{|\beta_i|} + 2t_1\sqrt{|\beta_i|}}, \tag{C10}$$

TABLE II. Integration bounds of expression (C5).

Range of $\Delta\tau_i$	t_1	t_2
$\frac{-T}{2}\left(\frac{\eta_i+\eta}{\eta_i\eta}\right) \le \Delta\tau_i < \frac{-T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert$	$\frac{-\Delta\tau_i}{2}-\frac{T}{2\eta}$	$\frac{\Delta\tau_i}{2}-\frac{T}{2\eta_i}$
$\frac{-T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert \le \Delta\tau_i \le \frac{T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert$ and $\eta_i < \eta$	$\frac{-\Delta\tau_i}{2}-\frac{T}{2\eta}$	$\frac{\Delta\tau_i}{2}+\frac{T}{2\eta}$
$\frac{-T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert \le \Delta\tau_i \le \frac{T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert$ and $\eta_i > \eta$	$\frac{\Delta\tau_i}{2}-\frac{T}{2\eta_i}$	$\frac{\Delta\tau_i}{2}+\frac{T}{2\eta_i}$
$\frac{T}{2}\lvert\frac{\eta_i-\eta}{\eta_i\eta}\rvert < \Delta\tau_i \le \frac{T}{2}\lvert\frac{\eta_i+\eta}{\eta_i\eta}\rvert$	$\frac{\Delta\tau_i}{2}-\frac{T}{2\eta_i}$	$\frac{-\Delta\tau_i}{2}+\frac{T}{2\eta}$

$$X_i = \frac{\gamma\alpha_i}{\sqrt{|\beta_i|}+2t_2\sqrt{|\beta_i|}}. \tag{C11}$$

Finally, the result of Eq. (C9) can be expressed and simplified with a complex form of the Fresnel integrals if v is different from v_i,

$$R(\tau,v) = \sum_i \frac{C_i D_i E_i}{2\sqrt{|\beta_i|}}(F(Y_i)-F(X_i)), \tag{C12}$$

where

$$F(u) = C(u) + j\gamma S(u), \tag{C13}$$

$$C(u) = \int_0^u \cos\left(\frac{\pi t^2}{2}\right)dt, \tag{C14}$$

$$S(u) = \int_0^u \sin\left(\frac{\pi t^2}{2}\right)dt. \tag{C15}$$

When the Doppler transformation of the reference signal matches exactly the Doppler transformation of the ith path, Eq. (21) is no longer valid. Equation (C5) becomes

$$R(\tau,u) = \sum_i C_i \int_{t_1}^{t_2} \exp(2j\pi tk\Delta\tau_i\eta^2)dt, \tag{C16}$$

where bounds of integration are given by R_1 and R_2, and when $\eta_i = \eta$ they satisfy

$$t_1 = \frac{|\Delta\tau_i|}{2} - \frac{T}{2\eta}, \tag{C17}$$

$$t_2 = \frac{T}{2\eta} - \frac{|\Delta\tau_i|}{2}, \tag{C18}$$

$$|\Delta\tau_i| \le \frac{T}{\eta}. \tag{C19}$$

The integration of formula (C16) finally gives the expression of the output matched-filter when Doppler transformation of the reference signal matches exactly with the Doppler transformation of the ith path

$$R(\tau,v) = \sum_i C_i\left(|\Delta\tau_i| - \frac{T}{\eta_i}\right)\frac{\sin(\xi_i)}{\xi_i}, \tag{C20}$$

where

$$\xi_i = \pi k\Delta\tau_i\eta_i(\eta_i|\Delta\tau_i| - T). \tag{C21}$$

APPENDIX D: DETERMINATION OF THE INTEGRATION BOUNDS

The bounds of integration of expression (C5) are called t_1 and t_2 and are given by R_1 and R_2. Different cases appear depending on both the length and the position of R_1 compared with R_2. When η_i is different from η, there are four possible values for each bound of integration and they may be sorted by the range of $\Delta\tau_i$, as summarized in Table II.

[1] Z. H. Michalopoulou, "Matched-impulse-response processing for shallow-water localization and geoacoustic inversion," J. Acoust. Soc. Am. **108**, 2082–2090 (2000).
[2] M. I. Taroudakis and G.-N. Makrakis, *Inverse Problems in Underwater Acoustics* (Springer-Verlag, New York, 2001).
[3] A. Baggeroer, W. Kuperman, and P. Mikhalevsky, "An overview of matched field methods in ocean acoustics," IEEE J. Ocean. Eng. **18**, 401–424 (1993).
[4] C. Gervaise, S. Vallez, Y. Stephan, and Y. Simard, "Robust 2d localization of low-frequency calls in shallow waters using modal propagation modelling," Can. Acoust. **36**, 153–159 (2008).
[5] C. Gervaise, S. Vallez, C. Ioana, Y. Stephan, and Y. Simard, "Passive acoustic tomography: New concepts and applications using marine mammals: A review," J. Mar. Biol. Assoc. U.K. **87**, 5–10 (2007).
[6] S. Qian and D. Chen, "Signal representation using adaptive normalized Gaussian functions," Signal Process. **36**, 1–11 (1994).
[7] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process. **41**, 3397–3415 (1993).
[8] H. Zou, Y. Chen, J. Zhu, Q. Dai, G. Wu, and Y. Li, "Steady-motion-based Dopplerlet transform: Application to the estimation of range and speed of a moving sound source," IEEE J. Ocean. Eng. **29**, 887–905 (2004).
[9] A. N. Guthrie, R. M. Fitzgerald, D. A. Nutile, and J. D. Shaffer, "Long-range low-frequency cw propagation in the deep ocean: Antigua-Newfoundland," J. Acoust. Soc. Am. **56**, 58–69 (1974).
[10] K. E. Hawker, "A normal mode theory of acoustic Doppler effects in the oceanic waveguide," J. Acoust. Soc. Am. **65**, 675–681 (1979).
[11] P. H. Lim and J. M. Ozard, "On the underwater acoustic field of a moving point source. I. Range-independent environment," J. Acoust. Soc. Am. **95**, 131–137 (1994).
[12] R. P. Flanagan, N. L. Weinberg, and J. G. Clark, "Coherent analysis of ray propagation with moving source and fixed receiver," J. Acoust. Soc. Am. **56**, 1673–1680 (1974).
[13] J. G. Clark, R. P. Flanagan, and N. L. Weinberg, "Multipath acoustic propagation with a moving source in a bounded deep ocean channel," J. Acoust. Soc. Am. **60**, 1274–1284 (1976).
[14] J. P. Hermand and W. I. Roderick, "Delay-Doppler resolution performance of large time-bandwidth-product linear fm signals in a multipath ocean environment," J. Acoust. Soc. Am. **84**, 1709–1727 (1988).
[15] C. L. Pekeris, "Theory of propagation of explosive sound in shallow water," *Propagation of Sound in the Ocean* (Geological Society of America, New York, 1948), Memoir 27, pp. 1–117.
[16] F. B. Jensen, W. A. Kuperman, and H. Schmidt, *Computational Ocean Acoustics* (AIP, New York, 1994).
[17] S. Kramer, "Doppler and acceleration tolerances of high-gain, wideband linear fm correlation sonars," Proc. IEEE **55**, 627–636 (1967).
[18] W. Adams, J. Kuhn, and W. Whyland, "Correlator compensation requirements for passive time-delay estimation with moving source or receivers," IEEE Trans. Acoust., Speech, Signal Process. **28**, 158–168 (1980).
[19] B. Harris and S. Kramer, "Asymptotic evaluation of the ambiguity functions of high-gain fm matched filter sonar systems," Proc. IEEE **56**, 2149–2157 (1968).
[20] N. F. Josso, C. Ioana, C. Gervaise, Y. Stephan, and J. I. Mars, "Motion effect modeling in multipath configuration using warping based lag-Doppler filtering," IEEE Trans. Acoust., Speech, Signal Process. **2009**,

2301–2304.
[21]G. Theuillon and Y. Stephan, "Geoacoustic characterization of the seafloor from a subbottom profiler applied to the BASE'07 experiment," J. Acoust. Soc. Am. **123**, 3108 (2008).
[22]N. Josso, C. Ioana, C. Gervaise, and J. I. Mars, "On the consideration of motion effects in underwater geoacoustic inversion," J. Acoust. Soc. Am. **123**, 3625 (2008).
[23]N. F. Josso, C. Ioana, J. I. Mars, C. Gervaise, and Y. Stephan, "Warping based lag-Doppler filtering applied to motion effect compensation in acoustical multipath propagation," J. Acoust. Soc. Am. **125**, 2541 (2009).

Acoustic mode radiation from the termination of a truncated nonlinear internal gravity wave duct in a shallow ocean area

Ying-Tsong Lin, Timothy F. Duda, and James F. Lynch
Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 26 January 2009; revised 20 June 2009; accepted 8 July 2009)

Horizontal ducting of sound between short-wavelength nonlinear internal gravity waves in coastal environments has been reported in many theoretical and experimental studies. Important consequences arising at the open end of an internal wave duct (the termination) are examined in this paper with three-dimensional normal mode theory and parabolic approximation modeling. For an acoustic source located in such a duct and sufficiently far from the termination, some of the propagating sound may exit the duct by penetrating the waves at high grazing angles, but a fair amount of the sound energy is still trapped in the duct and propagates toward the termination. Analysis here shows that the across-duct sound energy distribution at the termination is unique for each acoustic vertical mode, and as a result the sound radiating from the termination of the duct forms horizontal beams that are different for each mode. In addition to narrowband analysis, a broadband simulation is made for water depths of order 80 m and propagation distances of 24 km. Situations occur with one or more modes absent in the radiated field and with mode multipath in the impulse response. These are both consistent with field observations.
 [DOI: 10.1121/1.3203268]

PACS number(s): 43.30.Re, 43.20.Bi, 43.20.Mv, 43.30.Bp [JAC] Pages: 1752–1765

I. INTRODUCTION

Theoretical and experimental studies have revealed that propagating sound can be trapped and ducted between pairs of nonlinear internal waves of depression commonly observed in coastal areas. Here, we examine what happens when such a duct gradually or abruptly relaxes to the background conditions and the sound exits the duct. The horizontal ducting of sound between internal waves can be understood in terms of the horizontal rays of acoustic vertical modes bending and focusing in the duct, or by spiraling rays. Katsnel'son and Pereselkov[1] theoretically studied the effect using the theory of "horizontal rays and vertical modes."[2] Another study by Oba and Finette[3] which included results obtained with a computer code for three-dimensional (3D) sound propagation, FOR3D,[4] also predicted this ducting phenomenon. In another of their papers,[5] these investigators numerically investigated the consequences of this effect on horizontal array beamforming and also performed a modal decomposition of the sound field to explain its structure in the duct. Observations by Badiey *et al.*[6] and Franks *et al.*[7] from the SWARM'95 experiment off New Jersey confirmed these predictions. Typical internal wave packets capable of causing this effect have inter-wave distances of 300–1000 m and modal refractive index anomalies of a few tenths of a percent to 3%.[8]

Work to date has addressed these internal wave ducts as "infinite length pipelines" for sound, i.e., the internal wave crests are straight and endless. However, this is not the case in the real world. Many satellite and *in-situ* observations have shown internal wave packets (ducts) to be finite and localized, having total along-crest lengths as short as a few kilometers near an internal wave source region.[9] Prior work does not address what happens to internal-wave ducted sound when the duct terminates in the horizontal and the sound is emitted. The goal of this paper is to investigate this radiation from the duct termination. Analytical and numerical approaches are taken. In the analytical approach, an idealized internal-wave shape (square waveform) and the assumption of adiabatic mode propagation give a 3D normal mode solution for the sound field in the duct. Then Huygens' principle is adopted to calculate the radiation field. As for the numerical approach, a computer code[10] implementing a 3D parabolic approximation in Cartesian coordinates is employed to calculate the acoustic field and broadband impulses within and emitted from a duct. The use of this 3D parabolic approximation model allows more complicated (or more realistic) internal-wave shapes. Also, since this numerical model allows mode coupling, it can be used to verify the adiabatic mode assumption used in the analytical computation.

Both the analytical and numerical computations show horizontal interference patterns within the duct. Richly detailed sound radiation fields are predicted at locations far from the termination of a truncated internal wave duct, having mode-dependent patterns with strong azimuthal and temporal variability. Thus, effects of ducting are found at locations with no evidence of a ducting condition. Similar radiation effects were found in the data collected in the SW06 shallow water acoustic experiment.[11] Though we will not completely prove it in this paper (as this involves other analyses which are future work), the results shown here make a plausible explanation for at least some of the data variability measured in the field. As examples, the tempo-

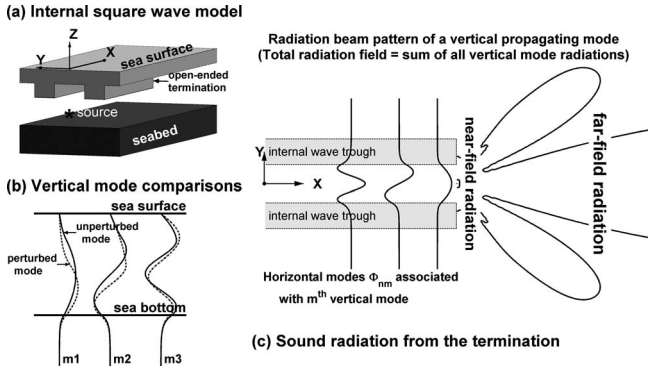


FIG. 1. Simplified internal-wave duct model used for theoretical analysis. (a) Internal waves of depression with square waveforms and parallel wavefronts are considered. The waves abruptly terminate to the background state. (b) Comparisons of acoustic vertical modes before and after being perturbed by the internal-square wave. (c) A depiction of sound radiation from the termination. Total radiation field equals to the sum of all vertical mode radiations.

rally varying modal contents of sound and mode multipath were observed and compare well qualitatively with the predictions made here.

The contents of this paper are as follows: Sec. II contains the derivation and evaluation of a 3D normal mode solution for an idealized environment. Section III contains a numerical parabolic equation solution for a more complex environment. Section IV contains field observation of the effects similar to those predicted by the theory. Section V has a discussion and Sec. VI contains a summary.

II. NORMAL MODE SOLUTION

The horizontal ducting of acoustic energy between internal waves can lead to a rather complicated horizontal eigenvalue problem if realistic detailed shapes of the internal waves that form the duct are considered. Also, the sound radiation from a real, irregular duct termination is not simple to handle by a theoretical means. So, rather than addressing a complicated case, we will begin with a simple duct model which will still retain most of the physics associated with the ducting and radiation effects. The detailed model configuration is provided below, along with the explanations for two assumptions used in deriving the normal mode solution.

A. Simplified physical model and assumptions

The model used for the theoretical analysis has two homogeneous water column layers bounded below by a homogeneous bottom, as shown in Fig. 1(a). The sound speed in the upper water column layer is slightly faster, and sound absorption is neglected in both the water column layers, but is included in the bottom. Two internal waves of depression are introduced by deepening the interface between the two aqueous layers from the ambient state to a perturbed state. Simplified internal waves with square and equal waveforms in the y direction, equal perturbation amplitudes, and parallel wave crests along the x-axis are used, forming an acoustic duct with slow normal velocity in the center, as will be shown. The waves abruptly terminate at a selected location in the x direction, forming an open end of the duct. Although the acoustic horizontal ducting modes in this square-wave approximate model may differ from the ones seen in realistic cases, the physical characteristics of these ducting modes can still be captured in the approximate model. In addition, if it is required, one can more carefully design an internal square-wave model to fit a realistic waveform such that the resultant horizontal ducting modes are comparable to the realistic ones. The coordinate origin of the Cartesian coordinate system is placed in the middle between the internal waves on the sea surface. The z axis is positive upward. At x locations less than termination position $x=L$ solutions valid in a ducted environment will be found. At $x>L$ solutions in a homogeneous layered environment will be computed using Huygens' principle, which is done by employing the Green's function method.

For the theoretical analysis, we make two following assumptions about sound propagation. (1) Mode propagation is adiabatic, with no energy exchange between modes. This assumption would be invalid if the addressed problem was sound propagation across the internal waves.[12,13] (2) Total transmission at the open-ended termination is assumed so that no sound energy is reflected back into the duct. This is justified by the fact that the termination width normal to the duct (300–1000 m across in a common situation) is very large compared to the underwater acoustic wavelength of frequencies greater than 50 Hz, which produces a large ka value (acoustic wavenumber multiplying the gap aperture) and thus should yield only tiny reflected sound by analogy with the unflanged, open pipe problem.[14] A final caution should be made here: for high grazing-angle penetration (at x near that of the source) the assumption of adiabatic mode propagation is violated, and the solution in terms of the resultant mode sum is not strictly correct. The horizontal leaky modes are included in the following derivation, but they only describe the barrier penetration and tunneling effect that occurs when the trapped sound strikes the internal waves at low grazing angle and radiates out of the duct in the y direction.

B. Derivation of an analytical solution

The sound pressure field excited by a harmonic point source located at (x_s,y_s,z_s) in the physical model described above is governed by the inhomogeneous 3D Helmholtz equation with a Dirac delta source function

$$\rho(z)\,\nabla\cdot\left(\frac{1}{\rho(z)}\,\nabla P(x,y,z)\right)+\frac{\omega^2}{c^2(x,y,z)}P(x,y,z)$$
$$=-4\pi\delta(x-x_s)\,\delta(y-y_s)\,\delta(z-z_s), \quad (1)$$

where $c(x,y,z)$ is the medium sound speed (this is a complex number with imaginary component enumerating absorption), $\rho(z)$ is the medium density (its horizontal variations are neglected), and $\omega=2\pi f$ is the acoustic frequency (f enumerates this in Hertz). To solve Eq. (1), begin with a vertical mode decomposition of the sound field

$$P(x,y,z)=\sum_m \Gamma_m(x,y)\Psi_m(x,y,z), \quad (2)$$

where $\Gamma_m(x,y)$ is the (complex) mode amplitude and $\Psi_m(x,y,z)$ is the mth vertical normal mode at (x,y) that

satisfies the following local mode equation with appropriate boundary conditions on the sea surface and the bottom:

$$\rho(z)\frac{d}{dz}\left(\frac{1}{\rho(z)}\frac{d}{dz}\Psi_m(x,y,z)\right)+\left(\frac{\omega^2}{c^2(x,y,z)}-\zeta_m^2(x,y)\right)\Psi_m(x,y,z)=0, \tag{3}$$

where ζ_m is the horizontal wavenumber of the mth vertical mode. Deriving analytic expressions for such vertical modes in the simplified, three-layer model considered here can be done by matching the interface conditions, and the details are omitted here. As shown in Fig. 1(b), the vertical mode functions are perturbed by the internal waves that comprise the duct. Also, the horizontal wavenumbers of vertical modes may be affected by the internal waves.

After replacing the sound pressure function in the 3D Helmholtz wave equation with the vertical mode decomposition and employing the assumption of adiabatic mode propagation, one can see that the vertical modal amplitude is governed by a two-dimensional (2D) Helmholtz equation, which is often called the horizontal refraction equation,[15] given by

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\Gamma_m+\zeta_m^2(x,y)\Gamma_m=-4\pi\frac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)}\delta(x-x_s)\delta(y-y_s). \tag{4}$$

This equation was employed in Ref. 5 for studying the focusing/defocusing structure in a straight and endless internal wave duct. Note how the horizontal wavenumber of the mth vertical mode, ζ_m, mimics the wavenumber $k=\omega/c$ in Eq. (1). In addition, the internal waves may alter the values ζ_m, yielding interface/boundary conditions across the waves in the y direction. In the model considered here the internal waves of depression have perturbed wavenumbers that are lower than the unperturbed values so that an incident vertical mode from the inside of the duct impinging horizontally onto the internal waves may encounter a situation where total reflection occurs and vertical modes are ducted in the horizontal plane, as shown in Fig. 1(c).

Solving the horizontal refraction equations for the vertical modal amplitudes allows the sound pressure field to be computed via mode summation. A two-step procedure is implemented. The first step is to obtain the ducted sound field between the simplified internal waves, and the second step is to apply Huygens' principle to obtain the radiation field due to the sound pressure distributed at the open-ended termination.

1. Ducted sound field

We begin by solving for the field at $x\leq L$, where ducted horizontal modes will exist. To solve Eq. (4) in this region, we can utilize many of the methods used in solving boundary value problems in shallow water acoustics, such as the well-known Pekeris waveguide mode problem. But, instead of looking at the constructive interferences of up- and down-going plane waves to get vertical modes, we are now seeking horizontal modes associated with each vertical mode Ψ_m. Due to barrier penetration and tunneling effects, which cause sound transmission through the internal wave "walls" and radiation outward in the y direction, Eq. (4) is not a proper Sturm–Liouville problem, so we cannot use an exact eigenfunction expansion to generate a solution. Instead, a wavenumber integration technique is employed, and a generalized eigenfunction expansion can be achieved for the vertical modal amplitude Γ_m if the branch-line integral is neglected. The detailed derivation is described below. Beforehand, note that the generalized expansion is of the form

$$\Gamma_m(x,y)=\sum_n A_{nm}(x)\Phi_{nm}(y), \tag{5}$$

where Φ_{nm} are the horizontal modes associated with each vertical mode Ψ_m, and A_{nm} are their (complex) amplitudes. A combination of indices (n,m) is required for the horizontal ducted modes. The complete solution for the ducted sound field will be of the form $P(x,y,z)=\Sigma_m\Sigma_n A_{nm}(x)\Phi_{nm}(y)\times\Psi_m(x,y,z)$. The horizontal mode functions only depend on y since the environment in the duct area does not have x-dependency. The amplitude A_{nm} depends on x only and will mainly vary due to modal phase delays and modal attenuation, with no geometric spreading loss for the trapped modes.

The wavenumber integration technique of solving Eq. (4) for the vertical modal amplitude Γ_m at $x\leq L$ is now described in detail. First, the assumption of total transmission at the open-ended termination allows us to reformat the problem to be an infinitely long duct problem, and the solution of the reformatted problem at the positions where $x\leq L$ (the internal waves terminate at $x=L$) is valid for representing the ducted sound field. Note that after we get the radiation field solution in Sec. II B 2, we will justify the assumption of total transmission. Since the horizontal wavenumber of the mth vertical mode, ζ_m, does not have x-dependency in the reformatted problem, we can apply the following plane-wave spectral integral or spatial Fourier transform to decompose the vertical modal amplitude in terms of a set of infinite plane waves:

$$\Gamma_m(x,y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}G_m(k_x,y)e^{ik_xx}dk_x, \tag{6a}$$

$$G_m(k_x,y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\Gamma_m(x,y)e^{-ik_xx}dx, \tag{6b}$$

where $G_m(k_x,y)$ is the plane-wave component of the mth vertical modal amplitude with wavenumber k_x. Applying the inverse transform operator to both sides of Eq. (4) yields a one-dimensional, y-dependent Helmholtz equation that is comparable to the depth-dependent Green's function equation in the wavenumber integration technique,[15]

$$\frac{d^2}{dy^2}G_m(k_x,y)+(\zeta_m^2-k_x^2)G_m(k_x,y)=-2\sqrt{2\pi}\frac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)}e^{-ik_xx_s}\delta(y-y_s), \tag{7}$$

which is subject to interface conditions on the internal-wave

"walls" that can be expressed as plane-wave reflection coefficients.

To determine $G_m(k_x,y)$ (and thence Γ_m) for $x \le L$, we first obtain a solution for the domain bounded by the internal waves $-D/2 \le y \le D/2$, where $y = \pm D/2$ are the inner boundaries of the internal waves, and then extend the bounded domain solution to the external domain by matching the interface condition at $y = \pm D/2$. Utilizing the endpoint method,[16] the plane-wave component $G_m(k_x,y)$ in the bounded domain ($|y| \le D/2$) can be found as

$$G_m(k_x,y) = \begin{cases} i\sqrt{2\pi}\dfrac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)} e^{-ik_x x_s} \dfrac{[e^{-ik_{y}y} + R_{Lm}e^{ik_{y}D}e^{ik_{y}y}][R_{Rm}e^{ik_{y}D}e^{-ik_{y}y_s} + e^{ik_{y}y_s}]}{k_y[1 - R_{Lm}R_{Rm}e^{2ik_{y}D}]} & \text{for } -D/2 \le y \le y_s, \\ i\sqrt{2\pi}\dfrac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)} e^{-ik_x x_s} \dfrac{[e^{-ik_{y}y_s} + R_{Lm}e^{ik_{y}D}e^{ik_{y}y_s}][R_{Rm}e^{ik_{y}D}e^{-ik_{y}y} + e^{ik_{y}y}]}{k_y[1 - R_{Lm}R_{Rm}e^{2ik_{y}D}]} & \text{for } y_s \le y \le D/2, \end{cases} \tag{8}$$

where $k_y^2 = \zeta_m^2 - k_x^2$, and R_{Lm} and R_{Rm} are the plane-wave reflection coefficients at $y = -D/2$ and $D/2$, respectively, as a function of k_x and the mth vertical modal phase speeds inside and outside the internal waves. The plane-wave reflection coefficients can be obtained using the usual formulas for plane-wave reflections from horizontally stratified multilayered media. Note that there is no horizontal density contrast in the system we are considering. To complete the solution for the vertical modal amplitude Γ_m, we insert Eq. (8) into Eq. (6a) and employ a complex contour integration technique. This immediately gives us the equation for the eigenvalues:

$$1 - R_{Lm}R_{Rm}\exp(2i\sqrt{\zeta_m^2 - k_x^2}D) = 0. \tag{9}$$

Since the plane-wave reflection coefficients R_{Lm} and R_{Rm} depend on which vertical mode is being considered, solving this eigenvalue equation leads to a set of horizontal mode wavenumbers for each vertical mode considered. There are two kinds of horizontal modes: trapped modes forming the ducted sound field and leaky modes forming the sound penetrating the internal waves and radiating laterally outward in the y direction. By selecting the Pekeris branch cut and neglecting the branch-line integral, the vertical modal amplitude Γ_m in the domain bounded by the internal waves ($-D/2 \le y \le D/2$) is found to be

$$\Gamma_m(x,y) = 2\pi i \frac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)} \sum_n \frac{\Phi_{nm}(y_s)}{\kappa_{nm} w_{nm}} e^{i\kappa_{nm}(x-x_s)} \Phi_{nm}(y), \tag{10a}$$

where w_{nm} is the normalization factor for the unnormalized horizontal mode function Φ_{nm},

$$w_{nm} = i[d(1 - R_{Lm}R_{Rm}e^{2ik_{y}D})/\, dk_y|_{k_x=\kappa_{nm}}^{k_y^2=\zeta_m^2-k_x^2}] \times [R_{Lm}e^{i\sqrt{\zeta_m^2-\kappa_{nm}^2}D}], \tag{10b}$$

$$\Phi_{nm}(y) = [e^{-i\sqrt{\zeta_m^2-\kappa_{nm}^2}y} + R_{Lm}e^{i\sqrt{\zeta_m^2-\kappa_{nm}^2}(y+D)}] \quad \text{for } -D/2 \le y \le D/2, \tag{10c}$$

and κ_{nm} is the x-component wavenumber of the nth horizontal mode resulting from the eigenvalue equation, Eq. (9), given that we have mth vertical mode excited. The normalization factor w_{nm} can be found numerically by employing a series of chain rules. This involves finding the first derivatives of the reflection coefficients with respect to k_y. In addition, taking a total derivative of the eigenvalue equation, Eq. (9), with respect to the frequency ω at $k_x = \kappa_{nm}$ yields the group slowness of the horizontal modes ($dk_x/d\omega$) and hence the group velocities. To extend the solution shown in Eq. (10a)–(10c) to $|y| > D/2$ with the assumption of adiabatic mode propagation, one can determine the unnormalized horizontal mode function outside the bounded domain by satisfying the continuity and smoothness conditions across all of the horizontal interfaces. For the horizontal ducted modes of interest, the horizontal trapped mode functions are sinusoids within the bounded domain and decay exponentially to vanish outward in the y direction. Note that, in the complex contour integration, without compensation of the branch-line integral for the exponential growth of the horizontal leaky modes in the y direction, the solution does not converge along the y-axis when the x-coordinate of a calculation point is close to the source. However, the boundary of this diverging area gets further and further out in the y direction as the x-coordinate of the calculation point moves away from the source, and so the branch-line integral can generally be neglected when the computation domain of interest expands reasonably in the x direction. After all, in the domain of small x and large y the sound field is dominated by the mode coupling effect occurring as the sound penetrates the internal waves at high grazing angles. This coupling is beyond the limitation of our 3D normal mode theory, so the branch-line integral failure is inconsequential.

2. Radiation field

The contribution of each vertical mode to the sound field at the termination of the duct ($x = L$) can now be computed from Eq. (10a)–(10c), and then used to obtain the radiation field from the termination by adopting Huygens' principle. Under the prior assumptions, the ducted sound will pass through the termination and radiate outward, with vertical modal energy staying in the same mode (adiabatic mode propagation). Hence, the total radiation field can be obtained by summing up all the vertical mode radiations.

The amplitude of the mth vertical mode at every termination position $(x=L)$ is given by Eq. (10a)–(10c). According to Huygens' principle, solutions at $(x>L)$ must be consistent with an infinite number of "modal" point sources adjoining each other across the termination, radiating the modal energy outward. Mathematically, the solution is found using the Green's function method. The required Green's function $\mathcal{G}_m(x,y;\xi,\eta)$, where (ξ,η) indicates the termination position, i.e., $\xi=L$, is symmetric and equals $\mathcal{G}_m(\xi,\eta;x,y)$ by the principle of reciprocity. This is governed by

$$\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)\mathcal{G}_m(\xi,\eta;x,y)+\zeta_m^2\mathcal{G}_m(\xi,\eta;x,y) = -\delta(\xi-x,\eta-y). \tag{11}$$

Note that since the mth vertical modal amplitude at the termination provides a Dirichlet boundary condition for the radiation field, a homogeneous Dirichlet boundary condition for the Green's function $(\mathcal{G}_m(\xi,\eta;x,y)=0)$ is required at $\xi=L$. The method of images yields

$$\mathcal{G}_m(x,y;\xi,\eta)=\frac{i}{4}H_0^{(1)}(\zeta_m\sqrt{(\xi-x)^2+(\eta-y)^2}) - \frac{i}{4}H_0^{(1)}(\zeta_m\sqrt{(\xi+x-2L)^2+(\eta-y)^2}), \tag{12}$$

where $H_0^{(1)}$ is the zeroth order Hankel function of the first kind. With this Green's function, it is straightforward to obtain the radiated mode amplitude at $x>L$, i.e., $\Gamma_m(x,y)=\int\Gamma_m(L,\eta)\partial\mathcal{G}_m(x,y;\xi,\eta)/\partial\xi|_{\xi=L}d\eta$, and $\Gamma_m(L,\eta)$ is the mth vertical modal amplitude across the termination. Substituting the necessary terms yields

$$\Gamma_m(x,y)=-\pi\frac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)}\zeta_m\sum_n\frac{\Phi_{nm}(y_s)}{\kappa_{nm}w_{nm}}e^{i\kappa_{nm}(L-x_s)} \times\int_{-\infty}^{\infty}\Phi_{nm}(\eta)\frac{x-L}{\sqrt{(x-L)^2+(y-\eta)^2}} \times H_1^{(1)}(\zeta_m\sqrt{(x-L)^2+(y-\eta)^2})d\eta, \tag{13}$$

where $H_1^{(1)}$ is the first order Hankel function of the first kind, resulting from differentiation of $H_0^{(1)}$. It is numerically straightforward to implement this Hankel integral, which is done in Sec. II C. Alternatively, with the assistance of the principal asymptotic form of the Hankel function for large arguments, the far-field approximation is given by

$$\Gamma_m(x,y)=2\pi e^{i(\pi/4)}\sqrt{\zeta_m}\frac{\Psi_m(x_s,y_s,z_s)}{\rho(z_s)} \times\sum_n\frac{\Phi_{nm}(y_s)}{\kappa_{nm}w_{nm}}e^{i\kappa_{nm}(L-x_s)}\Omega_{nm}(r,\theta), \tag{14}$$

where $r=\sqrt{(x-L)^2+y^2}$, $\theta=\tan^{-1}[y/(x-L)]$, $\Omega_{nm}(r,\theta)=r^{-1/2}e^{i\zeta_m r}\sin\theta\hat{\Phi}_{nm}(\theta)$, and $\hat{\Phi}_{nm}(\theta)=(2\pi)^{-1/2}\int_{-\infty}^{\infty}\Phi_{nm}(\eta)\times e^{-i2\pi u\eta}d\eta$, where the u in the exponent equals $\zeta_m\sin\theta/(2\pi)$.

The normal mode solutions for the ducted field and the radiation field have been obtained, and now we shall re-visit the prior assumption of total transmission at the open-ended termination. Remember that since the termination gap is very large compared to the underwater acoustic wavelength of interest in a common case, the ka value (acoustic wavenumber multiplying the gap aperture) is large and thus only tiny amount of reflection energy is yielded, which means that the assumption of total transmission should be valid. In addition, we can examine the continuity of the normal derivative of the pressure field solution (the smoothness) across the termination, as we notice there is only pressure continuity condition being used in deriving the radiation field solution. If the assumption of total transmission is sustained, the pressure field solution should be smooth across the termination, as shown in the next computational example.

C. Computational example

A numerical integration scheme is utilized to compute the analytic solution shown in Eq. (13), and the environmental model considered here is illustrated in Fig. 1(a). To remind the reader, a 3D Cartesian coordinate system (x-y-z) is chosen for positioning (see the figure for the orientation). The water column has two homogeneous layers; the sound speed in the upper layer is 1520 m/s, faster than the lower layer, where the sound speed is 1480 m/s. The density in the water column (in both layers) is 1.0 g/cm^3. The water depth is 80 m, and the thickness of the upper water layer is 20 m. Two nonlinear internal waves with square waveform disturb the water column, and their wavefronts are both parallel to the x-axis. These internal square waves have the same amplitude, 20 m, and wavelength, 200 m, and they both abruptly terminate at $x=20$ km. The gap between the internal waves is 300 m wide. The bottom is considered to be homogeneous, with sound speed of 1700 m/s, sound attenuation coefficient of 0.5 dB/λ, and density of 1.5 g/cm^3.

The first result shown here is the dependence of the radiation patterns of the vertical modes on the acoustic source position. A sinusoidal 100-Hz source is considered, and Fig. 2 shows the intensity contours of the first and second vertical modes excited by the source located at four different x-y positions in the gap between nonlinear internal waves. The internal waves form a horizontal duct at positions $0\leq x\leq 20$ km where the sound transmitted from the source is trapped. The waves terminate at $x=20$ km, and the trapped sound radiates away at this position. Note that the source depth is fixed ($z_s=-70$ m), and the internal waves have a square shape so that excitation of vertical modes is the same for each of the four source positions. There are more than two vertical modes excited by the source, but only the first two of them are shown in the plot. As shown in the figure, each vertical mode has its unique radiation pattern. That is because the across-duct energy distribution of each vertical mode at the termination is unique [see Eq. (13)]. Note also that the radiation pattern of each vertical mode differs dramatically for different source positions in the duct between the waves. The reason is that for a given vertical mode the excitation of the horizontal modes in the duct strongly de-

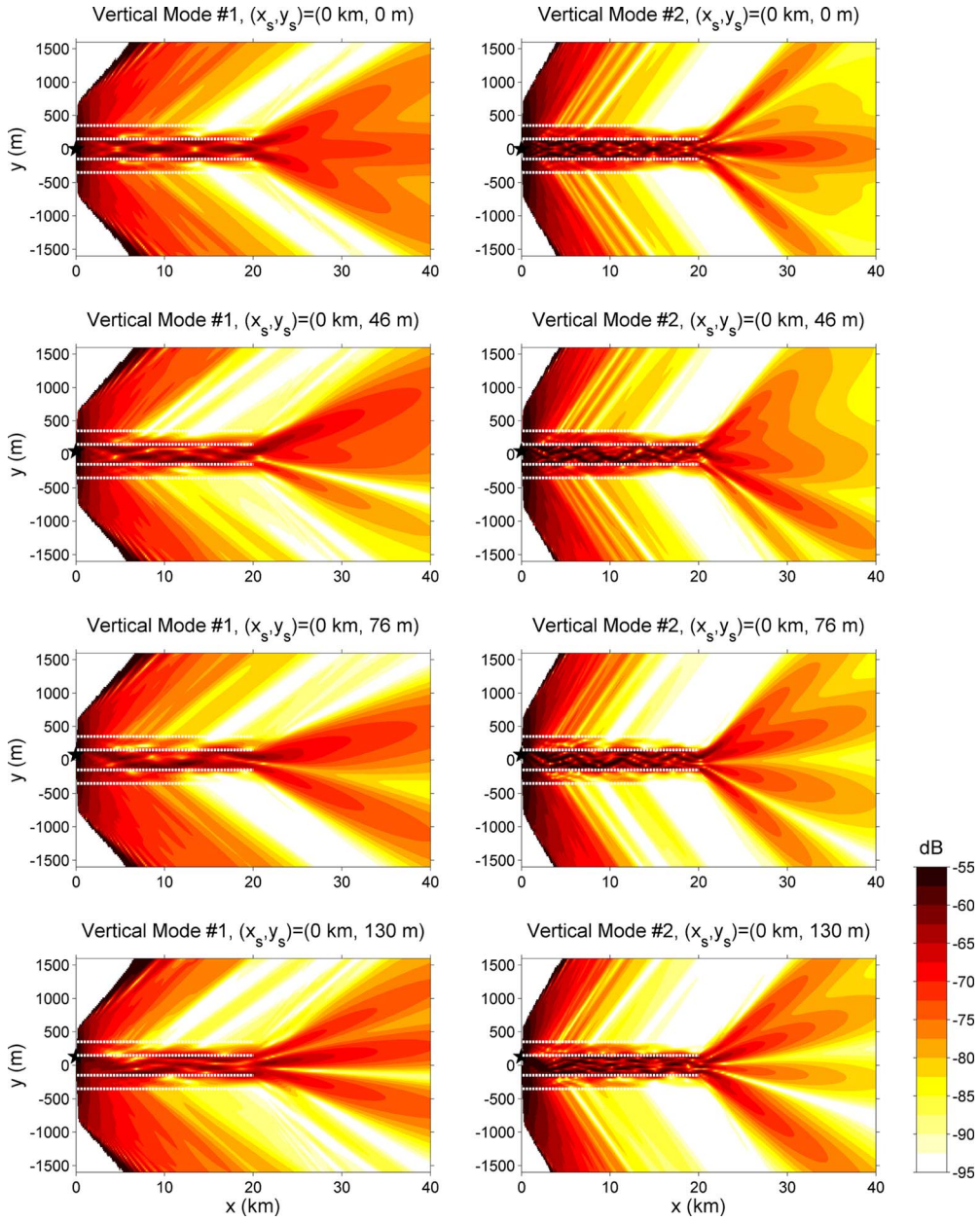


FIG. 2. (Color online) Intensity is shown for the first two vertical modes excited by a sinusoidal 100-Hz source located at four different y positions (shown from top to bottom). The source is always in the gap between the nonlinear internal waves, and the simplified model of Fig. 1 is used. Mode 1 intensity is shown on the left, mode 2 on the right. The source depth is fixed. The edges of internal waves are indicated by dashed lines. See the main text for the detailed environmental and acoustic source parameters.

pends on the source position in the y-axis [see Eq. (10a)–(10c)], so the across-duct energy distribution of the given vertical mode at the termination is a function of source position, and thus the radiation pattern is a function of source position.

The next computational result shows broadband sound radiation from the termination of an identical truncated internal square-wave duct. The previous harmonic source is replaced by a source emitting a broadband signal which contains 16 periods of a 100-Hz sine wave and is tapered by the Hann window. Note that 99.95% of the source energy is contained in the 25-Hz bandwidth centered at 100 Hz. This broadband source is fixed at a position of $(x_s, y_s, z_s) = (0\ \text{km}, 0\ \text{m}, -70\ \text{m})$, and a hydrophone array is placed at $x = 30$ km (10 km distant from the termination of the internal-wave duct). The array has a vertical component [vertical line array (VLA)] covering the whole water column and located at $y = 0$ m and a horizontal component [horizontal line array (HLA)] placed on the bottom and extending along the negative y-axis for 1000 m.

The internal waves form a horizontal acoustic duct at positions $0 \leq x \leq 20$ km. Horizontal modal dispersion in the duct is examined here because this phenomenon affects the broadband result. The dispersion of horizontal modes can be quantified by modal group velocities. Figure 3 shows how, for this duct configuration, the group velocities of the horizontal modes associated with the second vertical mode vary with the horizontal mode number and the frequency. The group velocities of horizontal modes have upper limits given by the group velocity computed using profiles in the wave

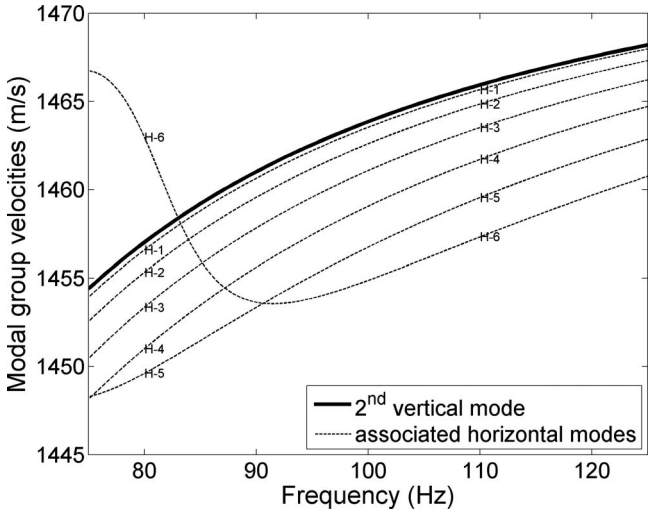


FIG. 3. Group velocities of the unperturbed second vertical mode (solid line) and six of its associated horizontal ducting modes (dashed lines) are shown for the simplified model of Fig. 1. The six horizontal modes are labeled by numbers.

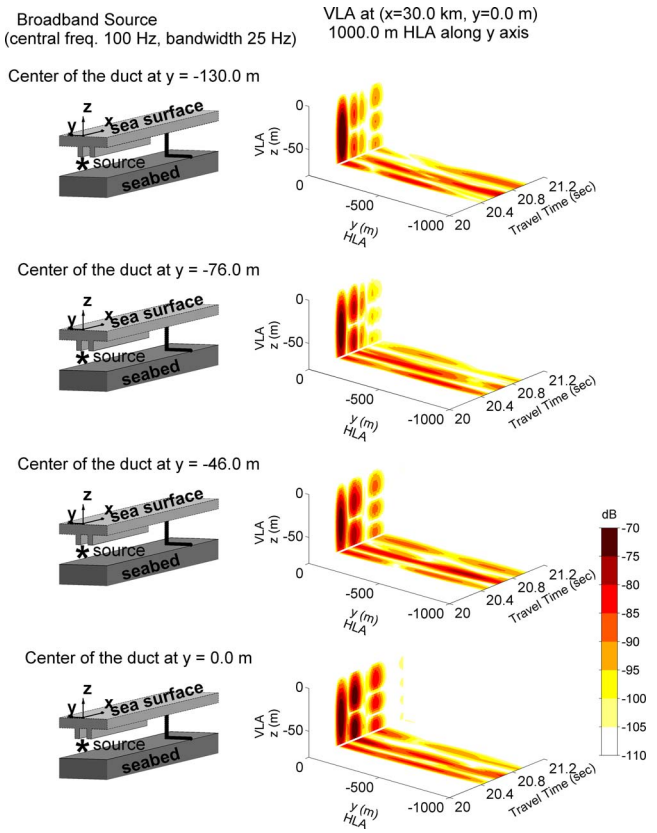


FIG. 4. (Color online) Broadband sound radiation from the termination of the internal-wave duct in the simplified mode. From top to bottom, the y position of the source with respect to the waves differs. The right panels illustrate the pulse intensity received at the horizontal and vertical arrays, denoting by thick lines in the left panels. See the main text for the detailed information about the model parameters.

crests (i.e., away from the wave duct region, which happens to be the profile in the duct itself for this idealized configuration). However, when the horizontal mode angle with respect to the internal-wave crests, determined from $\cos^{-1}(\kappa_{nm}/\zeta_m)$, nears the critical grazing angle for total internal reflection, the horizontal modal energy tunnels through the internal-wave trough, where the vertical modal phase and group velocities are higher, and thus the horizontal modal group velocity increases and exceeds that limit (see the sixth horizontal mode in Fig. 3).

The four panels in Fig. 4 illustrate the magnitude of the broadband acoustic time series received on the hydrophone arrays (both on the VLA and the HLA) when the truncated internal-wave duct is located at four different positions along the y-axis. As shown in the plot, each vertical mode arrival is well separated in time at the array; the arrivals on the VLA nicely represent the vertical mode shapes, and the arrivals on the HLA provide a means for us to observe the energy distribution of each vertical mode along the y-axis. Due to the change in source position in the internal-wave duct, modal radiation patterns at these four cases are different, and we do see the intensity of modal arrivals changing. This broadband source calculation does show the effect of horizontal modal dispersion (see the multiple arrivals of the second vertical mode in the top two panels of the plot). This is indeed a nice illustration in the time domain that the modal field now has vertical and horizontal mode numbers.

III. 3D PARABOLIC EQUATION MODELING

In the ocean, background sound speed profiles and nonlinear internal waves are not as simple as in the previous simplified model, and it would be challenging if not impossible to obtain an analytical solution for the 3D sound field. In order to handle a more realistic case, an acoustical propagation program [10] using the 3D parabolic approximation is employed here. The modeling technique utilized in this program is discussed briefly below, and the reader is referred to the technical report [10] for further details. Note that the parabolic approximation is more complete than the approach of Sec. II because it allows mode coupling, which has been observed to occur in the field, but is not expected to occur near the termination which we are modeling.

A. Split-step Fourier algorithm

The acoustical propagation program we use employs the split-step Fourier (SSF) technique[17] to solve the 3D parabolic acoustic wave equation (PE) for one-way propagating waves from a harmonic source in a Cartesian coordinate system. The SSF technique divides propagation over each distance increment through a heterogeneous sound speed environment into step-by-step "free space" propagation through a medium having a fixed reference wavenumber and periodically introduced (at each step) phase fluctuations consistent with departures from that fixed speed. The free space propagation is handled in the wavenumber domain, and the phase anomalies are introduced in the spatial domain. Amplitude effects such as absorption are introduced with the phase anomalies. Thus, each step (defined to be in the x direction) requires a 2D Fourier transform and an inverse 2D Fourier transform. Note that the wide-angle variant of the propagation operator[18] is used.

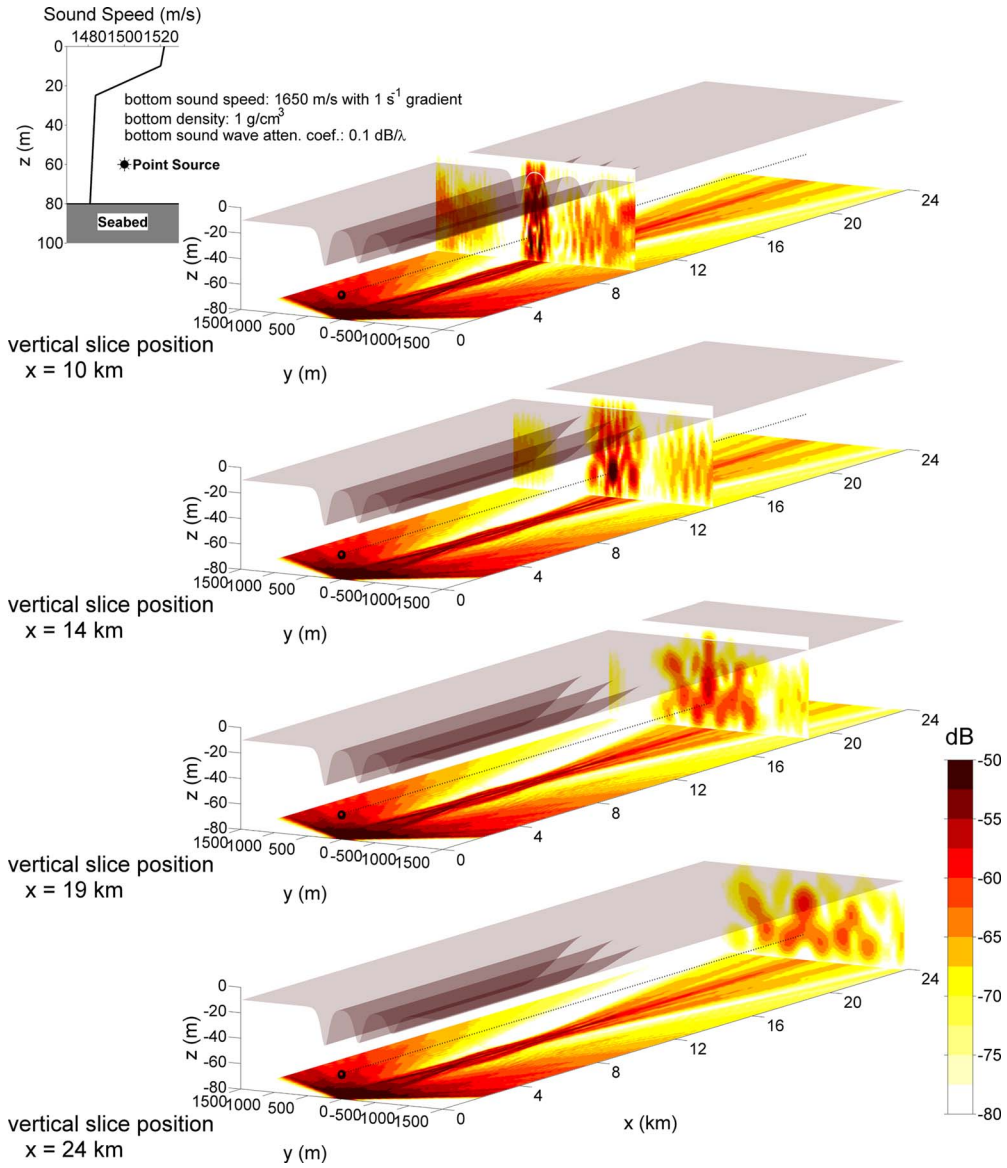


FIG. 5. (Color online) Numerical simulation of 3D sound propagation in an internal-wave duct model. (a) Upper left panel: environmental conditions in the numerical calculation. The sound speed profile represents a downward refracting situation in shallow water ocean. A sinusoidal 200-Hz source is located at 60-m depth. (b) Lower right panel: illustrations of 3D sound pressure fields. On the bottom of the visualization volume a 2D distribution of depth-averaged sound intensity is shown. The vertical slices show intensity on the slice. The transparent surface illustrates the displacement of the top of thermocline layer, i.e., internal waves.

B. Numerical simulation example

The environmental conditions for the numerical simulation using the Cartesian 3D PE program are illustrated in the upper-left panel of Fig. 5 and are described here. The water depth is 80 m, and a typical downward refracting sound speed profile observed in the ocean is considered. The bottom geoacoustic properties are listed in the figure. Note that the bottom density is set to be 1 g/cm^3 because the current version of this Cartesian 3D PE program does not allow medium density to vary; however, the conclusion drawn from the simulation result should still be valid. The nonlinear internal-wave chain moving in the y direction consists of three solitons, which depress the thermocline into the deeper water column. Although the solitons have different amplitudes, they all satisfy the Korteweg-de Vries (KdV) nonlinear internal-wave equation. Along the soliton wavefront (see the transparent surface shown in the lower-right panels of Fig. 5), the internal-wave waveform extends to $x=10$ km distant from the acoustic source with a permanent shape and then gradually diminishes to zero within a 4 km distance (14 km total).

The first numerical calculation is for narrowband sound propagation. The acoustic source, emitting 200-Hz sound, is placed between the biggest two solitons and at 60-m depth in the water, close to a node of the third acoustic mode. The calculation domain of the 3D PE program is configured in such a way that we can simulate sound propagating in the nonlinear internal-wave duct and radiating outward from the termination of the duct. A mode filtering is employed to obtain acoustic mode amplitudes at each (x,y) position from the full-field solution. The 3D visualization volume in Fig. 5 shows the simulation results. On the bottom of the visualiza-

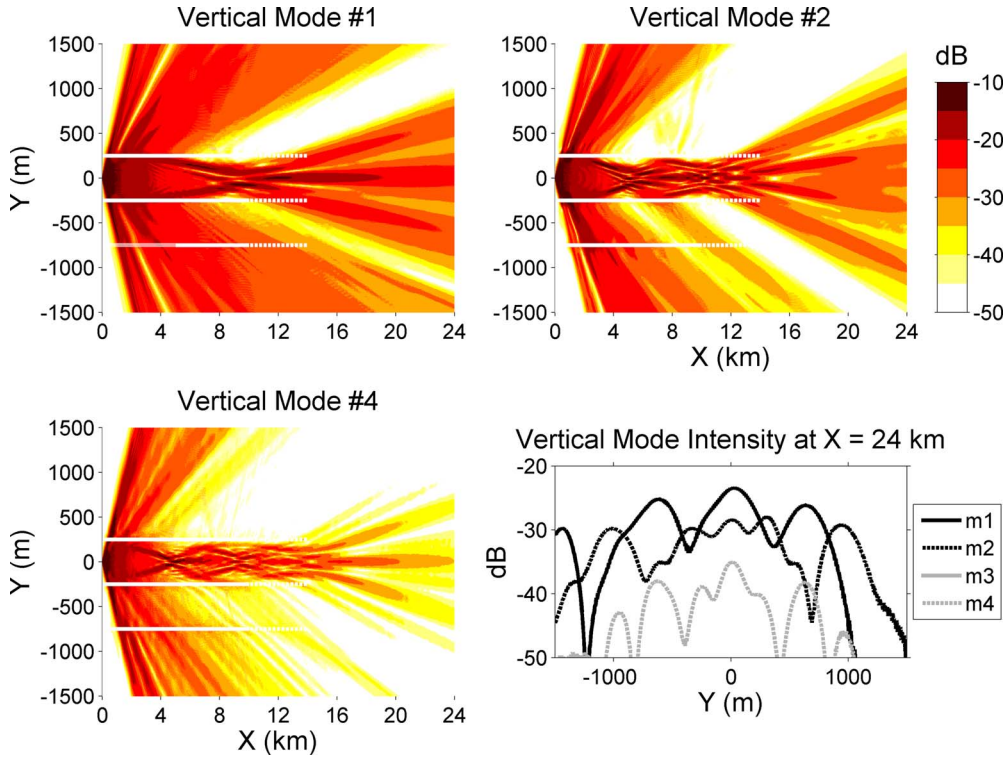


FIG. 6. (Color online) Vertical mode intensity is shown for propagation in the environmental model of Fig. 5. The solid lines are along internal-wave troughs, and the dashed lines indicate the "diminishing region" of the internal waves. See the main text for detailed discussions. At the lower right, the modal intensity at the right hand edge of other panels is plotted.

tion volume a 2D distribution of depth-averaged sound intensity is presented, and a vertical slice slides through the field to show detailed 3D structure. The mode amplitudes resulting from mode filtering are shown in Fig. 6 as horizontal contours, from which the following is observed.

Within the first 10-km propagation distance, despite some of the sound energy escaping at high angles with respect to the duct direction, the modal interference process yields a strong focusing effect in the internal-wave duct, which is consistent with what has been reported in the literature.[1,3,5–7] The radiation beam pattern, unique to each mode, starts to develop in the diminishing region of internal waves and becomes fully developed after the sound exits the duct with the features suggested by the simplified analytic model in Sec. II. Note that from the plot of mode amplitudes (Fig. 6) no mode-coupling effect is observed in both the ducted sound area and the radiation field, which confirms the assumption of adiabatic mode propagation used in deriving the normal mode solution.

The next case is broadband sound propagation, and the previous narrowband source is replaced by a source emitting a broadband signal which contains 16 periods of a 200-Hz sine wave and is tapered by a Hann window. Since 99.95% of the source energy is contained in the 50-Hz band centered at 200 Hz, one can calculate the fields for frequencies from 175 to 225 Hz (at N-Hz spacing) and sum them to produce a N^{-1}-second-long broadband reception at any receiving point. The four panels in Fig. 7 illustrate the magnitude of the broadband signals received at four VLAs distributed across the wavefront at the end of the propagation distance (X =24 km). The VLA locations are chosen to reveal the spatial variability of broadband sound radiation from the open end of the nonlinear internal-wave duct. A good reference VLA is the one located at the source axis (labeled as Y=0 m), where all vertical modes except for the third mode are distinct. Note that since the source is placed close to a depth null of the third mode, insignificant modal energy is excited. First we see that at Y=300 m the fifth mode fades out, and (more dramatically) at Y=1101 m only the second mode is observable. In addition, as seen in the previous analytical compu-

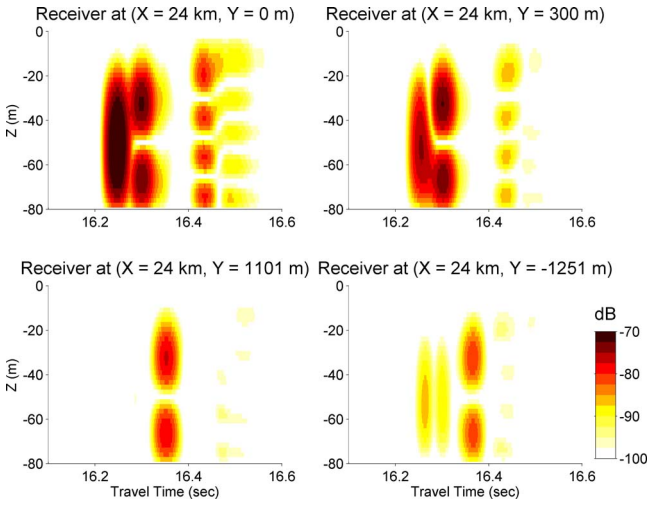


FIG. 7. (Color online) Simulation results for broadband sound (50-Hz bandwidth centered at 200-Hz) propagation in the environmental model of Fig. 5. These four panels show the intensity of the broadband signals received at four vertical hydrophone line arrays distributed across the wavefront at X =24 km.

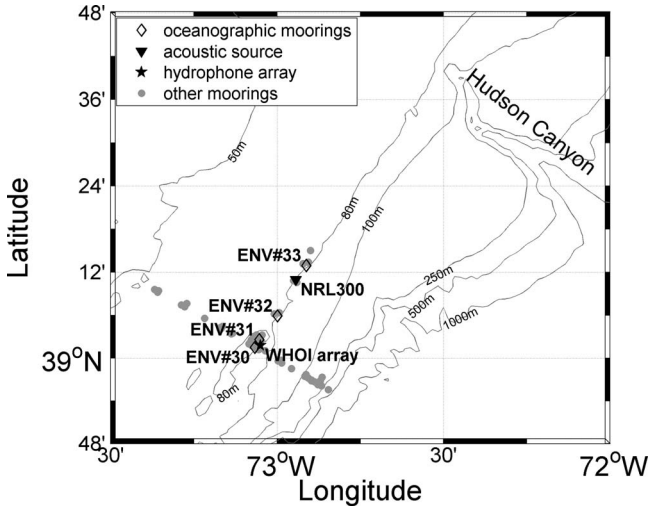


FIG. 8. Study area of the SW06 experiment. The depth of each isobath line is labeled. A total of 62 moorings was deployed in a "T" geometry to create an along-shelf path and an across-shelf path. The data collected from the six labeled moorings on the along-shelf path are used in this paper.

tation, multiple arrival of the same mode is also observed in this numerical calculation (mode 1 at $Y=-1251$ m). All of these plots truly convey the complexity of the sound field due to the modal radiation effect.

IV. EXPERIMENTAL DATA

In Secs. II and III theoretical and numerical analyses have been implemented to study acoustic mode radiation from the terminations of idealized square-wave and KdV-model internal-wave ducts. Here we shall examine acoustic signals measured in the coastal ocean for evidence of features similar to those seen in the computational examples. In the summer of 2006 a large multi-disciplinary experiment, SW06,[11] was conducted on the Mid-Atlantic Bight continental shelf at a location about 160 km east of the New Jersey coast and about 80 km southwest of the Hudson Canyon (Fig. 8). A total of 62 acoustic and oceanographic moorings was deployed in a "T" geometry to create an along-shelf track (following the 80-m isobath line) and an across-shelf track (depths changing from 50 to 500 m). In this paper, we will focus on the data collected from some of the moorings deployed on the along-shelf track. At the southern end of the along-shelf track, an underwater hydrophone-array system (labeled by "WHOI array" as it was deployed by the acoustics group from the Woods Hole Oceanographic Institution) was positioned. This system has two arrays: one is a 16-channel VLA covering the water column from depth 13.5 m to the bottom, and the other is a 465-m-long HLA with 32 uniformly-spaced hydrophones deployed on the bottom in a strict N-S orientation. There were many acoustic sources deployed in the water during the SW06 experiment, and the specific source of interest on the along-shelf track is a source transmitting a frequency modulated (FM) sweep sound centered at frequency of 300 Hz with 60-Hz bandwidth (labeled by "NRL300" as this source is owned by the U.S. Naval Research Laboratory). Water temperature data collected at the NRL300 source mooring and the WHOI array will be used to identify internal waves. Four oceanographic moorings deployed on the along-shelf track (labeled by ENV#30-33) also provide other useful environmental measurements, including water temperature and current speeds.

The environmental data indicate that a strong nonlinear internal-wave packet with a somewhat straight wavefront started passing through the along-shelf mooring track at about 10:00 Greenwich mean time (GMT) on August 20 (see Fig. 9). The angle between the wavefront and the along-shelf track was about 6 deg. The environmental data also indicate that this internal-wave packet terminated between moorings ENV#32 and ENV#31, and that the exact terminating point should be around 13 km distant from the NRL300 source and 5 km distant from the WHOI array. The NRL300 source was designed to transmit sound for 7.5 min at every hour and half-hour. At the next transmission (10:30 GMT) after the internal-wave packet reached the NRL300 source, the source was still in the packet. Hence it can be expected that the sound transmitted from the source would be trapped in the duct formed by the internal waves and propagate toward the termination of the internal-wave packet between moorings ENV#32 and ENV#31. The trapped sound would then radiate out from the termination and finally reach the WHOI array. In the following, we will see that the sound field received at the array was structured in a way that suggests that ducting and radiation from the duct had strong influences.

The NRL300 source linearly swept over 270–330 Hz in exactly 2.048 s every 4.096 s for 110 times, which yielded a 7.5-min-long transmission, at every hour and half-hour. Each sweep was tapered with a 0.2048-s amplitude taper (10% cosine taper) at the beginning and the end to allow gradually ramping on and off. A standard matched filter can compress the sweep to 33 ms, but with significant side-lobe ripples that could mask the multi-modal arrivals in this case. To reduce the side lobes, one can apply a taper window on the replica waveform of the matched filter[19] at the cost of increasing the (compressed) pulse length. In this data analysis the Hamming window function is applied, and the matched filter with such a tapered replica waveform produces 66-ms-long compressed pulses with side lobes less than −30 dB. Three of the 110 pulses are shown in Fig. 10, where one can clearly see complicated temporal and spatial variability. To further analyze the data and quantify the variability, the following modal analyses are performed.

Modal excitation by the NRL300 source during the transmission period encountering the truncated internal-wave packet is calculated by the acoustic normal mode program KRAKEN.[20] The *in situ* water sound speed measurements on the NRL300 source mooring and a bottom model provided by a previous study[21] are used in the program. The calculation results, shown in Fig. 11, indicate that the variation in modal energy excitation caused by the local water-column fluctuations is less than 3 dB, with the maximum variation occurring in the first mode. A series of broadband mode filterings is performed next to examine the vertical modal content of the received pulses on the WHOI VLA. The vertical mode functions used in the filter are calculated by the KRAKEN program, and the *in situ* water sound speed measurements on the WHOI VLA and the same bottom model

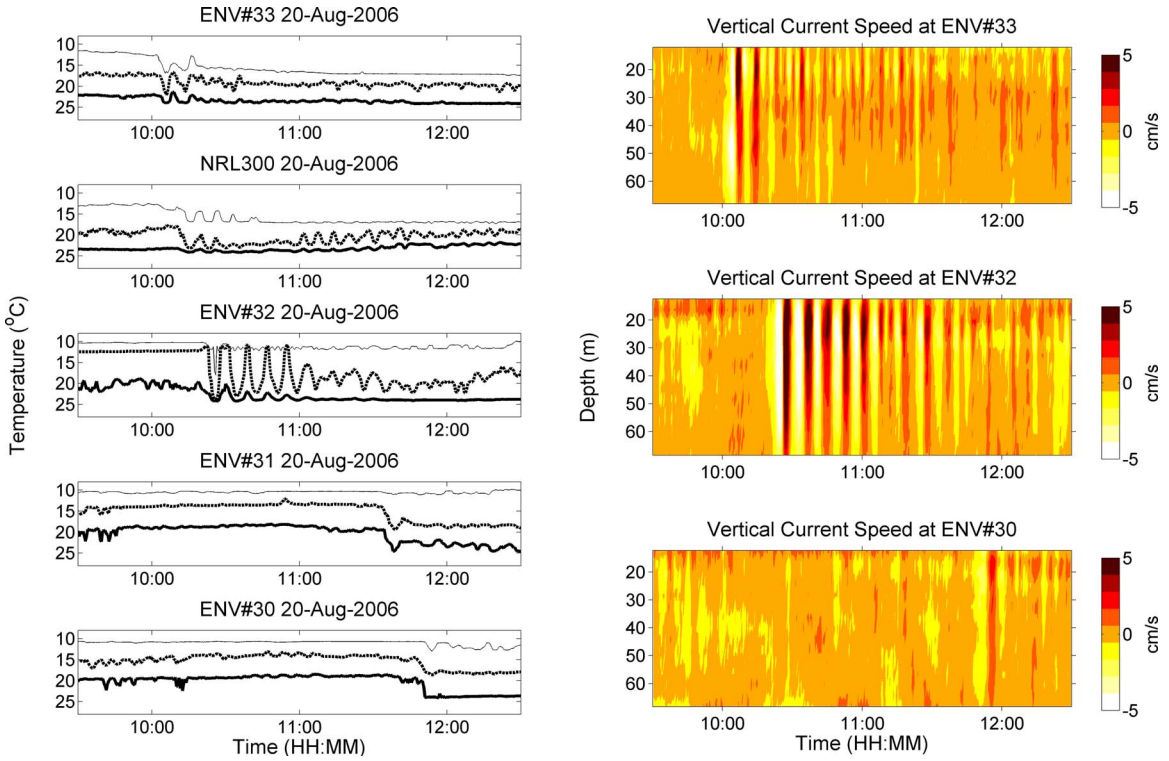


FIG. 9. (Color online) Environmental data collected from the moorings distributed along the 80-m-isobath track (total 30 km long) in the SW06 experiment on August 20 from 8:00 to 14:00 (GMT). Left panels: temperature data at three different depths about 10 m (thick line), 20 m (dashed line), and 30 m (thin line). Right panels: vertical current speeds measured by the acoustic Doppler current profilers. Both temperature and current speed data show that a truncated nonlinear internal-wave packet, terminating between moorings ENV#32 and ENV#31, passed through the mooring track.

for modal excitation calculation are used. Figure 12 shows the vertical mode functions of frequency 300 Hz at 10:32:30 GMT. The mode-filter outputs, shown in Fig. 13, reveal a very interesting mode-dependent variation pattern; one can see that modal arrival levels can vary up to 30 dB in 3 min, and also that modal arrivals alternately diminish. This is quite striking. Examining possible causes of this, given the good coverage of the WHOI VLA in the water column (see Fig. 12), there should not be a serious modal cross-talking defect in the mode filtering process, especially for the lowest three modes. Hence the modal variation seen in the filter outputs is not due to imperfect signal processing. Another possible explanation is variable modal excitation. A modal excitation calculation suggests that less than 3-dB variation should be expected, which cannot explain the 30-dB variations in modal arrival levels, not even if one placed considerable uncertainty on the modal excitation prediction. Invoking known mechanisms for mode coupling is also unreasonable because the fairly straight wavefront of the truncated internal-wave packet is only about 6 deg to the propagation path from the source to the receiver, which is smaller than the typical angles causing coupling (although the very largest waves measured in the SW06 area, 20-m amplitude, may cause coupling at 8 deg[22]). Thus, the modal arrivals are most consistent with time-variable of acoustic mode radiation from the internal-wave duct termination. As the internal-wave duct passed over the NRL300 source the relative source position in the duct changed, causing horizontal modal radiation pattern fluctuation in the manner illustrated by the examples.

V. DISCUSSIONS

The theoretical and numerical calculations clearly show that variable acoustic mode radiation from the termination of an internal-wave duct can be a significant source of spatial and temporal fluctuations of sound outside the duct. The spatial fluctuations result from the narrow angular extent of the radiation beams, for each mode, that emerge from the termination. One consequence is that the sound intensity in the areas not covered by radiation beams is reduced. The beams have temporal variability because passage of a truncated internal-wave duct over a fixed source causes the source position in the duct to vary in time, thus giving a time-dependence to the excitation of horizontal mode and thus to the emerging beams.

The real oceanic environment is often more complicated than the modeled environments used in this paper, and there are other acoustic effects that cause acoustic fluctuation patterns that are similar to the modal beam effect and that we shall be aware of. For instance, the effect of mode coupling on changing received modal intensity can be similar to the modal beam effect. To distinguish these two effects, one may need to carefully examine environmental data and determine which effect is more likely to occur in the environmental condition, as done here for the NRL300 signals. The following is another example. The calculations herein show that horizontal modal dispersion within the duct may cause multiple arrival of a mode (see the top two panels of Fig. 4 and the lower right panel of Fig. 7). However, in sound transmission data collected in the 1988 Hudson Canyon

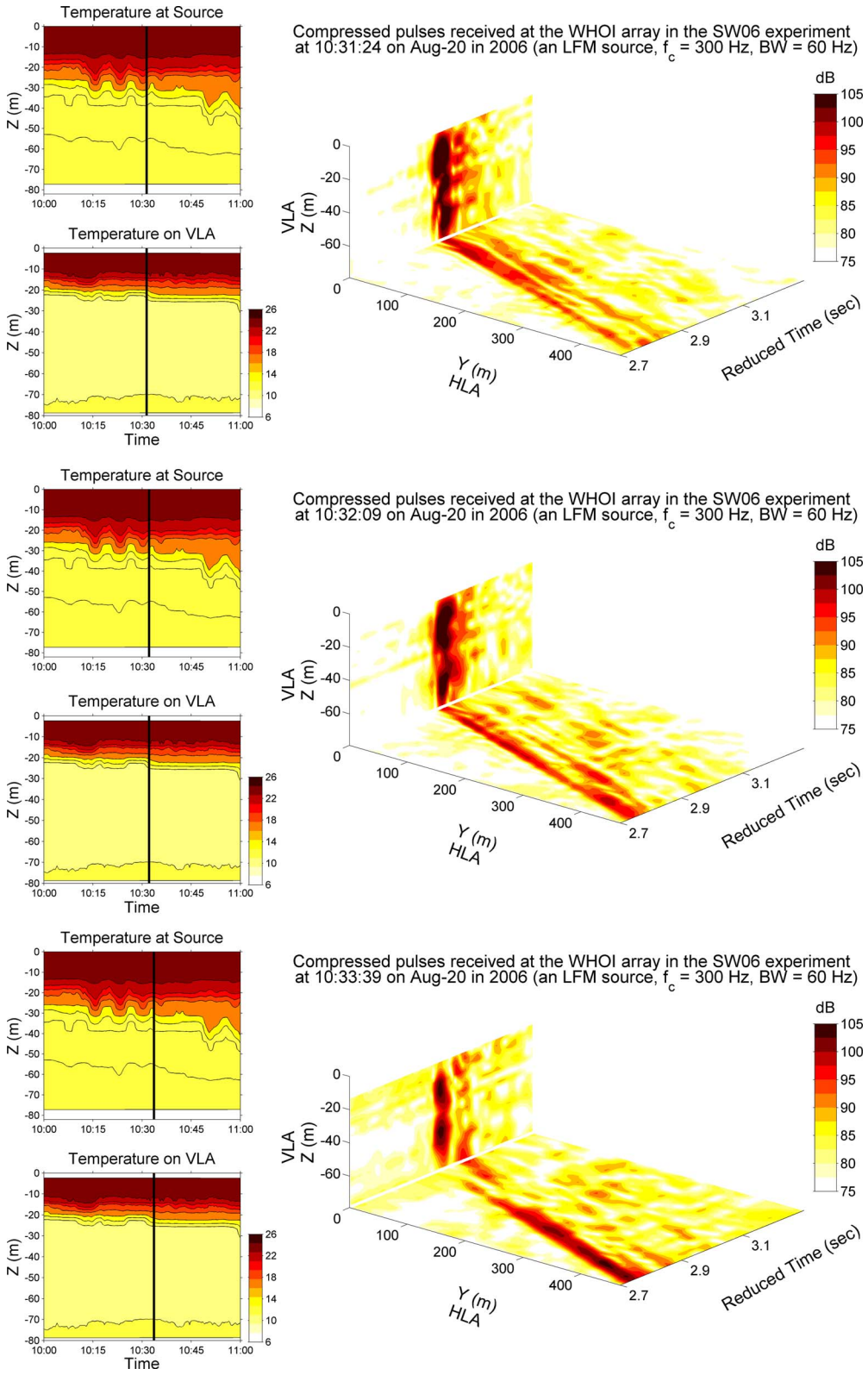


FIG. 10. (Color online) Compressed pulses of the NRL300 sweep signals received on the SW06 WHOI array. (Right) Pulse arrivals at three times are shown, from top to bottom. To the left of each pulse plot, temperature vs depth time series for the source and receiver positions are shown. The line indicates the condition at the time of transmission. At these times, the source was in the truncated internal-wave packet shown in Fig. 9.

Experiment,[23] Cederberg *et al.*[24] observed that a deep sub-bottom duct may produce two different modes with almost identical wavenumbers and with shapes that are similar in the water column. Separate arrival of the two modes may be mistaken for multiple arrival of the same mode, particularly when noise is present. This sub-bottom ducting effect presents an alternative explanation of the double (repeat) mode features.

During the SW06 experiment, other internal-wave ducting situations were observed, such as curved waves or frag-

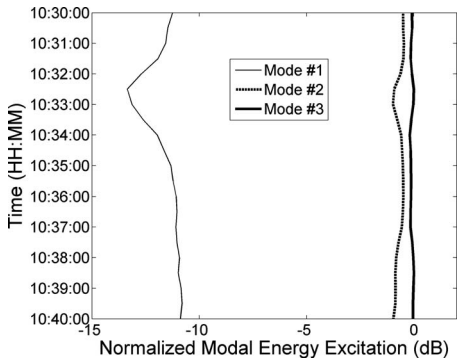


FIG. 11. Time series of theoretical estimates of the energy excitation of first three modes by the NRL300 source during the truncated internal-wave event shown in Fig. 9 on August 20.

mented waves. In the interesting curved wave situation, sound initially trapped in the duct may escape due to the change in the grazing angles of trapped sound. That is, horizontal modes that are sub-critical may become supercritical and escape. As for a fragmented internal-wave field, sound will encounter multiple processes of trapping and radiating, and the single truncation case reported in this paper may serve as a starting point toward understanding that multiple-scatter case. It is also worth mentioning that when an acoustic source is just outside and close to a truncated internal-wave duct, the significance of the mode radiation effect depends on how far the source is away from the duct. If it is close, a certain amount of energy may still be trapped in the duct after the sound refracts into the duct, and so at the termination one can still observe significant radiation effects.

The modal radiation may affect horizontal array beamforming. The simulations show that the termination acts like a sound projector, and horizontal beamforming may show incorrect sound source bearing directed toward the termination. Note that narrow beams are not plane waves, and standard beamforming with arrays larger than the beam width may be inconclusive. Also, because of the interference of the radiation patterns of different modes, the array coherent length in the radiation field may be much shorter than what would be expected at array locations far from internal waves

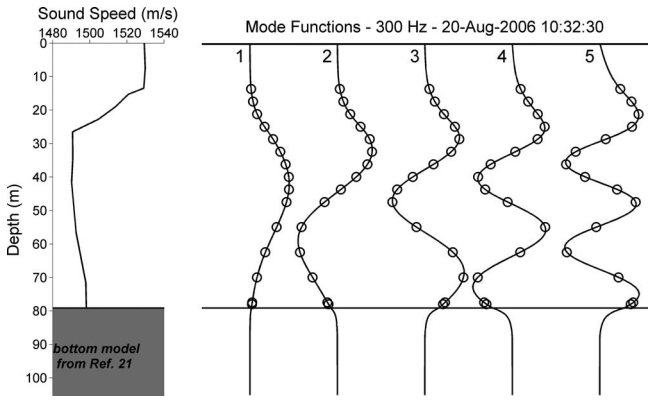


FIG. 12. Estimates of the mode functions of frequency 300 Hz at the location of WHOI VLA at 10:32:30 GMT on August 20 in the SW06 experiment. The *in situ* water-column sound speed profile and a bottom model derived from Ref. 21 are used for calculating the modes. The circles on the curves indicate the hydrophone depths on the VLA.

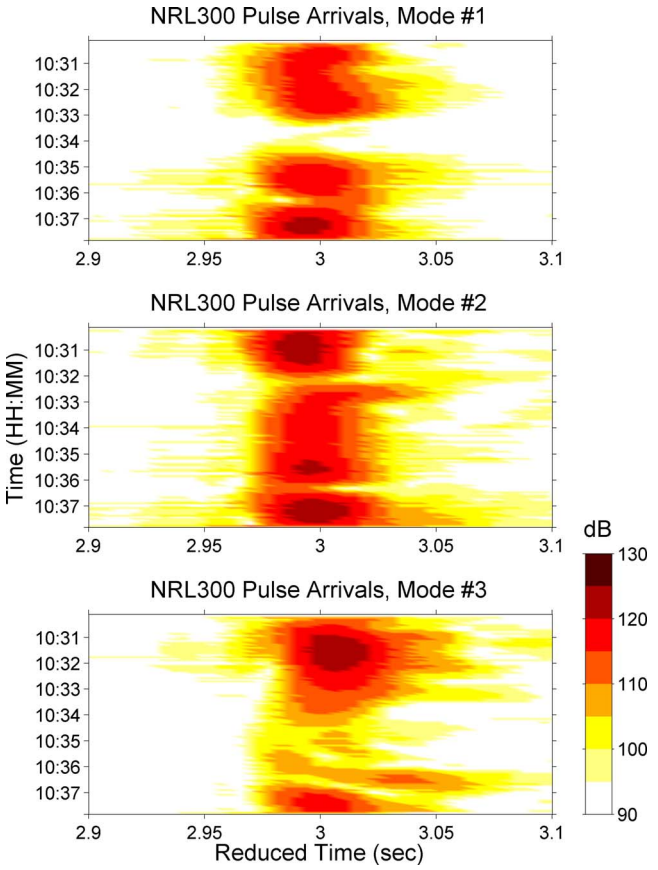


FIG. 13. (Color online) Broadband mode-filtered outputs of the received NRL300 pulses on the WHOI VLA during the truncated internal-wave event on August 20 in the SW06 experiment. The first three modes are shown, and a total of 110 pulses is analyzed. The mode intensities fluctuate in time.

(a typical value in the SW06 site is 15 wavelengths at 100 Hz, according to the measurement of Collis *et al.*[25] during quiescent periods. Further analysis of a larger portion of this data set shows higher values of 25–30 wavelengths during the first week of the experiment).

VI. CONCLUSIONS

The sound radiation from the termination of a truncated internal-wave duct has been studied with both analytical and numerical approaches. A 3D normal mode solution has been derived to describe the radiation field in a simplified truncated internal-wave duct model, and a computer code implementing 3D PE approximation is employed to investigate the radiation effects in a more realistic shallow-water ocean environment. Both of the analytical and numerical calculations predict anomalous sound radiation fields at locations far from the termination of a truncated internal-wave duct. The radiation patterns are unique to each mode and show strong spatial and temporal variability. The studies also show that the cause of this is a strong dependence on the position of an acoustic source in the duct, which changes in time for a moving wave and a moving or fixed source. Broadband simulations made in this paper reveal situations where one or more modes are completely absent on a fixed VLA because the strongly localized modal radiation beams cannot just

reach the fixed array and mode multipath arising from dispersive propagation within the horizontal waveguide.

ACKNOWLEDGMENTS

The 3D PE program used in this paper was derived from a trial version written by Dr. John Colosi of the Naval Postgraduate School, CA, USA, and the authors acknowledge his valuable efforts to the original code. The authors must also thank all of SW06 participants for their efforts to make this large multi-disciplinary shallow-water ocean experiment successful; heartfelt thanks must be offered to the acoustics and physical oceanography program managers from the Office of Naval Research (ONR), Ellen Livingston and Terry Paluszkiewicz, for their support. This work was supported under ONR Grant No. N00014-05-1-0482 and the ONR postdoctoral fellowship award, Grant No. N00014-08-1-0204.

[1]B. G. Katsnel'son and S. A. Pereselkov, "Low-frequency horizontal acoustic refraction caused by internal wave solitons in a shallow sea," Acoust. Phys. **46**, 684–691 (2000).
[2]R. Burridge and H. Weinberg, "Horizontal rays and vertical modes," in *Wave Propagation and Underwater Acoustics*, edited by J. B. Keller and J. S. Papadakis (Springer, Berlin, 1977), pp. 86–152.
[3]R. Oba and S. Finette, "Acoustic propagation through anisotropic internal wave fields: Transmission loss, cross-range coherence, and horizontal refraction," J. Acoust. Soc. Am. **111**, 769–784 (2002).
[4]D. Lee and M. H. Schultz, *Numerical Ocean Acoustic Propagation in Three Dimensions* (World Scientific, Singapore, 1995).
[5]S. Finette and R. Oba, "Horizontal array beamforming in an azimuthally anisotropic internal wave field," J. Acoust. Soc. Am. **114**, 131–144 (2003).
[6]M. Badiey, B. G. Katsnelson, J. F. Lynch, S. Pereselkov, and W. L. Siegmann, "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **117**, 613–625 (2005).
[7]S. D. Franks, M. Badiey, J. F. Lynch, and W. L. Siegmann, "Experimental evidence of three-dimensional acoustic propagation caused by nonlinear internal waves," J. Acoust. Soc. Am. **118**, 723–734 (2005).
[8]D. B. Reeder, T. F. Duda, and B. Ma, "Short-range acoustic propagation variability on a shelf area with strong nonlinear internal waves," in *Oceans'08 (Kobe) Conference Proceedings* (IEEE, New York, 2008).
[9]C. R. Jackson, *An Atlas of Internal Solitary-Like Waves and Their Properties*, 2nd ed. (Global Ocean Associates, Alexandria, VA, 2004) (available at http://www.internalwaveatlas.com/Atlas2_index.html (Last viewed 6/17/2009).
[10]T. F. Duda, "Initial results from a Cartesian three-dimensional parabolic equation acoustical propagation code," Technical Report No. WHOI-2006-041, Woods Hole Oceanographic Institution, Woods Hole, MA, 2006.
[11]D. J. Tang, J. N. Moum, J. F. Lynch, P. Abbot, R. Chapman, P. H. Dahl, T. F. Duda, G. Gawarkiewicz, S. Glenn, J. A. Goff, H. Graber, J. Kemp, A. Maffei, J. D. Nash, and A. Newhall, "Shallow Water '06, A joint acoustic propagation/nonlinear internal wave physics experiment," Oceanogr. **20**, 156–167 (2007).
[12]J. C. Preisig and T. F. Duda, "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269 (1997).
[13]J. Zhou, X. Zhang, and P. H. Rogers, "Resonant interaction of sound wave with internal solitons in the coastal zone," J. Acoust. Soc. Am. **90**, 2042–2054 (1991).
[14]H. Levine and J. Schwinger, "On the radiation of sound from an unflanged circular pipe," J. Acoust. Soc. Am. **73**, 383–405 (1948).
[15]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (AIP, Woodbury, NY, 1994).
[16]G. V. Frisk, *Ocean and Seabed Acoustic: A Theory of Wave Propagation* (Prentice-Hall, Englewood Cliffs, NJ, 1994).
[17]R. H. Hardin and F. D. Tappert, "Applications of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations," SIAM Rev. **15**, 423 (1973).
[18]D. J. Thomson and N. R. Chapman, "A wide-angle split-step algorithm for the parabolic equation," J. Acoust. Soc. Am. **74**, 1848–1854 (1983).
[19]T. F. Duda, "Analysis of finite-duration wide-band frequency sweep signals for ocean tomography," IEEE J. Ocean. Eng. **18**, 87–94 (1993).
[20]M. B. Porter, "The KRAKEN normal mode program," SACLANTECN Memorandum No. SM-245 (SACLANT Undersea Research Centre, La Spezia, Italy, 1991).
[21]Y.-M. Jiang, N. R. Chapman, and M. Badiey, "Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey self by inverting air gun data," J. Acoust. Soc. Am. **121**, 1879–1894 (2007).
[22]T. F. Duda, "Examining the validity of approximations to fully three-dimensional shallow-water acoustic propagation through nonlinear gravity waves," in *Oceans'07 (Aberdeen) Conference Proceedings* (IEEE, New York, 2007).
[23]W. M. Carey, J. Doutt, R. B. Evans, and L. M. Dillman, "Shallow-water sound transmission measurements on the New Jersey continental shelf," IEEE J. Ocean. Eng. **20**, 321–336 (1995).
[24]R. J. Cederberg, W. M. Carey, and W. L. Siegmann, "Modal analysis of geoacoustic influences on shallow-water propagation," IEEE J. Ocean. Eng. **22**, 237–244 (1997).
[25]J. M. Collis, T. F. Duda, J. F. Lynch, and H. A. Deferrari, "Observed limiting cases of horizontal field coherence and array performance in a time-varying internal wave field," J. Acoust. Soc. Am. **124**, EL97–EL103 (2008).

Wave scattering from encapsulated microbubbles subject to high-frequency ultrasound: Contribution of higher-order scattering modes

Jiusheng Chen
Department of Mechanical Engineering, University of Colorado, Boulder, Colorado 80309-0427

Kendall S. Hunter
Center for Bioengineering, University of Colorado, Anschutz Medical Campus, Aurora, Colorado 80045

Robin Shandas[a)]
Center for Bioengineering, University of Colorado, Anschutz Medical Campus, Aurora, Colorado 80045 and Department of Mechanical Engineering, University of Colorado, Boulder, Colorado 80309-0427

(Received 13 October 2008; revised 15 July 2009; accepted 16 July 2009)

The theoretical understanding of encapsulated microbubble response to high-frequency ultrasound (HFUS) excitation is still limited although some novel experimental HFUS contrast imaging techniques have been well developed. In this paper, the higher-order modal (HOM) contributions to the scattered field are studied for such microbubbles driven by 1–100 MHz ultrasound. An exact solution of all small-amplitude vibrational modes of a single encapsulated microbubble in water is given by the wave scattering theory (WST) method and compared to results obtained from Church's Rayleigh–Plesset-like model for the small-amplitude radial oscillation of a microbubble in an incompressible fluid. From numerical results, we show that the HOM field contribution is significant for scattering properties from individual Nycomed microbubbles with normalized frequency ≥ 0.2. It is also shown that the multiple scattering is strengthened for monodispersed Definity® microbubbles of 3 μm radius at frequencies >40 MHz. However, comparisons between the authors' analyses and known experimental data for polydispersed Definity® microbubbles indicate that the HOM contributions are insignificant in attenuation estimation at frequencies <50 MHz. In conclusion, the WST model analysis suggests that HOM scattering is an important consideration for single bubbles but may be less critical in the modeling of polydispersed Definity® bubbles at high frequencies. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203917]

PACS number(s): 43.35.Bf, 43.35.Ei, 43.20.Fn [CCC] Pages: 1766–1775

I. INTRODUCTION

Low-frequency ultrasound imaging systems clearly lack the spatial resolution to examine the microcirculation, although details obtained at this level may be important to both clinical and basic medical science, for example, in the hypertensive microcirculation (James *et al.*, 2006). Recently, microbubble contrast agents have been successfully extended to a variety of high-frequency ultrasound (HFUS) imaging systems and have made possible non-invasive slow blood flow measurements and targeted molecular imaging with high contrast in the microcirculation (Lanza *et al.*, 1997; Goertz *et al.*, 2007b; Goessling *et al.*, 2007; Needles *et al.*, 2008; Yeh *et al.*, 2008). In support of these imaging methods, the acoustic properties of microbubbles at higher frequencies have also been intensively studied through experimental observations of attenuation and nonlinear scattering activities (e.g., Moran *et al.*, 2002; Goertz *et al.*, 2006; Goertz *et al.*, 2007a; Cheung *et al.*, 2008). However, the understanding of microbubble dynamics at high frequencies is still incomplete, especially with regard to knowledge of resonant bubble sizes, attenuation prediction, and nonlinear scattering mechanisms. Similarly, the acoustic radiation forces on microbubbles subject to HFUS have not been investigated although some targeted applications have already been reported (Lanza *et al.*, 1997; Rychak *et al.*, 2007).

Multiple models have been employed for studying the dynamics of encapsulated microbubble contrast agents in the conventional frequency range (1–10 MHz). The most common models rely on Rayleigh–Plesset-like (RPL) ordinary differential equations and account for only bubble radial pulsation (de Jong *et al.*, 1994; Church, 1995; Hoff, 2001). Analytical solutions for both linear and nonlinear scattering from individual microbubbles encapsulated by a shell of Kelvin–Voigt viscoelastic solid are provided in such RPL models. Multiple scattering of bubbles was thus achieved in sound dispersion and attenuation (Church, 1995; Hoff, 2001) by Foldy's theory (Foldy, 1945). The RPL model has been extensively developed for various contrast microbubbles with different encapsulations and surrounding liquids (Sarkar *et al.*, 2005; Doinikov and Dayton, 2007); it has also been proven effective in predicting the sound properties of ultra-

[a)]Author to whom correspondence should be addressed. Also at The Children's Hospital, 13123 East 16th Ave., Cardiology B-100, Aurora, CO 80045. Electronic mail: robin.shandas@uchsc.edu

sound contrast agents in conventional ultrasound imaging (Frinking and de Jong, 1998; Coussios *et al.*, 2004). However, RPL models only consider the radial pulsation of bubbles and usually assume a uniform internal gas pressure, i.e., no inertial effect of the gas core. These approximations may not be appropriate for microbubbles subject to HFUS field. At higher frequencies, the wavelength decreases to the order of magnitude of some particle sizes and falls into an intermediate wavelength regime in which the higher-order contracting movements become important and anisotropic scattering begins to appear (McClements, 1996; Ye, 1996). Moreover, the shell properties for Definity® bubbles estimated by modeling experimental attenuation measurements with an RPL-based method differed between the low- and high-frequency regimes, as reported by Goertz *et al.* (2007a). This would imply that traditional modeling methods need to be improved for microbubble oscillation at high frequencies, with the higher-order modal (HOM) scattering contribution requiring particular attention.

The wave scattering theory (WST) has also been employed in the past decade in studies of the dynamics of encapsulated microbubble surrounded by water and oscillating in the linear regime (Ye, 1996; Allen *et al.*, 2001; Hu *et al.*, 2004; Chen and Zhu, 2005). The WST method was originally developed for object detection by the underwater acoustics industry and has been intensively used to predict acoustic scattering from both spherical and cylindrical scatterers immersed in various surrounding fluids over the past two decades (Ayres *et al.*, 1987; Gaunaurd and Werby, 1991; Hasheminejad and Safari, 2005; Mitri and Fellah, 2006). While the RPL equation is derived from a fluid velocity potential, through simplification it is reduced to a simple dynamic mass-spring system for the radial oscillation of the bubble without any shape oscillations. In contrast, the wave-based interaction between ultrasound and bubble in the WST method is retained in its description of multi-modal velocity potentials of the acoustic waves propagating through the surrounding fluid, the shell layer, and the internal gas. The first modal component of this velocity potential corresponds to the radial oscillation, and each additional mode represents a non-spherical oscillation mode of unique shape complexity. Under the quasi-equilibrium approximation and the small-amplitude assumption, all the oscillation modes are orthogonal and the total scattering of the bubble can be found from their linear summation. As a result, the WST method reveals directionality of the scattering. Ye (1996) first used the WST method to model the contrast microbubble as an individual elastic shell in inviscid water and qualitatively compared his results with Church's viscoelastic shell bubble model in incompressible, viscous water. He noted that the scattering by Albunex® bubbles from WST prediction could be highly anisotropic at frequencies above resonance. Allen *et al.* (2001) studied the reflectivity and scattering directivity of microbubble at high driving frequencies with the WST model and proposed the potential applicability of "shell" Lamb waves, which propagate as symmetric (flexural) and antisymmetric (extensional) modes of deformation in a curved plate idealized for shell. The double-layered shell, which is typically used for encapsulated drug delivery, has never been explored by the RPL model but was investigated using the WST model by Hu *et al.* (2004). The employment of WST provides an alternative method to predict the acoustic scattering properties of contrast microbubbles. However, previous WST studies provided only initial theoretical details, and the HOM components were actually found to be negligible for microbubbles in conventional ultrasound imaging. No comparisons between the WST and RPL models have been performed, and the multiple scattering of contrast microbubbles due to anisotropic effects has not been predicted with the WST model. The significance of the HOM scattering contribution in microbubbles under higher frequency excitation is still pending.

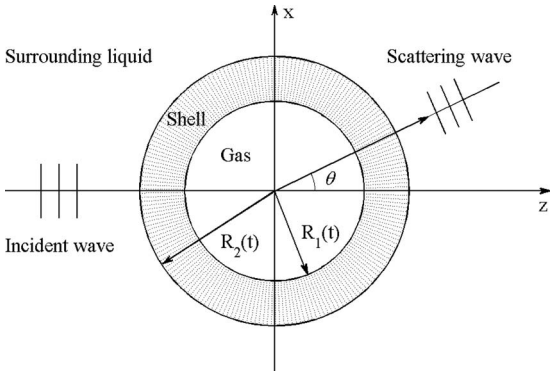


FIG. 1. Schematic sketch of the model.

The aim of this work is to assess the HOM contributions to scattering at both low- and high-frequency excitations as modeled by the WST model and to compare them to a dilatation-only model. This is accomplished with the following steps. First, the modeling of the same encapsulated microbubble by both WST and RPL methods is outlined. Second, the HOM contributions to the total scattering of the microbubble are compared within the WST method, and successively, the acoustic scattering properties of the microbubbles such as scattering cross section, attenuation, and acoustic radiation forces are simulated by both models and are compared to reveal the significance of HOM contributions. Finally, a discussion follows for physical interpretations of differences between the two models and the prospects of the WST method in modeling of microbubble response at high frequencies.

II. METHOD

The encapsulated microbubble model consists of a spherical shell located in a spherical coordinate system, shown in Fig. 1. An incident plane ultrasound wave excites the bubble in the direction $\theta=0$. The gas enclosed by the shell is assumed to be air, and the surrounding fluid is water. The bubble geometry is given by inner radius R_1, outer radius R_2, and shell thickness d. The shell is considered to be a thin monolayer of Kelvin–Voigt-type viscoelastic solid. The water is assumed to be inviscid and compressible for the wave scattering analyses. The incident ultrasound pressure is sufficiently small such that the vibration of the bubble can be approximated as a spherical oscillation within the linear regime. For simplicity, the damping due to liquid viscosity and

thermal conductivity, as well as surface tension effects, are ignored because they have been proven to be of little importance for micron-sized encapsulated microbubbles in the frequency range of 1–100 MHz by previous studies (Church, 1995; Hoff, 2001).

A. Outline of WST on encapsulated microbubbles

Ye's (1996) Albunex® bubble model consisted of a spherical gas-filled bubble coated with a layer of elastic solid; an exact solution for its wall motion was obtained based on the literature of Hasegawa *et al.* (1993). The vector wave potentials in the work of Hasegawa *et al.* are expanded with factors for each eigenterm using a differential form of the associated Legendre function. In contrast, other contributions (Ayres and Gaunaurd, 1987; Ayres *et al.*, 1987) expanded the vector potentials as sums of modal series without any differential forms of Legendre function. This so-called "Debye potential" expansion method has been widely used, and it has proved to be more numerically robust for modeling contrast microbubbles (Allen *et al.*, 2001). Allen *et al.* (2001) extended the calculations from elastic to viscoelastic shell models by directly applying the viscoelastic material parameters from Ayres *et al.* (1987) into the scattering matrix for an elastic shell. Chen and Zhu (2005) gave a rigorous derivation for the matrix elements of a Kelvin–Voigt viscoelastic shell model and proved that they are the same as those for the elastic shell model of Ayres *et al.* (1987). In this work, Chen and Zhu's results are quoted below to outline key steps and variables.

Let the potentials of incident and scattering waves in water be noted as Φ_i and Φ_s, respectively, and the potential inside the shell be Φ_a. Let Φ_2, Ψ_2 denote the longitudinal and shear wave potentials propagating in the solid layer, which satisfy the Helmholtz equations,

$$(\nabla^2 + k_d^2)\Phi_2 = 0, \quad (\nabla^2 + k_s^2)\Psi_2 = 0, \tag{1}$$

where k_d and k_s are acoustic wave numbers in the solid shell layer and given as complex forms by Ayres and Gaunaurd (1987). The Lamé parameters of various shell materials are given in the complex forms by $\lambda = \lambda_e + i\omega\lambda_v$ and $\mu = G_s + i\omega\mu_s$. The real and imagery parts represent elastic and viscous properties of the shell, respectively.

Expanding the acoustic field in each of the three media in terms of norm modes and taking the symmetry into account, it yields

$$\Phi_i = \Phi_0 \sum_{n=0}^{\infty} (2n+1)(i)^n j_n(k_1 r) P_n(\cos\theta) e^{-i\omega t}, \tag{2a}$$

$$\Phi_s = \Phi_0 \sum_{n=0}^{\infty} (2n+1)(i)^n a_n h_n^{(1)}(k_1 r) P_n(\cos\theta) e^{-i\omega t}, \tag{2b}$$

$$\Phi_a = \Phi_0 \sum_{n=0}^{\infty} (2n+1)(i)^n f_n j_n(k_3 r) P_n(\cos\theta) e^{-i\omega t} \tag{2c}$$

$$\Phi_2 = \Phi_0 \sum_{n=0}^{\infty} (2n+1)(i)^n [b_n j_n(k_d r) + c_n n_n(k_d r)] \times P_n(\cos\theta) e^{-i\omega t}, \tag{2d}$$

$$\Psi_2 = \Phi_0 \sum_{n=0}^{\infty} (2n+1)(i)^n [d_n j_n(k_s r) + e_n n_n(k_s r)] \times P_n(\cos\theta) e^{-i\omega t}, \tag{2e}$$

where k_1 and k_3 are the wave numbers in the water and in the gas, respectively, $P_n(\cdot)$ is the Legendre polynomial of order n, $j_n(\cdot)$ is the spherical Bessel function of order n, $n_n(\cdot)$ is the spherical Neumann function of order n, and $h_n^{(1)}(\cdot)$ is the spherical Hankel function of the first kind. Finally, the coefficients $a_n, b_n, c_n, d_n, e_n, f_n$ are unknowns determined by the boundary conditions noted below; a_n specifically is the scattering coefficient. Under this framework, the determination of the sound scattering from an encapsulated bubble comes down to the evaluation of the a_n.

Solution begins by applying the following approximate boundary conditions at the two spherical interfaces between the media: (1) the continuity of normal displacement u_r; (2) the continuity of normal stress τ_{rr}; (3) the tangential stress $\tau_{r\theta}$ must be zero; six equations are obtained:

At the water-shell interface ($r = R_2$),

$$u_r^1 + u_r^4 = u_r^2, \quad \tau_{rr}^1 + \tau_{rr}^4 = \tau_{rr}^2, \quad \tau_{r\theta}^2 = 0. \tag{3a}$$

At the shell-gas interface ($r = R_1$),

$$u_r^2 = u_r^3, \quad \tau_{rr}^2 = \tau_{rr}^3, \quad \tau_{r\theta}^2 = 0. \tag{3b}$$

Here, the superscripts denote the following: 1, incident waves; 2, waves propagating in the layer; 3, waves propagating in the inner air; 4, scattering waves.

Substituting the velocities and displacements in Eq. (6) of Ayres and Gaunaurd (1987) into the six boundary equations, the corresponding matrix equation is obtained,

$$\begin{bmatrix} 0 & \alpha_{12} & \alpha_{13} & \alpha_{14} & \alpha_{15} & \alpha_{16} \\ 0 & \alpha_{22} & \alpha_{23} & \alpha_{24} & \alpha_{25} & \alpha_{26} \\ 0 & \alpha_{32} & \alpha_{33} & \alpha_{34} & \alpha_{35} & 0 \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} & \alpha_{45} & 0 \\ \alpha_{51} & \alpha_{52} & \alpha_{53} & \alpha_{54} & \alpha_{55} & 0 \\ 0 & \alpha_{62} & \alpha_{63} & \alpha_{64} & \alpha_{65} & 0 \end{bmatrix} \begin{bmatrix} a_n \\ b_n \\ c_n \\ d_n \\ e_n \\ f_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \alpha_4 \\ \alpha_5 \\ 0 \end{bmatrix}. \tag{4}$$

From this matrix equation, we can solve for the scattering coefficients (a_n) by Cramer's rule. The elements of the matrices $\alpha_{ij}(i,j = 1-6)$ are also determined by the matrix equation, and their expressions can be found from appendices of previous literature (Ayres *et al.*, 1987; Chen and Zhu, 2005).

Using asymptotic expansions of Hankel functions, the scattering from an encapsulated microbubble in the far field can be expressed in terms of a modal series as follows:

$$f(\theta) = \sum_{n=0}^{\infty} f_n(\theta) = \frac{1}{ik_1} \sum_{n=0}^{\infty} (2n+1) a_n P_n(\cos\theta). \tag{5}$$

Here, $f(\theta)$ is the scattering (form) function. Each term $f_n(\theta)$ represents the nth partial wave of the scattering. Generally,

only the monopole scattering, i.e., the first term f_0, needs to be considered for Rayleigh scatterers ($k_1R_2 \ll 1$). It is also stated that the extinction cross section and the sound reflectivity of the bubble are proportional to the imaginary part of forward scattering $f(0)$ and the amplitude of backward scattering $f(\pi)$, respectively (Ye, 1996). The scattering function can also be linked to the scattering cross section,

$$\sigma_s(\theta) = 4\pi|f(\theta)|^2. \tag{6}$$

The reduced scattering cross section $\sigma_r(\theta)$ is defined here by $\sigma_s(\theta)$ multiplying a factor of $1/4\pi R_2^2$ for later convenience. Here, $\sigma_r(\theta)$ is noted as the "total" reduced scattering cross section since it is the infinite sum of all contributing scattering modes, according to Eqs. (5) and (6); in practice, however, the sum is truncated to an approximate representation. In order to study the modal contributions, we also use the symbol $\sigma_r^{(m)}(\theta)$ to represent the sum truncated to the first $m+1$ modes, i.e., mode 0 to mode m; thus, the contribution from mode m to the total scattering cross section $\sigma_r(\theta)$ in the WST model is $\sigma_r^{(0)}(\theta)$ for $m=0$ or $\sigma_r^{(m)}(\theta)-\sigma_r^{(m-1)}(\theta)$ for $m>0$. It is important to note that for a mode m, the sum still contains lower modal contributions; thus, the contribution from a single mode is coupled to the modes below it.

B. RPL equation for encapsulated microbubbles

The linear harmonic oscillation of an encapsulated microbubble can be written in a simple form using the RPL equation (Church, 1995; Hoff, 2001),

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = p_a(\rho_s R_{10}^2 \alpha)^{-1} \sin\ \omega t, \tag{7}$$

where x is the normalized small displacement to the first order of the inner shell wall, defined by $R_1 = R_{10}(1+x)$, $x \ll 1$ where R_{10} represents the initial inner shell radius; ω, ρ_s, and p_a are the angular driving frequency, the density of shell, and the acoustic pressure amplitude, respectively. The scattering cross section σ_s for our model can be given as

$$\sigma_s(R_{10},\omega) = 4\pi R_{10}^2 \frac{\rho_l^2}{\rho_s^2\alpha^2} \frac{\omega^4}{(\omega^2-\omega_0^2)^2 + (2\beta\omega)^2}, \tag{8}$$

where

$$\omega_0^2 = \frac{1}{\rho_s R_{10}^2 \alpha}\left[3\kappa p_0 + \frac{4G_s(R_{20}^3 - R_{10}^3)}{R_{20}^3}\right],$$

$$\alpha = \left[1 + \left(\frac{\rho_l - \rho_s}{\rho_s}\right)\frac{R_{10}}{R_{20}}\right],$$

$$\beta = \beta_{\rm ac} + \beta_{\rm sh}, \quad \beta_{\rm ac} = \frac{\rho_l}{\rho_s\alpha}\frac{\omega^2 R_{10}}{2c_0}\left[1 + \frac{\rho_l^2}{\rho_s^2\alpha^2}\frac{\omega^2 R_{10}^2}{c_0^2}\right]^{-1},$$

$$\beta_{\rm sh} = \frac{2\mu_s(R_{20}^3 - R_{10}^3)}{\alpha\rho_s R_{10}^2 R_{20}^3}.$$

Here, only the damping terms due to acoustic radiation $\beta_{\rm ac}$ and shell viscosity $\beta_{\rm sh}$ are considered. The above-listed expressions are simplified from a corresponding Church's RPL model (Church, 1995; Khismatullin, 2004). The compressibility of surrounding liquid is neglected here but can be implemented by some other RPL models in the low-frequency limit (Trilling, 1952; Keller and Miksis, 1980). Finally, the RPL model neglects the surface tension on both interfaces and the viscous effects of the surrounding liquid in order to compare with WST, which as noted above does not include these effects.

TABLE I. Parameter values for aqueous suspensions of Albunex®, Definity®, and Nycomed bubbles. Listed values are given by Church (1995) except those noted in the footnotes.

		(Value)			
	Symbol	Albunex®	Definity®	Nycomed	Unit
Shell	G_s	(88.8)	(190)[c,d]	(12)[e,f]	MPa
	μ_s	(1.77)	(0.07)[c,d]	⋯	Pa s
	λ_e	(6.1×10^4)[b]	(6.1×10^4)[b]	(6.1×10^4)[b]	MPa
	λ_v	(50)[b]	(50)[b]	⋯	Pa s
	$\omega\lambda_v$	⋯	⋯	(0.5×10^6)[a]	Pa
	$\omega\mu_s$	⋯	⋯	(0.1×10^5)[a]	Pa
	d	(15)	(0.9)[d]	$(0.05a)$[a,e]	nm
	ρ_s	(1100)	(1100)	(1100)	kg m^{-3}
Liquid	ρ_L	(1000)	(1000)	(1000)	kg m^{-3}
Gas	γ	(1.4)	(1.06)[c]	(1.1)[f]	⋯
	p_0	(0.1013)	(0.1013)	(0.1013)	MPa

[a]Allen *et al.*, 2001.
[b]Chen and Zhu, 2005.
[c]Cheung *et al.*, 2008.
[d]Goertz *et al.*, 2007a.
[e]Hoff, 2001.
[f]Khismatullin, 2004.

III. RESULTS

The simulation results of both WST and RPL models were performed using MATLAB® software (Version 14, The Mathworks Inc., and Natick, MA) and are compared in the following sections. Three types of contrast bubbles are simulated. We primarily examine Albunex® microbubbles, which have been best characterized in the literature, especially with regard to their shell material. In order to compare the theoretical models with the existing measurements, lipid-shelled Definity® microbubbles are also simulated. Finally, the polymer shell of Nycomed microbubbles is also simulated for discussion since they have a fixed thickness-to-radius ratio in contrast to the fixed thickness seen in Albunex® bubbles. All physical parameters used in the simulations are selected from previous works and provided in Table I if not specified elsewhere. In addition, the Lamé first parameters λ_e and λ_v have little influence on bubble scattering properties, and consistent values of 6.1×10^4 MPa and 50 Pa s are selected here (Chen and Zhu, 2005).

A. Scattering cross section: Modal contributions

The contributions of the first three modes to the reduced scattering cross section are examined by the term defined just below Eq. (6) and shown in Fig. 2. A single Albunex® bubble with radius of 2 μm is selected for investigation. Strong resonant peaks are observed in the figure at about 10, 56, and

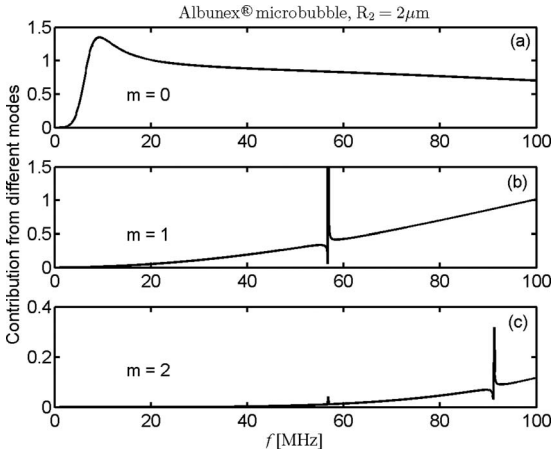


FIG. 2. Contribution of first three modes of WST model as a function of driving frequency. The Albunex® bubble is 2 μm in radius.

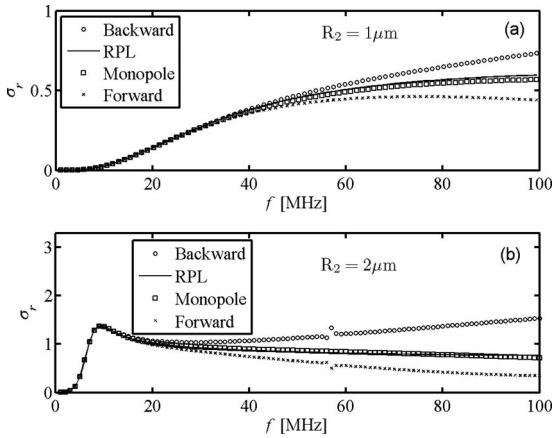


FIG. 3. Reduced scattering cross sections of Albunex® bubbles with radius (a) 1 μm and (b) 2 μm.

90 MHz corresponding to these modes; the mode 2 plot in Fig. 2(c) has a notable peak at the resonant frequency of mode 1 due to coupling. The figure also shows that mode 0, i.e., monopole scattering in Fig. 2(a), is far greater in magnitude than modes 1 and 2 and dominates the total scattering at low frequencies (<20 MHz). The contribution of mode 1 in Fig. 2(b) first becomes significant at 20 MHz and increases rapidly with the driving frequency, exceeding that of mode 0 above 80 MHz. At 100 MHz, modes 0 and 1 contribute 0.7 and 1.0 to the total reduced scattering cross section of about 1.9. The contribution of the third mode ($m=2$) in Fig. 2(c) becomes important at 60 MHz and is expected to be more significant than the first two modes at frequencies higher than 100 MHz. It is verified here that the higher modes contribute greatly to the total scattering of Albunex® bubbles at high frequencies. Moreover, further calculations for Definity® bubbles with radius between 1 and 6 μm indicate that, depending on bubble size and shell properties, the HOM can contribute greatly to the total scattering in the frequency range under 100 MHz. Following previous work, we explored multiple truncations of the WST modal sums; this analysis suggested that the first 30 modes are sufficient to approximately represent the total scattering at frequencies <100 MHz (error in scattering amplitude <0.1%). The HOM contribution can be easily seen from the differences between monopole (mode 0) and total (all modes) scattering of WST model.

B. Scattering cross section: Directionality

The reduced scattering cross sections of the Albunex® bubbles with radius either larger or less than the critical radius $R_c=1.46$ μm (Khismatullin, 2004) are shown in Fig. 3. To further understand the differences between the methods, we compare four terms of acoustic scattering: the backward scattering, the forward scattering, and the monopole scattering component from the WST model and the scattering from RPL model. It is apparent for a single Albunex® bubble with radius 1 μm in Fig. 3(a) that the four curves overlap at driving frequencies below 20 MHz, but significant differences begin to appear above 30 MHz. These differences in the magnitudes of reduced backward scattering and forward scattering clearly indicate directivity in the scattered field. Although the monopole is similar to the RPL model, it is obvious that its scattering also exhibits differences at high frequencies. From strong to weak, the four terms are in the order by backward scattering from WST > scattering from RPL > monopole scattering from WST > forward scattering from WST. For the large bubbles with 2 μm in radius, the resonances appear in Fig. 3(b) and act as boundaries to separate the low- and high-frequency ranges. Identical profiles are obtained for the four terms in the low-frequency range <13 MHz, but significant differences between them occur above resonance. Compared with Fig. 3(a), the scattering directivity in Fig. 3(b) is strengthened, and the additional peaks at higher frequencies begin to appear in both backward and forward scattering. These peaks in scattering cross sections are associated with various shell Lamb wave modes. Therefore, the results not only indicate that the monopole scattering component is dominant for the microbubble scattering at low driving frequencies, as was proved by Ye (1996), but also demonstrate that shell Lamb waves exist for Albunex® microbubbles at high frequencies, a finding of Allen *et al.* (2001). In addition, our results show that the RPL and WST models are identical for low ultrasound frequencies, but the HOM contribute greatly to the total at high frequencies. Further calculations show that the microbubbles with hard shells have more violent resonance peaks, which suggests the existence of stronger Lamb waves and greater HOM contributions at high frequencies.

C. Attenuation

The multiple scattering properties of contrast microbubbles are usually studied through the attenuation and dispersion of sound in suspensions. However, attenuation measurements at HFUS (>10 MHz) are limited. In order to compare our model predictions against prior experimental studies (Goertz *et al.*, 2007a), the Definity® bubbles were simulated as follows: The effective wave numbers in the microbubble suspensions following in the RPL and WST models were calculated by the multiple scattering theories of Foldy (1945) and Waterman and Truell (1961), respectively. Figure 4 is the result for monodispersed Definity® bubbles.

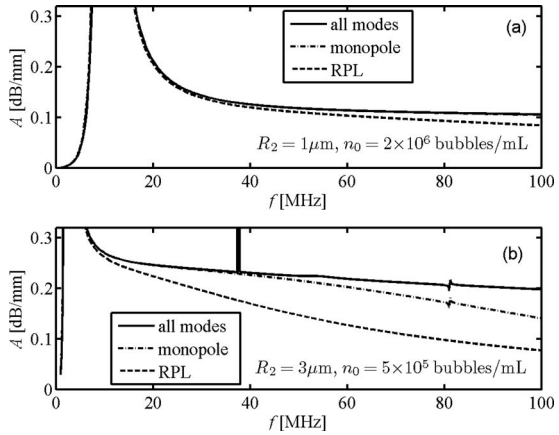


FIG. 4. Attenuation coefficients A in Definity® microbubble suspensions. The bubbles are identical in radii of (a) 1 μm and (b) 3 μm. The bubble concentrations are (a) 2×10^6 and (b) 5×10^5 bubbles/ml.

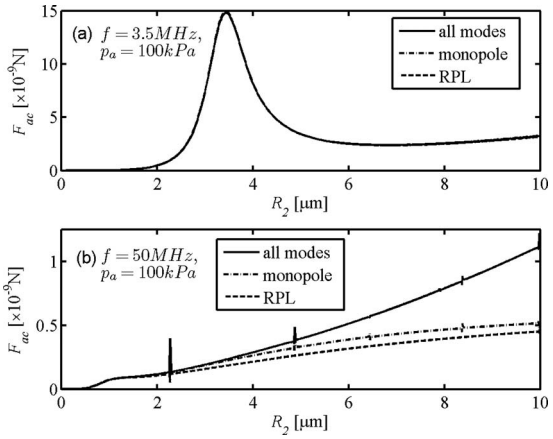


FIG. 6. Primary acoustic radiation forces on the Albunex® bubble in water as a function of radius under a driving ultrasound pulse wave with peak negative pressure p_a=100 kPa and central frequencies of (a) f=3.5 MHz and (b) f=50 MHz.

We see that the predicted attenuation from WST model is higher than that from the RPL model at frequencies above resonances; this difference becomes more significant as frequency increases, and larger differences were obtained for the larger bubble. Furthermore, the HOM contributions to attenuation in the WST model appear insignificant for smaller bubbles [1 μm in radius, Fig. 4(a)] but important for larger bubbles [3 μm in radius, Fig. 4(b)] at high frequencies $>$40 MHz; anyhow, both are less important than the differences between the attenuation due to isotropic scattering of the two models. Moreover, attenuation peaks similar to those seen in Fig. 3(b) appear at higher frequencies.

Figure 4 refers to monodispersed bubbles; however, in practice such sharp resonance peaks likely do not exist since the microbubbles have a broad size distribution ranging from $<$1 μm to 10 μm. Goertz *et al.* (2007a) measured the attenuation of polydisperse solutions of Definity® bubbles over a broad frequency range (2–50 MHz); these results are compared with corresponding attenuation values computed from our WST and RPL models and shown in Fig. 5. We specifically compare our simulations to results from their first vial of "native" bubbles (decantation time is 30 s), of which the size distribution and attenuation are given in their Figs. 5(a) and 4(a), respectively. This vial is selected because it contains a greater fraction of larger bubbles than other vials

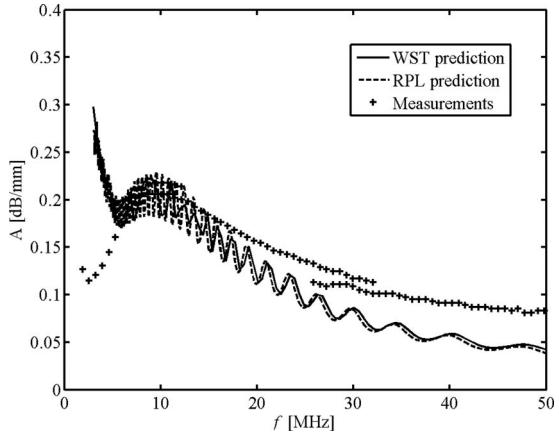


FIG. 5. Attenuation coefficients A from models and measurements;

(with longer decantation times or after filters), and so the HOM contributions from bubble scattering, if present, should be most apparent. Some observations and conclusions can be reached: First, the model predictions agree fairly well at frequencies $<$50 MHz, but the HOM contribution is negligible at this broad frequency range, in contrast to what is seen in monodisperse simulations; second, the models offer good prediction near resonance at around 10 MHz but model-measurement agreement degrades at frequencies $<$7 and $>$20 MHz. The predicted attenuation peak at low frequencies ($<$7 MHz), which should correspond to the large bubbles of radii 6–7 μm, has shifted greatly to a lower frequency range ($<$3 MHz) in the measured results. Measured attenuation is also far more than predicted at frequencies $>$20 MHz. These disagreements reveal the failure of both prediction methods and will be discussed in next section.

D. Primary acoustic radiation force

Accurately predicting the acoustic radiation force on the microbubble subject to the HFUS field is another concern introduced by the WST model. The primary radiation forces on the encapsulated microbubbles are calculated by means of Eq. (12) of Dayton *et al.* (1997) and Eqs. (23), (24), and (28) of Hasegawa *et al.* (1993) for RPL and WST models, respectively. The acoustic radiation forces as a function of bubble radius are shown for the cases of both low frequency (3.5 MHz) and high frequency (50 MHz) in Figs. 6(a) and 6(b). The two models predict approximately identical profiles of acoustic radiation force at 3.5 MHz but predict remarkable differences for bubbles larger than 2 μm in radius at 50 MHz. The differences increase with the bubble radius as well. Resonant peaks are also shown in the WST result in Fig. 6(b) and may correspond to natural frequencies of vibrational modes. Unlike the conclusions from Fig. 4, the difference in the results of acoustic radiation forces between monopole and multi-modal scattering is much larger than that between monopole scattering and RPL scattering.

IV. DISCUSSION

Our simulation results highlight four novel points regarding the acoustic response of contrast microbubbles at high frequencies as follows. (1) The HOM contribution is not negligible at high frequencies. (2) The results of isotropic scattering from both models are also different at high frequencies. (3) Both models cannot well-predict attenuation in suspensions with dispersed Definity bubbles. (4) These differences are exhibited in multiple acoustic scattering properties of microbubbles including not only scattering cross sections and attenuation but also acoustic radiation forces, which are key parameters of various currently available contrast techniques. The results may lead to deeper understanding of the microbubble dynamics, as well as optical usage and potential applications of contrast microbubbles at HFUS.

It is seen in Fig. 3 that monopole scattering from WST model and scattering from current RPL model are slightly different from each other. This difference is more evident in Figs. 4 and 6(b) and may be due to the following two reasons. First, the effect of liquid compressibility is ignored in Eq. (8) but is included in WST model, which, based on prior work (Church, 1995; Hoff, 2001), should lead to considerable errors at high frequencies. Second, in the RPL model, it is assumed that the air pressure stays uniform, but the WST model takes the gas inertial effects into consideration. At high frequencies, the time period is so short that the internal energy in the air may become spatially nonuniform during the oscillation. As a result of these issues, we believe the WST model may be more reasonable to predict microbubble responses to HFUS field.

The gas inertial effects can be further discussed by the modeling of thermal behavior within the microbubble. It has been previously reported that the selection of polytropic exponent κ depends on the encapsulation properties and was given by complex expressions (Hoff, 2001). Selected value for κ in RPL models varies from 1.0 to γ, representing isothermal to adiabatic behavior of gas (Sarkar *et al.*, 2005; Goertz *et al.*, 2007a). An empirical value of 1.1 is also found to best match numerical and analytical solutions for the case of Nycomed and Albunex® microbubbles within conventional frequency range (Khismatullin, 2004). The modeling of heat conduction through the Albunex® bubble wall in WST is investigated by the comparisons in Fig. 7. It is seen from Figs. 7(a) and 7(b) that the resonance slightly shifts when the WST plot is compared to both isothermal and empirical RPL plots; in Fig. 7(c), the WST result best matches for an adiabatic RPL model. It has been demonstrated that whether the compression process is isothermal or adiabatic for free bubble oscillation depends on the ratio of thermal diffusion length to bubble radius defined by the parameter $\chi=D_0/\omega R_2^2$ (Prosperetti *et al.*, 1988). The large and small values of χ indicate nearly isothermal and adiabatical behavior of gas, respectively. Obviously, the value of χ also has a reciprocal relation to the driving frequency. In other words, the value of χ will be much smaller and thus represents the adiabatic behavior of gas at high frequencies. The WST is shown here to display adiabatic gas behavior and may thus be a good model for encapsulated microbubbles subject to high frequencies.

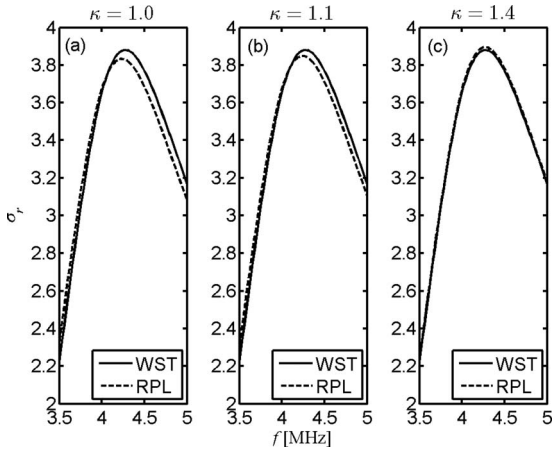


FIG. 7. Reduced scattering cross sections of single Albunex® bubble as a function of driving frequency. The polytropic exponents are (a) 1.0 (left), (b) 1.1 (center), and (c) 1.4 (right).

The failure of both models in predicting the attenuation of Definity® suspensions at higher frequencies in Fig. 5 excludes the HOM contribution, liquid compressibility, and gas inertial effects. Alternative explanations here may link to the material properties of lipid shells, which are different from those of polymer and albumin encapsulation (Sarkar *et al.*, 2005; Doinikov and Dayton, 2007). Disagreement above 20 MHz could also be attributed to nonlinear shell response due to shear-thinning and strain-softening material properties (Tsiglifis and Pelekasis, 2008; Doinikov *et al.*, 2009). However, it is necessary to state that the importance of HOM scattering should be reexamined for other types of contrast agents, especially those with a greater number percentage of large bubbles or encapsulated by thick, hard shells, or driven at even higher frequencies >100 MHz. One good example is Imagent® bubbles, which have a mean size of 6.0 μm. Moreover, from the shift of the measured attenuation peak at frequencies <7 MHz, we believe that further investigation is warranted to examine problems such as theoretical modeling of lipid shells and experimental design of acoustic measurements.

The significance of the HOM contribution to the primary acoustic radiation forces on the microbubble subject to the HFUS has been exhibited in Fig. 6(b) for larger bubbles. As is known, the second scattering mode ($m=1$) has a dipole pattern in scattering directivity and results in a unidirectional radiation force that enables the detection of bubble translational motion. It is thus indicated that bubble manipulation by ultrasound will be violently impacted by the anisotropic scattering for high-frequency targeted contrast applications such as drug delivery and molecular imaging. It is also no doubt that the frequencies applied in upcoming microbubble-assisted ultrasound techniques will get higher, and thus it will lead to considerable HOM contributions for even smaller microbubbles.

Microbubble usage in HFUS imaging may also be optimized by means of WST model. As introduced above, the monopole scattering is dominant only if the normalized fre-

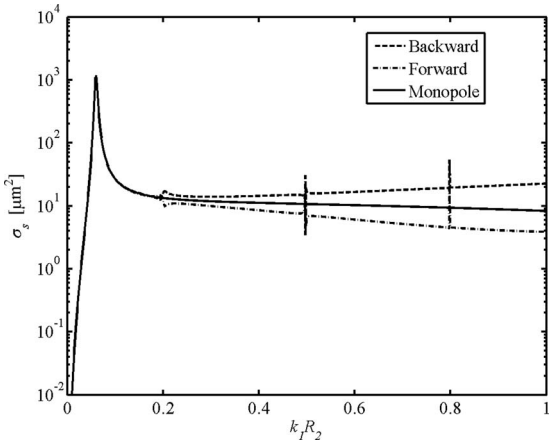


FIG. 8. Scattering cross sections of single Nycomed bubble as a function of normalized frequency.

quency $k_1R_2 \ll 1$, indicating that such scattering is on par with the scattering from radial pulsation predicted by the RPL model for contrast microbubbles at frequencies <10 MHz. However, this condition is partially violated for contrast microbubbles at higher frequencies, and in such intermediate wavelength regimes the HOM contribution becomes significant. Figure 8 shows the scattering cross sections for various terms as a function of normalized frequency for Nycomed bubbles, which has a fixed value of 5% for thickness-to-radius ratio (Hoff, 2001). It is seen that the differences due to the HOM start at $k_1R_2=0.2$. Further calculations lead to a range of start points of k_1R_2 from 0.1 to 0.4 for other types of shells. This result is instructive for optimized usage of contrast bubbles in HFUS applications. A contour plot of k_1R_2 is given in Fig. 9. Given a certain frequency, one may use the plot to find the corresponding threshold of bubble radius below which the HOM contribution is negligible. For example, the HOM contribution to scattering in bubbles with radius <1 μm is negligible at 50 MHz if the start point $k_1R_2=0.2$ (Nycomed bubbles). This indicates that the WST analyses instead of the RPL method should be applied for predictions on single scattering of bubbles >1 μm

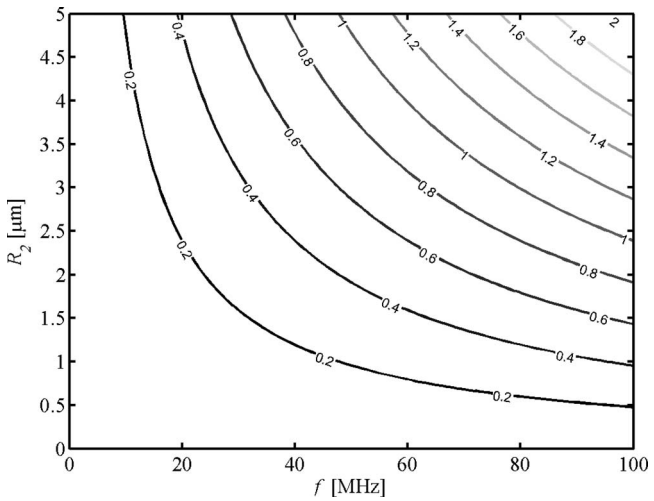


FIG. 9. Contour plot of normalized frequency.

in radius and multiple scattering for suspensions of which a great portion of the bubbles >1 μm in radius.

Shell Lamb waves are essential signals of the contrast microbubble. Simulation results have shown that optimized design of material properties and geometry of the shell can maximize the response of the Lamb wave for a specific frequency (Allen *et al.*, 2001). This may make possible novel HFUS contrast techniques in ophthalmology, dermatology and oncology and so on, based on detection of shell Lamb waves within the total scattering spectrum of microbubbles seeded in complex structures of microcirculation. In underwater acoustics, the modal verification and peak detection are realized by the partial wave decomposition of the scattering form function in the analyses of resonance scattering theory (Gaunaurd and Werby, 1991). This approach may be useful for detection of Lamb waves for encapsulated microbubbles and is a topic of ongoing study.

The WST model holds both advantages and disadvantages against various RPL models in modeling of encapsulated microbubble dynamics at high frequencies. Besides the HOM contribution, it has also been shown previously that liquid compressibility is no longer negligible for Mach numbers comparable to unity (Hoff, 2001). For time-harmonic scattering, the radial velocity of the bubble shell outer interface is $\dot{R}_2 \sim 2\pi f R_2$, and thus the Mach number of the fluid is $M=2\pi f R_2/c_0=k_1R_2$. Developed RPL models, such as the Triling or Keller–Miksis equations, are restricted by $M<1$, but WST method works for any k_1R_2. In addition, studies on the microbubble responses at high frequencies can be extended to more practical linear WST models. For example, a viscous-liquid WST model can be used to evaluate bubbles in blood. However, the nonlinear oscillation of microbubbles has never been incorporated in WST models and is under further investigation. To date, the RPL models continue to be used for the evaluation of nonlinear effects.

ACKNOWLEDGMENTS

This work is supported by grants from NIH (Grant Nos. T32-HL072738, SCCOR-HL081506, and K24-HL084923) and NSF (Grant No. CTS 0421461). Jiusheng Chen would like to thank Professor Thomas L. Geers for useful discussion on WST.

NOMENCLATURE

R_1 = Inner radius of the shell
R_2 = Outer radius of the shell
d = Shell thickness
θ = Scattering angle
Φ_i = Potential of the incident waves
Φ_s = Potential of the scattering waves
Φ_a = Wave potential in the air
Φ_2 = Scalar potential in the shell
Ψ_2 = Vector potential in the shell
μ_s = Shear viscosity of the shell material
G_s = Shear modulus of the shell material
λ_e = Elastic Lamé's first parameter
λ_v = Viscous Lamé's first parameter

k_1 = Wave number in water
k_s = Wave number of the shear waves in the shell
k_d = Wave number of the longitudinal waves in the shell
k_3 = Wave number in the air
$P_n(\cdot)$ = Legendre polynomial of order n
$j_n(\cdot)$ = Spherical Bessel function of order n
$n_n(\cdot)$ = Spherical Neumann function of order n
$h_n^{(1)}(\cdot)$ = Spherical Hankel function of the first kind of order n
a_n = Scattering coefficient
u_r = Normal displacement
τ_{rr} = Normal stress
$\tau_{r\theta}$ = Tangential stress
α_{ij} = Elements of the scattering matrix $(i,j=1-6)$
$f(\theta)$ = Scattering function in the far field
$f_n(\theta)$ = Scattering function of mode n
f_0 = Monopole scattering function
$f(0)$ = Forward scattering function
$f(\pi)$ = Backward scattering function
σ_s = Scattering cross section
σ_r = Reduced scattering cross section
$\sigma_r^{(m)}$ = σ_r calculated by truncation of first $(m+1)$ modes
ρ_s = Density of the shell material
ρ_l = Density of water
β_r = Damping due to acoustic radiation
β_{sh} = Damping due to shell viscosity
β = $\beta_r+\beta_{\mathrm{th}}$, total damping
p_a = Acoustic pressure amplitude
p_0 = Hydrostatic pressure in the surrounding liquid
R_{10} = Inner radius of the shell at $t=0$
R_{20} = Outer radius of the shell at $t=0$
f = Driving frequency
ω = $2\pi f$, angular driving frequency
ω_0 = Resonant angular frequency
c_0 = Sound speed in water
κ = Polytropic exponent
γ = Ratio of specific heats
A = Sound attenuation coefficient
F_{ac} = Primary acoustic radiation force
D_0 = Gas thermal diffusivity

Allen, J. S., Kruse, D. E., and Ferrara, K. W. (**2001**). "Shell waves and acoustic scattering from ultrasound contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 409–418.

Ayres, V. M., and Gaunaurd, G. C. (**1987**). "Acoustic-resonance scattering by viscoelastic objects," J. Acoust. Soc. Am. **81**, 301–311.

Ayres, V. M., Gaunaurd, G. C., Tsui, C. Y., and Werby, M. F. (**1987**). "The effects of lamb waves on the sonar cross-sections of elastic spherical-shells," Int. J. Solids Struct. **23**, 937–946.

Chen, J. S., and Zhu, Z. M. (**2005**). "Sound scattering characteristics of bubbles with viscoelastic shells," Acta Acust. **30**, 385–392.

Cheung, K., Couture, O., Bevan, P. D., Cherin, E., Williams, R., Burns, P. N., and Foster, F. S. (**2008**). "In vitro characterization of the subharmonic ultrasound signal from Definity microbubbles at high frequencies," Phys. Med. Biol. **53**, 1209–1223.

Church, C. C. (**1995**). "The effects of an elastic solid-surface layer on the radial pulsations of gas-bubbles," J. Acoust. Soc. Am. **97**, 1510–1521.

Coussios, C. C., Holland, C. K., Jakubowska, L., Huang, S. L., MacDonald, R. C., Nagaraj, A., and McPherson, D. D. (**2004**). "In vitro characterization of liposomes and Optison® by acoustic scattering at 3.5 MHz," Ultrasound Med. Biol. **30**, 181–190.

Dayton, P. A., Morgan, K. E., Klibanov, A. L. S., Brandenburger, G., Nightingale, K. R., and Ferrara, K. W. (**1997**). "A preliminary evaluation of the effects of primary and secondary radiation forces on acoustic contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 1264–1277.

de Jong, N., Cornet, R., and Lancée, C. T. (**1994**). "Higher harmonics of vibrating gas-filled microspheres. 1. Simulations," Ultrasonics **32**, 447–453.

Doinikov, A. A., and Dayton, P. A. (**2007**). "Maxwell rheological model for lipid-shelled ultrasound microbubble contrast agents," J. Acoust. Soc. Am. **121**, 3331–3340.

Doinikov, A. A., Haac, J. F., and Dayton, P. A. (**2009**). "Modeling of nonlinear viscous stress in encapsulating shells of lipid-coated contrast agent microbubbles," Ultrasonics **49**, 269–275.

Foldy, L. L. (**1945**). "The multiple scattering of waves. 1. General theory of isotropic scattering by randomly distributed scatterers," Phys. Rev. **67**, 107–119.

Frinking, P. J. A., and de Jong, N. (**1998**). "Acoustic modeling of shell-encapsulated gas bubbles," Ultrasound Med. Biol. **24**, 523–533.

Gaunaurd, G. C., and Werby, M. F. (**1991**). "Sound scattering by resonantly excited, fluid-loaded, elastic spherical-shells," J. Acoust. Soc. Am. **90**, 2536–2550.

Goertz, D. E., de Jong, N., and van der Steen, A. F. W. (**2007a**). "Attenuation and size distribution measurements of Definity™ and manipulated Definity™ populations," Ultrasound Med. Biol. **33**, 1376–1388.

Goertz, D. E., Frijlink, M. E., de Jong, N., and Steen, A. F. W. V. (**2006**). "High frequency nonlinear scattering from a micrometer to submicrometer sized lipid encapsulated contrast agent," Ultrasound Med. Biol. **32**, 569–577.

Goertz, D. E., Frijlink, M. E., Tempel, D., Bhagwandas, V., Gisolf, A., Krams, R., de Jong, N., and van der Steen, A. F. W. (**2007b**). "Subharmonic contrast intravascular ultrasound for vasa vasorum imaging," Ultrasound Med. Biol. **33**, 1859–1872.

Goessling, W., North, T. E., and Zon, L. I. (**2007**). "Ultrasound biomicroscopy permits in vivo characterization of zebrafish liver tumors," Nat. Methods **4**, 551–553.

Hasegawa, T., Hino, Y., Annou, A., Noda, H., Kato, M., and Inoue, N. (**1993**). "Acoustic radiation pressure acting on spherical and cylindrical-shells," J. Acoust. Soc. Am. **93**, 154–161.

Hasheminejad, S. M., and Safari, N. (**2005**). "Acoustic scattering from viscoelastically coated spheres and cylinders in viscous fluids," J. Sound Vib. **280**, 101–125.

Hoff, L. (**2001**). *Acoustic Characterization of Contrast Agents for Medical Ultrasound Imaging* (Kluwer Academic, Boston, MA).

Hu, Y. T., Qin, S. P., and Jiang, Q. (**2004**). "Characteristics of acoustic scattering front a double-layered micro shell for encapsulated drug delivery," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 808–820.

James, M. A., Tullett, J., Hemsley, A. G., and Shore, A. C. (**2006**). "Effects of aging and hypertension on the microcirculation," Hypertension **47**, 968–974.

Keller, J. B., and Miksis, M. (**1980**). "Bubble oscillations of large-amplitude," J. Acoust. Soc. Am. **68**, 628–633.

Khismatullin, D. B. (**2004**). "Resonance frequency of microbubbles: Effect of viscosity," J. Acoust. Soc. Am. **116**, 1463–1473.

Lanza, G. M., Wallace, K. D., Fischer, S. E., Christy, D. H., Scott, M. J., Trousil, R. L., Cacheris, W. P., Miller, J. G., Gaffney, P. J., and Wickline, S. A. (**1997**). "High-frequency ultrasonic detection of thrombi with a targeted contrast system," Ultrasound Med. Biol. **23**, 863–870.

McClements, D. J. (**1996**). "Principles of ultrasonic droplet size determination in emulsions," Langmuir **12**, 3454–3461.

Mitri, F. G., and Fellah, Z. E. A. (**2006**). "Amplitude-modulated acoustic radiation force experienced by elastic and viscoelastic spherical shells in progressive waves," Ultrasonics **44**, 287–296.

Moran, C. M., Watson, R. J., Fox, K. A. A., and McDicken, W. N. (**2002**). "In vitro acoustic characterisation of four intravenous ultrasonic contrast agents at 30 MHz," Ultrasound Med. Biol. **28**, 785–791.

Needles, A., Goertz, D. E., Karshafian, R., Cherin, E., Brown, A. S., Burns, P. N., and Foster, F. S. (**2008**). "High-frequency subharmonic pulsed-wave Doppler and color flow imaging of microbubble contrast agents," Ultrasound Med. Biol. **34**, 1139–1151.

Prosperetti, A., Crum, L. A., and Commander, K. W. (**1988**). "Nonlinear bubble dynamics," J. Acoust. Soc. Am. **83**, 502–514.

Rychak, J. J., Graba, J., Cheung, A. M. Y., Mystry, B. S., Lindner, J. R., Kerbel, R. S., and Foster, F. S. (**2007**). "Microultrasound molecular imaging of vascular endothelial growth factor receptor 2 in a mouse model of tumor angiogenesis," Mol. Imaging **6**, 289–296.

Sarkar, K., Shi, W. T., Chatterjee, D., and Forsberg, F. (**2005**). "Characterization of ultrasound contrast microbubbles using in vitro experiments and viscous and viscoelastic interface models for encapsulation," J. Acoust. Soc. Am. **118**, 539–550.

Trilling, L. (**1952**). "The collapse and rebound of a gas bubble," J. Appl. Phys. **23**, 14–17.

Tsiglifis, K., and Pelekasis, N. A. (**2008**). "Nonlinear radial oscillations of encapsulated microbubbles subject to ultrasound: The effect of membrane constitutive law," J. Acoust. Soc. Am. **123**, 4059–4070.

Waterman, P. C., and Truell, R. (**1961**). "Multiple scattering of waves," J. Math. Phys. **2**, 512–517.

Ye, Z. (**1996**). "On sound scattering and attenuation of Albunex® bubbles," J. Acoust. Soc. Am. **100**, 2011–2028.

Yeh, C. K., Lu, S. Y., and Chen, Y. S. (**2008**). "Microcirculation volumetric flow assessment using high-resolution, contrast-assisted images," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **55**, 74–83.

Sound radiation quantities arising from a resilient circular radiator

Ronald M. Aarts[a] and Augustus J. E. M. Janssen
Philips Research Europe, HTC 36 (WO-02), NL-5656AE Eindhoven, The Netherlands

(Received 3 November 2008; revised 11 July 2009; accepted 27 July 2009)

Power series expansions in ka are derived for the pressure at the edge of a radiator, the reaction force on the radiator, and the total radiated power arising from a harmonically excited, resilient, flat, circular radiator of radius a in an infinite baffle. The velocity profiles on the radiator are either Stenzel functions $(1-(\sigma/a)^2)^n$, with σ the radial coordinate on the radiator, or linear combinations of Zernike functions $P_n(2(\sigma/a)^2-1)$, with P_n the Legendre polynomial of degree n. Both sets of functions give rise, via King's integral for the pressure, to integrals for the quantities of interest involving the product of two Bessel functions. These integrals have a power series expansion and allow an expression in terms of Bessel functions of the first kind and Struve functions. Consequently, many of the results in [M. Greenspan, J. Acoust. Soc. Am. **65**, 608–621 (1979)] are generalized and treated in a unified manner. A foreseen application is for loudspeakers. The relation between the radiated power in the near-field on one hand and in the far field on the other is highlighted. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3206580]

PACS number(s): 43.38.Ar, 43.20.Bi, 43.20.Px, 43.40.At [JGM] Pages: 1776–1787

I. INTRODUCTION

The analytical theory of sound radiation for the case of a harmonically excited, circular piston in an infinite baffle is firmly established in the literature. There are chapters in text books,[1–5] survey papers,[6,7] and many research papers from older[8–13] and more recent[14–21] years devoted to this subject. A big effort has been made to find series or convenient integral expressions for the sound pressure at all field points on or in front of the disk and the baffle. The single integral approach, in which the pressure is expressed via Rayleigh's integral[2] or by other means as a single integral with bounded integration limits, occurs in Refs. 10, 11, 13, and 15 and is reviewed, with emphasis on numerical work, in Ref. 16. The spherical-wave-expansion approach uses Gegenbauer's addition theorem[22] to express the pressure as a series of integrals involving spherical Bessel functions and Legendre polynomials and can be found in Refs. 15, 18–20, and 23. Furthermore, King's integral[9] is used in Refs. 6, 12, 14, 17, and 19 and this yields expressions for the pressure in the form of an infinite integral involving the product of two Bessel functions and an exponential factor. In order that this integral can be used conveniently for computations, it is often necessary to employ somewhat more advanced complex function theory. Finally, in Ref. 24 series expansions following from Rayleigh's integral[2] are given for the on-disk pressure, and in Ref. 19 King's integral is used to develop a double-series expansion for the on-disk pressure.

In the present paper, single-series expressions are developed for quantities associated with the pressure (rather than for the pressure in the field itself) from King's integral. The velocity profile on the radiator is allowed here to be non-uniform. A set of results of this type, for the reaction force and the total radiated power, was obtained by Greenspan.[6] Greenspan considered, what are called in the present paper, Stenzel functions[8] $s^{(n)}(\sigma)=(1-(\sigma/a)^2)^n$, $0\le\sigma\le a$ $(n=0,1,2)$, and (infinitely supported) Gaussians and established series expansions and closed-form expressions for the quantities just mentioned. These results were derived from King's integral in an *ad hoc* manner with an impressive amount of analytical skill.

Greenspan's results will be generalized to velocity profiles of the Stenzel type of arbitrary order n and of the Zernike[25] type, see below, and linear combinations of the latter. The acoustical quantities considered here are edge pressure, reaction force, total power, and directivity. Zernike functions have the form $R^0_{2n}(\sigma/a)=P_n(2(\sigma/a)^2-1)$, with P_n the Legendre polynomial of degree n. Linear combinations of both Stenzel functions and Zernike functions can be used to approximate any radially symmetric velocity profile. In this respect, Zernike functions are much more effective, in terms of the required number of coefficients and amplitudes of these, than the Stenzel functions. Zernike functions have been considered recently[21] by the authors with respect to their potential and efficacy for forward computation of the on-axis and far-field pressure from a non-uniform velocity profile on the piston in terms of its Zernike expansion coefficients. Here it has been very helpful that Zernike functions are orthogonal and that many velocity profiles have Zernike coefficients that can be found in analytical form, see Ref. 21, Appendix A. Moreover, the inverse problem of estimating a velocity profile in terms of its expansion coefficients from on-axis pressure data has been considered in Ref. 21. Results for quantities related to the pressure that admit an analytical treatment via King's integral were, however, not presented in

[a]Author to whom correspondence should be addressed. Electronic mail: ronald.m.aarts@philips.com. Also at: Technical University Eindhoven, Den Dolech 2, PT3.23, P.O. Box 513, NL-5600MB Eindhoven, The Netherlands.

Ref. 21. This is done in the present paper: for the quantities of interest it will be shown that King's integral yields an attractive power series expansion.

Stenzel functions, with half-integer rather than integer order n, have been considered recently by Mellow[19] in the context of King's integral for the pressure. In Ref. 19, these Stenzel type functions have been used to expand a non-uniform velocity profile for the purpose of computing the pressure everywhere on the radiator. This leads via King's integral to certain double series expansions for the pressure on the radiator that are more complicated than what is obtained here for the pressure at the edge of the radiator.

This paper is organized as follows. In Sec. II the basic formulas and functions are presented, and the results of this paper are discussed globally. In Secs. III–V the formulas for the pressure at the edge, the reaction force, and the total power and directivity are presented and discussed. These formulas are given for the case that v is a general (integer-order) Stenzel function or a linear combination of Zernike functions. In Sec. VI, it is briefly indicated how the power can be estimated from near-field on-axis pressure data via the inverse method developed in Ref. 21. In Sec. VII the conclusions are presented. Finally, in Appendix A the required integrals are computed in the form of power series in ka, and in Appendix B the convergence properties of these series are discussed.

II. BASIC FORMULAS AND OVERVIEW

The radiated pressure is given by Rayleigh's integral[2,26] as

$$p(\underline{r},t) = \frac{i\rho_0 ck}{2\pi} e^{i\omega t} \int_S v(\underline{r}_s) \frac{e^{-ikr'}}{r'} dS, \tag{1}$$

where ρ_0 is the density of the medium, c is the speed of sound in the medium, $k=\omega/c$ is the wave number, and ω is the radial frequency of the vibrating surface S for which a disk of radius a in an infinite baffle is taken. Furthermore, $\underline{r}$ is a field point, $\underline{r}_s$ is a point on the disk S, $r'=|\underline{r}-\underline{r}_s|$ is the distance between $\underline{r}$ and $\underline{r}_s$, and $v(\underline{r}_s)$ is the normal component of a (not necessarily uniform) velocity profile; see Fig. 1 for the geometry and notations. The time variable t in $p(\underline{r},t)$ and the harmonic factor $\exp(i\omega t)$ in front of the integral in Eq. (1) will be omitted in the sequel. The average velocity on S is denoted by V_s; hence the volume velocity V is given by

$$V = \int_S v(\underline{r}_s) dS = \pi a^2 V_s \tag{2}$$

[the definition of V is in agreement with Greenspan's definition and notation,[6] Eq. (2); in the present paper only V_s will be used].

In the main body of this paper, the velocity profile $v(\underline{r}_s)$ is assumed to be radially symmetric and is written as $v(\sigma)$, $0\leq\sigma\leq a$ [with $v(\sigma)=0$, $\sigma>a$]. In the notations of Fig. 1, King's result[9] reads

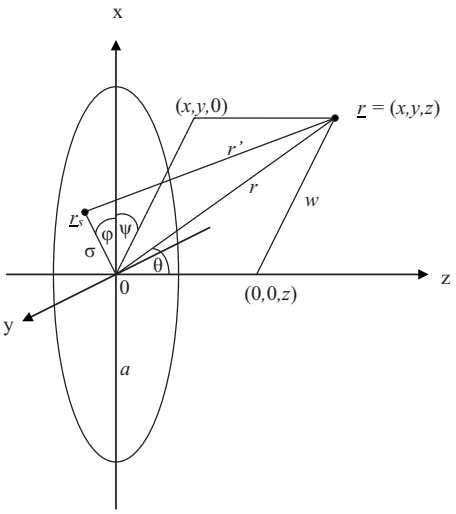


FIG. 1. Set-up and notations. The piston is surrounded by an infinite rigid baffle. $\underline{r}_s=(x_s,y_s,0)=(\sigma\cos\varphi,\sigma\sin\varphi,0)$, $\underline{r}=(x,y,z)=(r\sin\theta\cos\psi, r\sin\theta\sin\psi, r\cos\theta)$, $w=r\sin\theta=(x^2+y^2)^{1/2}$, $z=r\cos\theta$, $r=|\underline{r}|=(x^2+y^2+z^2)^{1/2}=(w^2+z^2)^{1/2}$, $r'=|\underline{r}-\underline{r}_s|=(r^2+\sigma^2-2\sigma w\cos(\psi-\varphi))^{1/2}$.

$$p(\underline{r}) = \frac{i\rho_0 ck}{2\pi} \int_S v(\sigma) \frac{e^{-ikr'}}{r'} dS$$

$$= i\rho_0 ck \int_0^\infty \frac{e^{-z(u^2-k^2)^{1/2}}}{(u^2-k^2)^{1/2}} J_0(wu) V(u) u du, \tag{3}$$

where

$$(u^2-k^2)^{1/2} = \begin{cases} i\sqrt{k^2-u^2}, & 0\leq u\leq k, \\ \sqrt{u^2-k^2}, & k\leq u<\infty, \end{cases} \tag{4}$$

with $\sqrt{}$ non-negative, and

$$V(u) = \int_0^a J_0(u\sigma) v(\sigma) \sigma d\sigma, \quad u\geq 0, \tag{5}$$

is the Hankel transform (of order 0) of v and $J_n(z)$ is the Bessel function of order n, where $n=0$ in Eqs. (3) and (5).

In this paper, two sets of functions for use in conjunction with King's result will be highlighted. The first set of functions are Stenzel's function $s^{(n)}(\sigma)$ given by

$$v(\sigma) = s^{(n)}(\sigma) = (n+1)V_s(1-(\sigma/a)^2)^n, \quad 0\leq\sigma\leq a, \tag{6}$$

with $n=0,1,\ldots$. To the authors' knowledge, Stenzel[8] was the first author to write on these functions in relation to sound radiation [the profiles in Eq. (6) were considered extensively by Greenspan,[6] but no reference to work of Stenzel can be found in Ref. 6; also see Ref. 7, p. 14]. The Hankel transform $S^{(n)}(u)$ of $s^{(n)}$ is given by[8]

$$S^{(n)}(u) = a^2(n+1)!2^n V_s \frac{J_{n+1}(au)}{(au)^{n+1}}. \tag{7}$$

A second set of functions considered in this paper is the set of Zernike functions

$$R_{2n}^0(\sigma/a) = P_n(2(\sigma/a)^2-1), \quad 0\leq\sigma\leq a. \tag{8}$$

Any radially symmetric velocity profile $v(\sigma)$, $0\leq\sigma\leq a$, allows an (orthogonal) expansion as

$$v(\sigma) = V_s \sum_{n=0}^{\infty} u_n R_{2n}^0(\sigma/a), \tag{9}$$

where $u_0=1$ [this follows from the definition in Eq. (2) and the fact that $\int_0^a R_{2n}^0(\sigma/a)\sigma d\sigma = \frac{1}{2}a^2\delta_{n0}$]. The Hankel transform $V(u)$ of $v(\sigma)$ in Eq. (9) is given by[21,25,27]

$$V(u) = V_s \sum_{n=0}^{\infty} u_n (-1)^n \frac{a}{u} J_{2n+1}(ua). \tag{10}$$

After inserting Eq. (7) or Eq. (10) into Eq. (3), integrals of the form

$$\int_0^{\infty} e^{-z(u^2-k^2)^{1/2}} \frac{J_\nu(wu)J_\mu(au)}{u^\lambda (u^2-k^2)^{1/2}} u du \tag{11}$$

with $\nu=0$ appear. These integrals seem too complicated to allow a completely analytic treatment. However, there are cases with $z=0$ and $w=a$ that allow such a treatment. These cases occur as follows:

(a) for the pressure at the edge with $\underline{r}=(a\cos\psi, a\sin\psi, 0)$, and $w=a$, see Sec. III;
(b) in the computation of the reaction force $F=\int_S p dS$ on the radiator, see Sec. IV;
(c) in the computation of the radiated power $P=\int_S p v^* dS$, see Sec. V.

The quantity in (a) was expressed in terms of Bessel and Struve functions by Warren[28] for the case $n=0$, and this result was proved and generalized (to the cases $n=0,1$) by McLachlan.[24] The quantities mentioned in (b) and (c) were expressed by Rayleigh[26] [case (b), $n=0$] and Greenspan[6] [cases (b) and (c), $n=0,1,2$] for the case of Stenzel radiators $s^{(n)}$ in terms of Bessel and Struve functions, and Greenspan gave the first few power series coefficients of reaction force and power as a function of ka. Moreover, Greenspan obtained similar results for the (infinitely supported) Gaussian. In the present paper the quantities in (a)–(c) will be computed as power series for all Stenzel cases $v=s^{(n)}$ and for all terms and cross terms pertaining to $v(\sigma)$ in Eq. (9). The coefficients of these power series are organized in such a way that the close connection between the integral expression on one hand and the Bessel functions of the first kind and the Struve functions on the other hand is immediately apparent. Finally, the series are cast into single-series format which makes them convenient to use. Greenspan's results have been used as a check of correctness of the formulas here. To facilitate this, the formulas have been brought into the same form as Greenspan's results. Also a number of cross-checks have been carried out.

The main results of this paper follow from the power series expansions in ka of the integrals that appear in Eq. (11) with $w=a$ and $z=0$. These power series are derived in Appendix A. In Appendix B the convergence behavior of these expansions are considered. It thus appears that all series provide 10^{-6} absolute accuracy when they are truncated at a summation index $\geqslant 2eka+10$ and when $ka \leqslant 12$ (machine precision 10^{-15}).

III. PRESSURE AT THE EDGE

According to Eq. (3), the pressure p_{edge} at an edge point $(a\cos\psi, a\sin\psi, 0)$ of the radiator is obtained by taking $z=0$ and $w=a$. Thus

$$p_{\text{edge}} = i\rho_0 ck \int_0^{\infty} \frac{J_0(au)V(u)}{(u^2-k^2)^{1/2}} u du, \tag{12}$$

with $V(u)$ the Hankel transform in Eq. (5) of $v(\sigma)$.

A. Stenzel functions

With $v(\sigma)=s^{(n)}(\sigma)$, see Eq. (6), for which the Hankel transform $S^{(n)}(u)$ is given by Eq. (7), the pressure $p_{\text{edge}}^{(n)}$ at the edge is given by

$$\frac{p_{\text{edge}}^{(n)}}{\rho_0 c V_s} = \frac{ik(n+1)!2^n}{a^{n-1}} \int_0^{\infty} \frac{J_0(au)J_{n+1}(au)}{u^n (u^2-k^2)^{1/2}} du. \tag{13}$$

The integral in Eq. (13) has been evaluated in Appendix A 1 and the result is

$$\begin{aligned}
\frac{p_{\text{edge}}^{(n)}}{\rho_0 c V_s} &= (n+1)!\left(\frac{2}{a}\right)^n ka \left[\int_0^k \frac{J_0(au)J_{n+1}(au)}{u^n\sqrt{k^2-u^2}} du \right. \\
&\quad \left. + i\int_k^{\infty} \frac{J_0(au)J_{n+1}(au)}{u^n\sqrt{u^2-k^2}} du \right] \\
&= \frac{1}{2}(n+1)!ka \left[\sum_{j=0}^{\infty} \frac{(-1)^j (2j+2)_n (ka)^{2j+1}}{\Gamma^2(n+j+2)} \right. \\
&\quad \left. + i\sum_{j=0}^{\infty} \frac{(-1)^j (2j+1)_n (ka)^{2j}}{\Gamma^2(n+j+3/2)} \right] \\
&= -\frac{1}{2}(n+1)!\sum_{\ell=1}^{\infty} \frac{(\ell)_n (-ika)^\ell}{\Gamma^2\left(n+\frac{1}{2}\ell+1\right)}.
\end{aligned} \tag{14}$$

Here Γ is the Gamma function and $(x)_n$ is Pochhammers's symbol as defined in Eq. (A13).

The middle expression for $p_{\text{edge}}^{(n)}$ in Eq. (14) is convenient for expressing $\Re p_{\text{edge}}^{(n)}$ and $\Im p_{\text{edge}}^{(n)}$ in terms of Bessel functions of the first kind and Struve functions, respectively, see Eqs. (A4) and (A5). Thus for the case that $n=0$, it is seen that [using $(x)_0=1$]

$$\frac{p_{\text{edge}}^{(0)}}{\rho_0 c V_s} = \frac{1}{2}[1 - J_0(2ka) + i\mathbf{H}_0(2ka)], \tag{15}$$

a result given by Warren[28] (without proof) and proved and discussed by McLachlan.[24] For the case that $n=1$, the coefficients in the two series in the middle expression for $p_{\text{edge}}^{(1)}$ in Eq. (14) must be manipulated. Thus one has [using $(x)_1=x$]

$$\sum_{j=0}^{\infty} \frac{(-1)^j (2j+2)z^{2j+1}}{\Gamma^2(j+3)} = 2\frac{1-J_0(2z)}{z^3} - 2\frac{J_1(2z)}{z^2}, \tag{16}$$

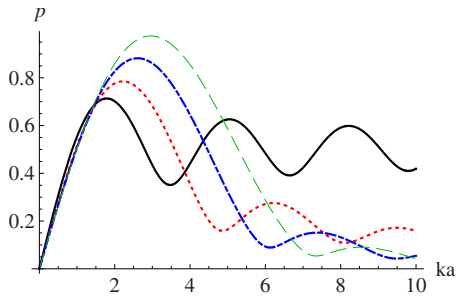


FIG. 2. (Color online) The pressure at the edge $|p_{\text{edge}}^{(n)}/\rho_0 c V_s|$ vs ka, for the rigid piston ($n=0$) (solid curve), the simply supported radiator ($n=1$) (dotted curve), and the first two clamped radiators ($n=2,3$) (dash-dotted and dashed curves, respectively).

$$\sum_{j=0}^{\infty} \frac{(-1)^j (2j+1) z^{2j}}{\Gamma^2(j+5/2)} = \frac{2}{z^3}\mathbf{H}_0(2z) - \frac{2}{z^2}\left(\frac{4}{\pi} - \mathbf{H}_1(2z)\right), \tag{17}$$

where $\mathbf{H}_n(z)$ is the Struve function of order n, and with $z=ka$ it follows that

$$\frac{p_{\text{edge}}^{(1)}}{\rho_0 c V_s} = 2\left\{\frac{1-J_0(2z)}{z^2} - \frac{J_1(2z)}{z} + i\left[\frac{1}{z^2}\mathbf{H}_0(2z) - \frac{1}{z}\left(\frac{4}{\pi} - \mathbf{H}_1(2z)\right)\right]\right\}. \tag{18}$$

This agrees with the result in Ref. 24, Eq. (34), except for the overall factor 2 [due to the definition of $v^{(n)}$ in Eq. (6)] and the signs in front of $J_1(2z)/z$ and $\mathbf{H}_1(2z)$. Since $p_{\text{edge}}^{(1)} \to 0$ as $k \to 0$, the correct signs are as in Eq. (18). In Fig. 2 $|p_{\text{edge}}^{(n)}/\rho_0 c V_s|$ vs ka is plotted [using the last formula of Eq. (14)] for the rigid piston ($n=0$), the simply supported radiator ($n=1$) and the first two clamped radiators ($n=2,3$).

It is observed that the coefficient of $(ka)^2$ in $p_{\text{edge}}^{(n)}$ equals $\frac{1}{2}\rho_0 c V_s$ (independent of n). Thus $\Re[p_{\text{edge}}^{(n)}] \approx \frac{1}{2}\rho_0 c V_s (ka)^2$ for small ka.

B. Zernike functions

With $v(\sigma)$ a linear combination of Zernike functions $R_{2n}^0(\sigma/a)$ as in Eq. (9), for which the Hankel transform $V(u)$ is given by Eq. (10), the pressure p_{edge} at the edge is given by

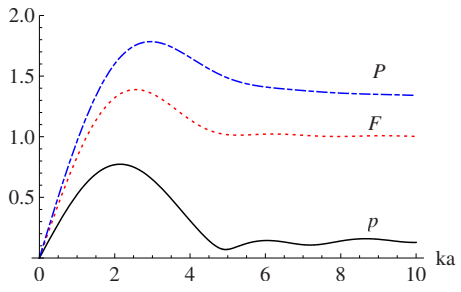


FIG. 3. (Color online) p: the pressure at the edge $|p_{\text{edge}}/\rho_0 c V_s|$ vs ka, (solid curve) using Eq. (20). F: the reaction force $|F/\pi\rho_0 c V_s a^2|$ vs ka (dotted curve) using Eqs. (30) and (31). P: the power $|P/\pi\rho_0 c V_s^2 a^2|$ vs ka (dash-dotted curve), using Eqs. (39) and (40). All curves for a truncated Gaussian velocity profile $\exp(-\alpha(\sigma/a)^2)$, $\alpha=2$, approximated using three Zernike coefficients $u_0=1.0000$, $u_1=-0.9392$, and $u_2=0.3044$.

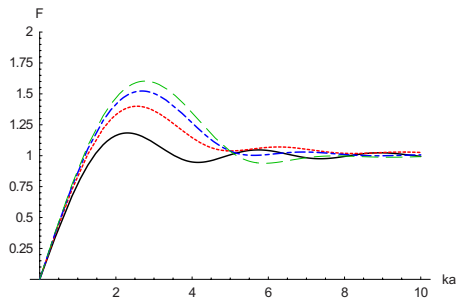


FIG. 4. (Color online) The force on the radiator $|F^{(n)}/\pi\rho_0 c V_s a^2|$ vs ka, for the rigid piston ($n=0$) (solid curve), the simply supported radiator ($n=1$) (dotted curve), and the first two clamped radiators ($n=2,3$) (dash-dotted and dashed curves, respectively).

$$\frac{p_{\text{edge}}}{\rho_0 c V_s} = ika \sum_{n=0}^{\infty} (-1)^n u_n \int_0^{\infty} \frac{J_0(au) J_{2n+1}(au)}{(u^2-k^2)^{1/2}} du. \tag{19}$$

The integrals in Eq. (19) have been evaluated in Appendix A 2, with the result that

$$\begin{aligned} & i\int_0^{\infty} \frac{J_0(au)J_{2n+1}(au)}{(u^2-k^2)^{1/2}} du \\ &= \int_0^k \frac{J_0(au)J_{2n+1}(au)}{\sqrt{k^2-u^2}} du + i\int_k^{\infty} \frac{J_0(au)J_{2n+1}(au)}{\sqrt{u^2-k^2}} du \\ &= \frac{1}{2}\sum_{j=0}^{\infty} (-1)^j \left(\frac{(j+1)_n}{\Gamma(2n+j+2)}\right)^2 (ka)^{2n+2j+1} \\ &\quad + \frac{1}{2} i(-1)^n \sum_{j=0}^{\infty} (-1)^j \left(\frac{(-j+1/2)_n}{\Gamma(j+n+3/2)}\right)^2 (ka)^{2j} \\ &= \frac{-(-1)^n}{2ka} \sum_{\ell=1}^{\infty} \left(\frac{\left(-\frac{1}{2}\ell+1\right)_n}{\Gamma\left(n+\frac{1}{2}\ell+1\right)}\right)^2 (-ika)^{\ell}. \end{aligned} \tag{20}$$

In Fig. 3 $|p_{\text{edge}}/\rho_0 c V_s|$ vs ka is plotted [solid curve, using the last formula of Eq. (20)] for a Gaussian velocity profile $\exp(-\alpha(\sigma/a)^2)$, $\alpha=2$ and truncated at $\sigma=a$, approximated using three Zernike coefficients $u_0=1.0000$, $u_1=-0.9392$, $u_2=0.3044$.

The case with $u_0=1$, $u_1=u_2=\cdots=0$ in Eqs. (19) and (20) yields the same result as Eq. (14) with $n=0$, as it should. Observe also that the real part in Eq. (20) has $O[(ka)^{2n+1}]$-behavior as $ka \to 0$. As a consequence of this and the fact that $u_0=1$ by definition, the coefficient of $(ka)^2$ in p_{edge} equals $\frac{1}{2}\rho_0 c V_s$ for any profile $v(\sigma)$.

The formulas on the last lines of Eqs. (14) and (20) can further be checked against one another, because

$$s^{(n)}(\sigma) = V_s \sum_{j=0}^{n} (-1)^j \frac{2j+1}{j+1} \frac{\binom{n}{j}}{\binom{n+j+1}{n}} (n+1) R_{2n}^0(\sigma/a), \tag{21}$$

see Ref. 21, Eq. (10). For instance, when $n=1$, one has $s^{(1)}(\sigma) = V_s(R_0^0(\sigma/a) - R_2^0(\sigma/a))$, and one has to check that

$$\frac{\ell}{\Gamma^2\left(\frac{1}{2}\ell+2\right)} = \frac{1}{2}\left[\frac{1}{\Gamma^2\left(\frac{1}{2}\ell+1\right)} - \frac{\left(-\frac{1}{2}\ell+1\right)^2}{\Gamma^2\left(\frac{1}{2}\ell+2\right)}\right], \tag{22}$$

which indeed holds. Accordingly, Fig. 2 (and Figs. 4–6)

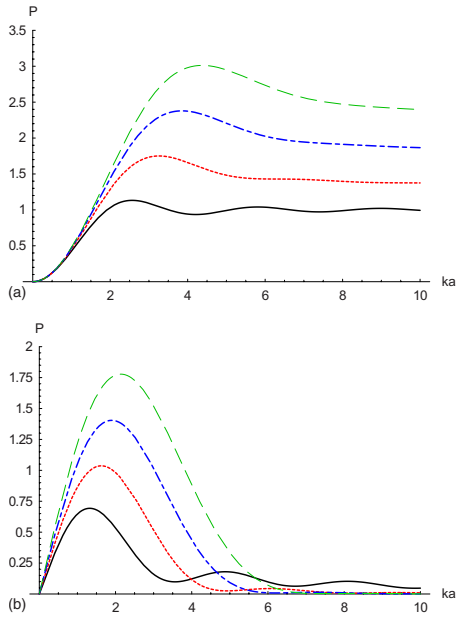


FIG. 5. (Color online) The power of the radiator $P^{(n)}/\pi\rho_0 c V_s^2 a^2$ vs ka, for the rigid piston ($n=0$) (solid curve), the simply supported radiator ($n=1$) (dotted curve), and the first two clamped radiators ($n=2,3$) (dash-dotted and dashed curves, respectively). (a) Real part and (b) imaginary part.

could have been produced equally well within the framework of Zernike expansions.

IV. REACTION ON RADIATOR

According to Eq. (3), the total reaction force F on the radiator is given by

$$F = \int_S p dS = \int_0^a \int_0^{2\pi} p((\sigma\cos\psi, \sigma\sin\psi, 0))\sigma d\sigma d\psi$$
$$= 2\pi \int_0^a i\rho_0 ck \int_0^\infty \frac{J_0(\sigma u)V(u)}{(u^2-k^2)^{1/2}} u du \sigma d\sigma. \tag{23}$$

Because $\int_0^a J_0(\sigma u)\sigma d\sigma = a u^{-1} J_1(au)$, there results

$$F = 2\pi i \rho_0 ck a \int_0^\infty \frac{J_1(au)V(u)}{(u^2-k^2)^{1/2}} du. \tag{24}$$

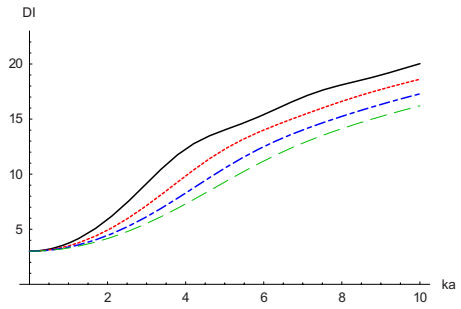


FIG. 6. (Color online) The directivity index $\mathrm{DI}=10\log_{10} D$ (dB) vs ka, for the rigid piston ($n=0$) (solid curve), the simply supported radiator ($n=1$) (dotted curve), and the first two clamped radiators ($n=2,3$) (dash-dotted and dashed curves, respectively).

A. Stenzel functions

With $v(\sigma)=s^{(n)}(\sigma)$ and its Hankel transform $V^{(n)}(u)$ given by Eqs. (6) and (7), the total reaction force $F^{(n)}$ on the radiator is given by

$$\frac{F^{(n)}}{\pi\rho_0 c V_s a^2} = (n+1)!\left(\frac{2}{a}\right)^{n+1} ika \int_0^\infty \frac{J_1(au)J_{n+1}(au)}{u^{n+1}(u^2-k^2)^{1/2}} du. \tag{25}$$

The integral at the right-hand side of Eq. (25) has been evaluated in Appendix A 1, and the result is that

$$\frac{F^{(n)}}{\pi\rho_0 c V_s a^2} = (n+1)!\left(\frac{2}{a}\right)^{n+1} ka\left[\int_0^k \frac{J_1(au)J_{n+1}(au)}{u^{n+1}\sqrt{k^2-u^2}} du + i\int_k^\infty \frac{J_1(au)J_{n+1}(au)}{u^{n+1}\sqrt{u^2-k^2}} du\right]$$
$$= (n+1)!\left[\sum_{j=0}^\infty \frac{(-1)^j(2j+3)_n (ka)^{2j+2}}{\Gamma(n+j+2)\Gamma(n+j+3)} + i\sum_{j=0}^\infty \frac{(-1)^j(2j+2)_n (ka)^{2j+1}}{\Gamma(n+j+3/2)\Gamma(n+j+5/2)}\right]$$
$$= -(n+1)!\sum_{\ell=1}^\infty \frac{(\ell+1)_n (-ika)^\ell}{\Gamma\left(n+\frac{1}{2}\ell+1\right)\Gamma\left(n+\frac{1}{2}\ell+2\right)}. \tag{26}$$

The middle expression for $F^{(n)}$ in Eq. (26) can be used to express $\Re F^{(n)}$ and $\Im F^{(n)}$ in terms of Bessel functions of the first kind and Struve functions, respectively. This yields

$$\frac{F^{(0)}}{\pi\rho_0 c V_s a^2} = 1 - \frac{J_1(2z)}{z} + i\frac{\mathbf{H}_1(2z)}{z}, \tag{27}$$

$$\frac{F^{(1)}}{\pi\rho_0 c V_s a^2} = 1 - \frac{6J_1(2z) - 4zJ_0(2z) - 2z}{z^3} - i\frac{4z\mathbf{H}_0(2z) - 6\mathbf{H}_1(2z)}{z^3}, \tag{28}$$

$$\frac{F^{(2)}}{\pi\rho_0 c V_s a^2} = 1 - 24\frac{(5-z^2)J_1(2z) - \frac{7}{2}zJ_0(2z) - \frac{3}{2}}{z^5} + 24i\frac{(5-z^2)\mathbf{H}_1(2z) - \frac{7}{2}z\mathbf{H}_0(2z) - \frac{2}{3\pi}z}{z^5}, \tag{29}$$

in which $z=ka$. This is in complete agreement with Ref. 6, Eqs. (35), (40), and (41), where it is recalled that Greenspan's V is equal to $\pi a^2 V_s$. Equation (27) is discussed in many texts; recently a simple and effective approximation of $\mathbf{H}_1(z)$ which is valid for all z is developed in Ref. 29.

The expression for $F^{(n)}$ on the last line of Eq. (26) is in the form of a power series in $-ika$. In Ref. 6, Eqs. (35a), (40b), and (41b), the first few terms of the power series of $F^{(n)}$, $n=0,1,2$, have been displayed. It turns out that this is in complete agreement with what Eq. (26) gives for these cases. Furthermore, it can be checked directly from Eq. (26) that the coefficient of $(ka)^2$ in $F^{(n)}$ equals $\frac{1}{2}\pi\rho_0 c V_s a^2$ (inde-

pendent of n and in agreement with Greenspan's observation for $s^{(n)}$, $n=0,1,2$). Thus $\Re[F^{(n)}]\approx\frac{1}{2}\pi\rho_0 c V_s a^2(ka)^2$ for small ka. In Fig. 4 the force on the radiator $F^{(n)}/\pi\rho_0 c V_s a^2$ vs ka is plotted [using the last formula of Eq. (26)] for the rigid piston ($n=0$), the simply supported radiator ($n=1$), and the first two clamped radiators ($n=2,3$).

B. Zernike functions

With $v(\sigma)$ a linear combination of Zernike functions $R^0_{2n}(\sigma/a)$ as in Eq. (9) with Hankel transform $V(u)$ as given in Eq. (10), the total reaction force F is given by

$$\frac{F}{\pi\rho_0 c V_s a^2}=2ik\sum_{n=0}^{\infty}(-1)^n u_n\int_0^{\infty}\frac{J_1(au)J_{2n+1}(au)}{u(u^2-k^2)^{1/2}}du. \quad (30)$$

The integrals at the right-hand side of Eq. (30) have been evaluated in Appendix A 2, with the result that

$$\begin{aligned}
&2ik\int_0^{\infty}\frac{J_1(au)J_{2n+1}(au)}{u(u^2-k^2)^{1/2}}du\\
&=2k\int_0^{k}\frac{J_1(au)J_{2n+1}(au)}{u\sqrt{k^2-u^2}}du+2ik\int_k^{\infty}\frac{J_1(au)J_{2n+1}(au)}{u\sqrt{u^2-k^2}}du\\
&=\sum_{j=0}^{\infty}\frac{(-1)^j(j+1)_n(j+2)_n}{\Gamma(j+2n+2)\Gamma(j+2n+3)}(ka)^{2(n+j+1)}\\
&\quad+i(-1)^n\sum_{j=0}^{\infty}\frac{(-1)^j\left(-j-\frac{1}{2}\right)_n\left(-j+\frac{1}{2}\right)_n}{\Gamma(j+n+3/2)\Gamma(j+n+5/2)}(ka)^{2j+1}\\
&=-(-1)^n\sum_{\ell=1}^{\infty}\frac{\left(-\frac{1}{2}\ell\right)_n\left(-\frac{1}{2}\ell+1\right)_n}{\Gamma\left(\frac{1}{2}\ell+n+1\right)\Gamma\left(\frac{1}{2}\ell+n+2\right)}(-ika)^{\ell}. \quad (31)
\end{aligned}$$

In Fig. 3 $|F/\pi\rho_0 c V_s a^2|$ vs ka is plotted [dotted curve, using the last formula of Eq. (31)] for the same Gaussian velocity profile as was considered in Sec. III B.

The case with $u_0=1$, $u_1=u_2=\cdots=0$, in Eqs. (30) and (31) yields the same as Eq. (26) with $n=0$. The middle expression for F in Eq. (31) can be used to express $\Re F$ and $\Im F$ in terms of Bessel functions of the first kind and Struve functions, respectively. As with the Stenzel functions in Eqs. (27)–(29), this soon gets cumbersome. Finally, the single-series expressions on the last lines of Eqs. (26) and (31) can be checked against one another based on the Zernike representation in Eq. (21) of $s^{(n)}(\sigma)$, just as this was done with the two single-series expressions for p_{edge} in Sec. III.

Equation (31) shows that the real part has a non-zero coefficient for $(ka)^2$ only when $n=0$. Accordingly, the coefficient of $(ka)^2$ in F of Eq. (30) is equal to $\frac{1}{2}\pi\rho_0 c V_s a^2$ (because $u_0=1$ by definition), no matter what $v(\sigma)$ is. This is in agreement with the observation of Greenspan,[6] Sec. IV, Eqs. (35a), (40b), and (41b), that this holds for Stenzel functions $s^{(n)}$, $n=0,1,2$.

V. POWER OUTPUT AND DIRECTIVITY

The power is defined as the intensity pv^* integrated over the plane $z=0$. Thus, because v vanishes outside S,

$$P=\int_S p(\sigma)v^*(\sigma)dS, \quad (32)$$

where $p(\sigma)=p(\sigma\cos\psi,\sigma\sin\psi,0)$ is the pressure at an arbitrary point on S. According to Eq. (3) with $z=0$ and $w=\sigma$,

$$p(\sigma)=i\rho_0 ck\int_0^{\infty}\frac{V(u)}{(u^2-k^2)^{1/2}}J_0(\sigma u)udu, \quad 0\leqslant\sigma<\infty, \quad (33)$$

assumes the form of a Hankel transform, viz., of the function $i\rho_0 ckV(u)(u^2-k^2)^{-1/2}$, where $V(u)$ is the Hankel transform of $v(\sigma)$. By using Parseval's theorem for Hankel transforms in Eq. (32), it follows that

$$P=2\pi i\rho_0 ck\int_0^{\infty}\frac{V(u)V^*(u)}{(u^2-k^2)^{1/2}}udu. \quad (34)$$

A. Stenzel functions

With $v(\sigma)=s^{(n)}(\sigma)$ and $V(u)=S^{(n)}(u)$ as given by Eqs. (6) and (7), the power $P^{(n)}$ is given by

$$\frac{P^{(n)}}{\pi\rho_0 V_s^2 a^2}=2\left((n+1)!\left(\frac{2}{a}\right)^n\right)^2 k\int_0^{\infty}\frac{J_{n+1}^2(au)}{u^{2n+1}(u^2-k^2)^{1/2}}du. \quad (35)$$

The integral at the right-hand side of Eq. (35) has been evaluated in Appendix A 1, and the result is that

$$\begin{aligned}
\frac{P^{(n)}}{\pi\rho_0 V_s^2 a^2}&=2\left((n+1)!\left(\frac{2}{a}\right)^n\right)^2 k\left[\int_0^{k}\frac{J_{n+1}^2(au)}{u^{2n+1}\sqrt{k^2-u^2}}du\right.\\
&\quad\left.+i\int_k^{\infty}\frac{J_{n+1}^2(au)}{u^{2n+1}\sqrt{u^2-k^2}}du\right]\\
&=((n+1)!2^n)^2\left[\sum_{j=0}^{\infty}\frac{(-1)^j(j+3/2)_n(ka)^{2j+2}}{\Gamma(n+j+2)\Gamma(2n+j+3)}\right.\\
&\quad\left.+i\sum_{j=0}^{\infty}\frac{(-1)^j(j+1)_n(ka)^{2j+1}}{\Gamma(n+j+3/2)\Gamma(2n+j+5/2)}\right]\\
&=-((n+1)!2^n)^2\sum_{\ell=1}^{\infty}\frac{\left(\frac{1}{2}(\ell+1)\right)_n(-ika)^{\ell}}{\Gamma\left(n+\frac{1}{2}\ell+1\right)\Gamma\left(2n+\frac{1}{2}\ell+2\right)}.
\end{aligned} \quad (36)$$

The middle expression for $P^{(n)}$ in Eq. (36) can be used to express the real and imaginary parts of $P^{(n)}$ in Bessel functions of the first kind and Struve functions, respectively. The case that $n=0$ in Eq. (36) yields the same integral as the one that occurs in the expression of Eq. (25) for $F^{(n)}$ with $n=0$, and so Eq. (27) can be used yielding

$$\frac{P^{(0)}}{\pi\rho_0 c V_s^2 a^2}=1-\frac{J_1(2z)}{z}+i\frac{\mathbf{H}_1(2z)}{z}. \quad (37)$$

For the cases $n=1,2$, Ref. 6, Eqs. (45) and (47) express $P^{(n)}$ in terms of Bessel and Struve functions; these results have been checked against what Eq. (36) gives for $n=1,2$ and complete agreement has been observed. [This check has been

carried out since, due to the very complicated nature of the resulting expressions, Greenspan had some doubts about correctness of his Eq. (47).]

The single-series expression on the last line of Eq. (36) has also been checked against Ref. 6, Eqs. (45b) and (47b), where the first few terms are displayed: there is complete agreement. A further observation is that the coefficient of $(ka)^2$ in $P^{(n)}$ equals $\frac{1}{2}\pi\rho_0 c V_s^2 a^2$ (independent of n). Thus $\Re[P^{(n)}] \approx \frac{1}{2}\pi\rho_0 c V_s^2 a^2 (ka)^2$ for small ka. In Fig. 5 the real and imaginary parts of the power of the radiator $P^{(n)}/\pi\rho_0 c V_s^2 a^2$ vs ka is plotted [using the last formula of Eq. (36)] for the rigid piston ($n=0$), the simply supported radiator ($n=1$), and the first two clamped radiators ($n=2,3$). Note that the power shown in Fig. 5 reaches a fixed value for large ka values. With the general approach developed in Secs. V B and V D [in particular, Eqs. (39), (52), and (53)], it follows that these fixed values are given as

$$\lim_{ka\to\infty} \frac{P^{(n)}}{\pi\rho_0 c V_s^2 a^2} = 2\int_0^1 [(n+1)(1-\rho^2)^n]^2 \rho d\rho = \frac{(n+1)^2}{2n+1}. \tag{38}$$

However, for real sources like loudspeakers, the power will decay. This is because the velocity of the radiator will decrease for higher frequencies, in particular, for loudspeakers above their resonance frequency.

B. Zernike functions

With $v(\sigma)$ a linear combination of Zernike functions $R_{2n}^0(\sigma/a)$ as in Eq. (9) with Hankel transform $V(u)$ as given in Eq. (10), the power P is given by

$$\frac{P}{\pi\rho_0 c V_s^2 a^2} = 2ik \sum_{n_1,n_2=0}^{\infty} (-1)^{n_1+n_2} u_{n_1} u_{n_2}^* \cdot \left[\int_0^\infty \frac{J_{2n_1+1}(ua) J_{2n_2+1}(ua)}{u(u^2-k^2)^{1/2}} du \right]. \tag{39}$$

The integrals at the right-hand side of Eq. (39) have been evaluated in Appendix A 2, with the result that

$$\begin{aligned} 2ik\int_0^\infty \frac{J_{2n_1+1}(ua)J_{2n_2+1}(ua)}{u(u^2-k^2)^{1/2}} du &= 2k\int_0^k \frac{J_{2n_1+1}(ua)J_{2n_2+1}(ua)}{u\sqrt{k^2-u^2}} du \\ &\quad + 2ik\int_k^\infty \frac{J_{2n_1+1}(ua)J_{2n_2+1}(ua)}{u\sqrt{u^2-k^2}} du \\ &= \sum_{j=0}^\infty \frac{(-1)^j (j+1)_N (j+2n_2+2)_M}{\Gamma(j+2n_1+2)\Gamma(j+2N+3)} (ka)^{2(N+j+1)} \\ &\quad + i(-1)^M \sum_{j=0}^\infty \frac{(-1)^j (-j-1/2)_M (-j+1/2)_N}{\Gamma(j+M+3/2)\Gamma(j+N+5/2)} (ka)^{2j+1} \\ &= -(-1)^M \sum_{\ell=1}^\infty \frac{\left(-\frac{1}{2}\ell\right)_M \left(-\frac{1}{2}\ell+1\right)_N (-ika)^\ell}{\Gamma\left(\frac{1}{2}\ell+M+1\right)\Gamma\left(\frac{1}{2}\ell+N+2\right)}, \end{aligned} \tag{40}$$

where $M=|n_1-n_2|$ and $N=n_1+n_2$.

The case with $u_0=1$, $u_1=u_2=\cdots=0$ in Eqs. (39) and (40) yields the same as Eq. (36) with $n=0$. The middle expression for the integral in Eq. (40) can be used to express real and imaginary parts in terms of Bessel functions of the first kind and Struve functions, respectively, but this gets out of hand quite soon. Finally, the single-series expressions in Eqs. (36), (39), and (40) can be checked against one another using the Zernike expansion for $s^{(n)}$ in Eq. (21); this has been observed to give the same results for the case $s^{(1)}$. In Fig. 3 the normalized power $|P/\pi\rho_0 c V_s^2 a^2|$ (dash-dotted curve) is plotted using Eqs. (39) and (40) for a truncated Gaussian velocity profile $\exp(-\alpha(\sigma/a)^2)$, $\alpha=2$, approximated using three Zernike coefficients $u_0=1.0000$, $u_1=-0.9392$, and $u_2=0.3044$.

The series expansion in Eq. (40) shows that the coefficient of $(ka)^2$ is non-zero for $N=n_1+n_2=0$ only [this is so since $(0)_N \neq 0$ for $N=0$ only]. Since by definition $u_0=1$, this shows that the coefficient of $(ka)^2$ in P in Eq. (39) is equal to $\frac{1}{2}\rho_0 c \pi V_s^2 a^2$, no matter what $v(\sigma)$ is.

C. Power evaluated from the far field

Usually one calculates the power of the radiator from p and v values at the radiator itself, but due to the conservation of energy the power can also be computed from sound field values at any distance from the radiator. Below it is shown that the power can be calculated in the far field with the techniques described in the preceding sections. The power as defined in Eq. (32) should be equal to the integral of pv^* over any surface Σ in $z\geqslant 0$ containing the disk $\sigma\leqslant a$. Here v and p are related to one another in $z\geqslant 0$ according to

$$v = \frac{-1}{ik\rho_0 c}\frac{\partial p}{\partial \underline{n}} \quad \underline{n} \text{ normal to } \Sigma. \tag{41}$$

The imaginary part of P, the wattless component, manifests itself only close to the radiator and has no physical significance in the far field; it is thus customary to consider the real part of P only, especially when the non-rigid part of Σ in $z>0$ is in the far field. It is shown by Bouwkamp[12] that

$$\Re\left[\int_\Sigma pv^* d\Sigma\right] = \frac{1}{2}\int_\Sigma (pv^* + p^* v) d\Sigma = 0. \tag{42}$$

Thus, taking for Σ the surface S_R of the hemisphere $x^2+y^2+z^2\leqslant R^2$, $z\geqslant 0$ with $R\geqslant a$, together with the disk $x^2+y^2\leqslant R^2$, $z=0$, one finds that

$$\Re\left[\int_S pv^* dS\right] = \Re\left[\int_{S_R} pv^* dS_R\right]. \tag{43}$$

The right-hand side of Eq. (43) will now be considered when $R\to\infty$. According to Blackstock[1] it holds in the far field that

$$p = O(1/r), \quad v = (\rho_0 c)^{-1} p(1+O(1/r)). \tag{44}$$

Hence,

$$\Re\left[\int_{S_R} pv^* dS_R\right] = (\rho_0 c)^{-1} \int_{S_R} |p|^2 dS_R + O(R^{-1}). \tag{45}$$

From the Rayleigh representation of p in Eq. (1), it follows with the usual approximation arguments that

$$p(\underline{r}) \approx i\rho_0 ck \frac{e^{-ikr}}{r} \int_0^a v(\sigma) J_0(k\sigma \sin\theta) \sigma d\sigma$$
$$= i\rho_0 ck \frac{e^{-ikr}}{r} V(k \sin\theta) \tag{46}$$

with V the Hankel transform of v as before. Then taking spherical coordinates in the integral at the right-hand side of Eq. (45) and letting $R \to \infty$, it follows that

$$\Re\left[\int_{S_R} pv^* dS\right] = 2\pi\rho_0 ck^2 \int_0^{\pi/2} |V(k \sin\theta)|^2 \sin\theta d\theta. \tag{47}$$

By changing integration variables in the integral at the right-hand side of Eq. (47) according to $u = k \sin\theta$, $0 \leqslant u \leqslant k$, the final result becomes

$$\Re\left[\int_{S_R} pv^* dS\right] = 2\pi\rho_0 ck \int_0^k \frac{|V(u)|^2}{\sqrt{k^2 - u^2}} u du. \tag{48}$$

Compare Eq. (34).

D. Directivity

From the far-field expression in Eq. (46) for $p(\underline{r})$, $\underline{r} = (r\cos\psi \sin\theta, r \sin\psi \sin\theta, r\cos\theta)$, there results the directivity

$$D = \frac{4\pi |V(0)|^2}{\int_0^{2\pi} \int_0^{\pi/2} |V(k \sin\theta)|^2 \sin\theta d\psi d\theta}$$
$$= \frac{2|V(0)|^2}{\int_0^{\pi/2} |V(k \sin\theta)|^2 \sin\theta d\theta}, \tag{49}$$

see Ref. 2, Sec. 8.9. This gives rise to the same integral as in Eq. (47). By Eqs. (2) and (5) it holds that $V(0) = \frac{1}{2} a^2 V_s$, and by Eq. (34) it holds that

$$\int_0^{\pi/2} |V(k \sin\theta)|^2 \sin\theta d\theta = \frac{1}{2\pi\rho_0 ck^2} \Re[P]. \tag{50}$$

In Fig. 6 the directivity index ($\mathrm{DI} = 10 \log_{10} D$) vs ka is plotted [using Eq. (36), last formula, and Eqs. (49) and (50)] for the rigid piston ($n=0$), the simply supported radiator ($n=1$), and the first two clamped radiators ($n=2,3$).

Consider the case that $ka \to 0$. By the observation at the end of Secs. V A and V B, it holds that $\Re[P] \approx \frac{1}{2}\pi\rho_0 c V_s^2 a^2 (ka)^2$. Therefore, as $ka \to 0$

$$D \approx \frac{2\left(\frac{1}{2}a^2 V_s\right)^2}{\frac{1}{2\pi\rho_0 ck^2} \frac{1}{2} \pi\rho_0 c V_s^2 a^2 (ka)^2} = 2, \tag{51}$$

or 3 dB, which is the same for a rigid piston[2] or a hemispherical source on an infinite baffle and this is supported by Fig. 6.

Next consider the case that $ka \to \infty$, in the general setting of Sec. V B. Now, by Ref. 22, 11.4.6,

$$\int_0^k \frac{J_{2n_1+1}(ua) J_{2n_2+1}(ua)}{u\sqrt{k^2 - u^2}} du = a \int_0^{ka} \frac{J_{2n_1+1}(v) J_{2n_2+1}(v)}{v\sqrt{(ka)^2 - v^2}} dv$$
$$\approx \frac{a}{ka} \int_0^\infty \frac{J_{2n_1+1}(v) J_{2n_2+1}(v)}{v} dv$$
$$= \frac{1}{k} \frac{\delta n_1 n_2}{2(2n_1+1)}. \tag{52}$$

Thus from Eq. (39)

$$\Re[P] \approx 2\pi\rho_0 c V_s^2 a^2 \sum_{n=0}^\infty \frac{|u_n|^2}{2(2n+1)}$$
$$= 2\pi\rho_0 c a^2 \int_0^1 |v(a\rho)|^2 \rho d\rho, \tag{53}$$

where Parseval's theorem for Zernike expansions $v(\sigma) = V_s \Sigma_{n=0}^\infty u_n R_{2n}^0(\sigma/a)$ has been used. It thus follows that, as $ka \to \infty$,

$$D \approx \frac{2\left(\frac{1}{2}a^2 V_s\right)^2}{\frac{1}{2\pi\rho_0 ck^2} 2\pi\rho_0 c a^2 \int_0^1 |v(a\rho)|^2 \rho d\rho}$$
$$= \frac{\frac{1}{2}(ka)^2 V_s^2}{\int_0^1 |v(a\rho)|^2 \rho d\rho}. \tag{54}$$

In case that $v = s^{(n)}$, the last member of Eq. (54) is given by $(2n+1)(n+1)^{-1}(ka)^2$; in Ref. 2, end of Sec. 8.9, the result for the case $n=0$ is given.

VI. ESTIMATING POWER FROM NEAR-FIELD ON-AXIS MEASUREMENTS

In Ref. 21, a method has been introduced recently to estimate a radially symmetric velocity profile $v(\sigma)$, $0 \leqslant \sigma \leqslant a$, from on-axis sound pressure data, in terms of Zernike expansion coefficients. The basis of this method is the explicit formula, see Ref. 21, Eq. (17),

$$p((0,0,r)) = \tfrac{1}{2}\rho_0 c V_s (ka)^2 \sum_{n=0}^\infty \gamma_n(k,r) u_n \tag{55}$$

for the on-axis pressure, where the $\gamma_n(k,r)$ are explicitly given in terms of spherical Bessel and Hankel functions. The formulas in Sec. V B then show how these estimated coefficients give rise to a means to compute the power and directivity.

VII. CONCLUSIONS

Greenspan's results[6] on acoustic quantities related to the pressure due to a velocity profile on a piston radiator in an infinite baffle are treated in a unified way. By expanding the velocity profile in terms of Zernike functions, the pressure at the edge of a radiator, the reaction force on the radiator, the power output, and directivity of the radiator can be expressed in an attractive way as power series in ka. Since many velocity profiles have a representation in terms of Zernike functions with explicitly computable coefficients, the results of this paper constitute a considerable generalization of Greenspan's results.

APPENDIX A: EVALUATION OF BESSEL INTEGRALS

In this appendix, the integrals

$$i\int_0^\infty \frac{J_{m+1}(au)J_{n+1}(au)}{u^{n+m+1}(u^2-k^2)^{1/2}}du \tag{A1}$$

are evaluated. These integrals occur in relation to the pressure at the edge ($m=-1$), the reaction on the radiator ($m=0$), and the power if the velocity profile $v(\sigma)$ is a Stenzel function $s^{(n)}(\sigma)$ or a linear combination of Stenzel functions. Furthermore, the integrals

$$i\int_0^\infty \frac{J_m(au)J_{n+1}(au)}{(u^2-k^2)^{1/2}}du, \tag{A2}$$

$$i\int_0^\infty \frac{J_{m+1}(au)J_{n+1}(au)}{u(u^2-k^2)^{1/2}}du, \tag{A3}$$

with integer n, $m\geqslant 0$ such that $n-m$ even and $\geqslant 0$ are evaluated. These integrals occur in connection with the pressure at the edge [$m=0$ in Eq. (A2)], reaction on the radiator [$m=0$ in Eq. (A3)], and the power [general m, n in Eq. (A3)] if the velocity profile is a linear combination of Zernike functions.

The integrals are evaluated in the form $\Re+i\Im$, where $\Re$ and $\Im$ arise from the integration ranges $[0,k]$ and $[k,\infty)$, respectively. These $\Re$ and $i\Im$ parts are given as power series in ka in a form from which the close relationship with Bessel functions of the first kind,

$$J_\nu(x)=\left(\frac{1}{2}x\right)^\nu\sum_{j=0}^\infty\frac{\left(-\frac{1}{4}x^2\right)^j}{\Gamma(j+1)\Gamma(j+\nu+1)}, \tag{A4}$$

and Struve functions,

$$\mathbf{H}_\nu(x)=\left(\frac{1}{2}x\right)^{\nu+1}\sum_{j=0}^\infty\frac{\left(-\frac{1}{4}x^2\right)^j}{\Gamma\left(j+\frac{3}{2}\right)\Gamma\left(j+\nu+\frac{3}{2}\right)}, \tag{A5}$$

is apparent. Furthermore, the two series for $\Re$ and $i\Im$ are reorganized and combined into a concise single power series in $-ika$ for the various integrals.

A formula in terms of hypergeometric functions ${}_3F_4$ of the integrals

$$\left(\int_0^k+\int_k^\infty\right)u^{\alpha-1}(u^2-k^2)^{\beta-1}J_\mu(au)J_\nu(au)du \tag{A6}$$

can be found in Ref. 30, Sec. 2.12.32, items 3 and 8. These formulas are quite complicated; there is no indication of where a proof can be found, several degenerations and simplifications occur due to special values of α,β,μ,ν to which Eqs. (A1)–(A3) restrict, and no attention is paid to bringing the results into an attractive form. The results obtained here have been checked against the results in Ref. 31 for the special values of α,β,μ,ν that occur here. The method of the proofs used is taken from Ref. 31, Sec. 13.6.

1. Evaluation of Equation (A1)

Let n, $m=0,1,\ldots$. The case where $m=-1$ in Eq. (A1) is dealt with at the end of this section. It holds that

$$i\int_0^\infty\frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}(u^2-k^2)^{1/2}}du=\int_0^k\frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{k^2-u^2}}du+i\int_k^\infty\frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{u^2-k^2}}du. \tag{A7}$$

There is the integral representation, see Ref. 31, Sec. 13.6,

$$J_\mu(au)J_\nu(ua)=\frac{1}{2\pi i}\int_{-\infty i}^{\infty i}\frac{\Gamma(-s)\Gamma(\mu+\nu+2s+1)\left(\frac{1}{2}au\right)^{\mu+\nu+2s}}{\Gamma(\mu+s+1)\Gamma(\nu+s+1)\Gamma(\mu+\nu+s+1)}ds, \tag{A8}$$

where the integration contour has the poles of $\Gamma(-s)$ on its right and those of $\Gamma(\mu+\nu+2s+1)$ on its left (thus $-\frac{1}{2}(\mu+\nu+1)<\Re(s)<0$).

For the first integral in Eq. (A7), second line, the result of Eq. (A8) is used together with

$$\int_0^k\frac{u^\alpha}{\sqrt{k^2-u^2}}du=\frac{1}{2}k^\alpha\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\alpha+\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\alpha+1\right)},\quad \Re(\alpha)>-1. \tag{A9}$$

This yields

$$\int_0^k\frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{k^2-u^2}}du=\frac{1}{2}\Gamma\left(\frac{1}{2}\right)\left(\frac{1}{2}a\right)^{n+m+1}\frac{1}{2\pi i}\times\int_{-\infty i}^{\infty i}\frac{\Gamma(-s)\Gamma(n+m+2s+3)\left(\frac{1}{2}ka\right)^{2s+1}}{\Gamma(n+s+2)\Gamma(m+s+2)\Gamma(n+m+s+3)}\times\frac{\Gamma(s+1)}{\Gamma\left(s+\frac{3}{2}\right)}ds, \tag{A10}$$

where the two occurring integrals have been interchanged. The integration in Eq. (A10) is closed to the right, thereby enclosing all poles of $\Gamma(-s)$ at $s=j$ with residues $(-1)^{j+1}/j!$, and it follows from Cauchy's theorem that

$$\int_0^k \frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{k^2-u^2}}du$$

$$= \frac{1}{2}\Gamma\left(\frac{1}{2}\right)\left(\frac{1}{2}a\right)^{n+m+1}\sum_{j=0}^{\infty}\frac{\Gamma(2j+n+m+3)}{\Gamma(j+m+2)\Gamma(j+3/2)}$$

$$\times\frac{(-1)^j\left(\frac{1}{2}ka\right)^{2j+1}}{\Gamma(j+n+2)\Gamma(j+n+m+3)}. \quad \text{(A11)}$$

Now assume that $n \geqslant m$. Then it holds that

$$\frac{\Gamma(n+m+2j+3)}{\Gamma(m+j+2)\Gamma(j+3/2)} = \frac{2^{2m+2j+2}}{\Gamma\left(\frac{1}{2}\right)}(2m+2j+3)_{n-m}\left(j+\frac{3}{2}\right)_m, \quad \text{(A12)}$$

where Pochhammer's symbol $(x)_\ell$ has been used,

$$(x)_\ell = \frac{\Gamma(x+\ell)}{\Gamma(x)}, \quad (x)_0 = 1,$$

$$(x)_\ell = x(x+1)\cdot\cdots\cdot(x+\ell-1), \quad \ell = 1,2,\ldots. \quad \text{(A13)}$$

Therefore,

$$\int_0^k \frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{k^2-u^2}}du$$

$$= 2^{m-n-1}a^{n+m+1}\sum_{j=0}^{\infty}\frac{(-1)^j(2m+2j+3)_{n-m}(j+3/2)_m}{\Gamma(j+n+2)\Gamma(j+n+m+3)}$$

$$\times(ka)^{2j+1}. \quad \text{(A14)}$$

Next, for the second integral on the second line of Eq. (A7), the plan of the proof is the same, except that now

$$\int_k^\infty \frac{u^\alpha}{\sqrt{u^2-k^2}}du = \frac{1}{2}k^\alpha\frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(-\frac{1}{2}\alpha\right)}{\Gamma\left(\frac{1}{2}-\frac{1}{2}\alpha\right)}, \quad \Re(\alpha) < 0 \quad \text{(A15)}$$

is used. This yields the same expression as in Eq. (A10), except that the $\Gamma(s+1)/\Gamma(s+3/2)$ just in front of ds has to be replaced by $\Gamma(-s-1/2)/\Gamma(-s)$, thereby canceling the $\Gamma(-s)$ just behind the integral sign. Now the poles of $\Gamma(-s-1/2)$ at the points $s=j-1/2$ have to be taken into account, and this yields [using Eq. (A12) with $j-1/2$ instead of j]

$$\int_k^\infty \frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}\sqrt{u^2-k^2}}du$$

$$= 2^{m-n-1}a^{n+m+1}\sum_{j=0}^{\infty}\frac{(-1)^j(2j+2m+2)_{n-m}(j+1)_m}{\Gamma(j+n+3/2)\Gamma(j+n+m+5/2)}(ka)^{2j}. \quad \text{(A16)}$$

This yields a power series for the real part in Eq. (A7) per Eq. (A14) and the imaginary part in Eq. (A7) per Eq. (A16) that shows a close relationship with the Bessel and Struve functions in Eqs. (A4) and (A5), respectively. It is actually possible to express the results of Eqs. (A14) and (A16) systematically in terms of Bessel and Struve functions by manipulating the polynomial of degree n in j occurring in the numerator of the coefficients in the series in Eqs. (A14) and (A16).

The case that $m=-1$ can be dealt with in a completely similar fashion, except that in Eq. (A12) the definition $(x)_{-1}=(x-1)^{-1}$ of Pochhammer's symbol with subscript $\ell=-1$ should be used [this is consistent with the Γ-function based definition in Eq. (A13)]. With some further rewriting, this then yields the identity of the quantities in the second and third members between [ ] in Eq. (14) in the main text.

A single power series in $-ika$ for Eq. (A7) follows on combining Eqs. (A14) and (A16). Doing the administration with the $(-1)^j=(-i)^{2j}$, there results

$$i\int_0^\infty \frac{J_{n+1}(au)J_{m+1}(au)}{u^{n+m+1}(u^2-k^2)^{1/2}}du$$

$$= -\frac{1}{k}2^{m-n-1}a^{n+m}$$

$$\times\sum_{\ell=1}^{\infty}\frac{(\ell+2m+1)_{n-m}\left(\frac{1}{2}(\ell+1)\right)_m}{\Gamma\left(\frac{1}{2}\ell+n+1\right)\Gamma\left(\frac{1}{2}\ell+n+m+2\right)}(-ika)^\ell. \quad \text{(A17)}$$

2. Evaluation of Equations (A2) and (A3)

Let n, m be integers $\geqslant 0$ with $n-m$ even and $\geqslant 0$. The proof for the integral in Eq. (A2) follows the same reasoning as for the integral in Eq. (A1). Letting

$$p = \frac{n-m}{2}, \quad q = \frac{n+m}{2}, \quad \text{(A18)}$$

it is found without any particular problem that

$$\int_0^k \frac{J_{n+1}(au)J_m(au)}{\sqrt{k^2-u^2}}du$$

$$= \frac{1}{2}\sum_{j=0}^{\infty}\frac{(-1)^j(j+1)_q(j+m+1)_p(ka)^{2(j+q)+1}}{\Gamma(n+j+2)\Gamma(j+2q+2)}. \quad \text{(A19)}$$

For the integration range $[k,\infty)$ in the integral in Eq. (A2), the method based on Eqs. (A8) and (A15) yields an expression as in the second member of Eq. (A10), except that the $\Gamma(s+1)/\Gamma(s+3/2)$ appearing in front of ds should be replaced by $\Gamma(-s-q-1/2)/\Gamma(-s+q)$, thereby canceling all poles of the $\Gamma(-s)$ appearing just behind the integral sign in Eq. (A10). It is then found upon some further administration with Γ-functions that

$$\int_k^\infty \frac{J_{n+1}(au)J_m(au)}{\sqrt{u^2-k^2}}du$$

$$= \frac{1}{2}\sum_{k=0}^{\infty}\frac{(-1)^{j+p}(-j+1/2)_q(-j+1/2)_p}{\Gamma(j+p+3/2)\Gamma(j+q+3/2)}(ka)^{2j}. \quad \text{(A20)}$$

From Eqs. (A19) and (A20), a single-series expression for Eq. (A2) can be derived. To this end, the summation index j in Eq. (A19) is changed into $j-q-1$ with $j=q+1$, $q+2$,…. Next, it is observed that $n-q=p$, $m-q=-p$, and that

$$(j-q)_q=(-1)^q(-j+1)_q, \quad (j-p)_p=(-1)^p(-j+1)_p. \tag{A21}$$

As a consequence of $(-j+1)_q=0$ for $j=1,\ldots,q$ it follows that the new summation index j can be taken to range from 1 to ∞. This all yields the result

$$\int_0^k \frac{J_{n+1}(au)J_m(au)}{\sqrt{k^2-u^2}}du = -\frac{(-1)^p}{2ka}\sum_{j=1}^{\infty}\frac{(-1)^j(-j+1)_p(-j+1)_q(ka)^{2j}}{\Gamma(j+p+1)\Gamma(j+q+1)}. \tag{A22}$$

Then combining Eqs. (A20) and (A22) while administrating the $(-1)^j=(-i)^{2j}$ yields the single-series expression

$$i\int_0^{\infty} \frac{J_{n+1}(au)J_m(au)}{(u^2-k^2)^{1/2}}du = -\frac{(-1)^p}{2ka}\sum_{\ell=1}^{\infty}\frac{\left(-\frac{1}{2}\ell+1\right)_p\left(-\frac{1}{2}\ell+1\right)_q}{\Gamma\left(\frac{1}{2}\ell+p+1\right)\Gamma\left(\frac{1}{2}\ell+q+1\right)}(-ika)^{\ell}. \tag{A23}$$

The treatment of the integral in Eq. (A3) is entirely similar. There results

$$\begin{aligned} i\int_0^{\infty} \frac{J_{n+1}(au)J_{m+1}(au)}{u(u^2-k^2)^{1/2}}du &= \int_0^k \frac{J_{n+1}(au)J_{m+1}(au)}{u\sqrt{k^2-u^2}}du + i\int_k^{\infty} \frac{J_{n+1}(au)J_{m+1}(au)}{u\sqrt{u^2-k^2}}du \\ &= \frac{1}{2k}\sum_{j=0}^{\infty}\frac{(-1)^j(j+1)_q(j+m+1)_p}{\Gamma(j+n+2)\Gamma(j+2q+3)}(ka)^{2(q+1+j)} \\ &\quad + \frac{1}{2}ia(-1)^p\sum_{j=0}^{\infty}\frac{(-1)^j(-j-1/2)_p(-j+1/2)_q}{\Gamma(j+p+3/2)\Gamma(j+q+5/2)}(ka)^{2j} \\ &= -\frac{(-1)^p}{2k}\sum_{\ell=1}^{\infty}\frac{\left(-\frac{1}{2}\ell\right)_p\left(-\frac{1}{2}\ell+1\right)_q}{\Gamma\left(\frac{1}{2}\ell+p+1\right)\Gamma\left(\frac{1}{2}\ell+q+2\right)}(-ika)^{\ell}. \end{aligned} \tag{A24}$$

Here n and m are integers $\geqslant 0$ with $n-m$ even and $\geqslant 0$, and p and q given by Eq. (A18).

The results for the integrals in Eqs. (A2) and (A3) just given have been checked by using[22] $z^{-1}J_{m+1}(z)=(2(m+1))^{-1}(J_m(z)+J_{m+2}(z))$ in Eq. (A3).

APPENDIX B: CONVERGENCE ANALYSIS OF THE SERIES

The series in Eqs. (14), (26), and (36) for the case of Stenzel functions on one hand and those in Eqs. (20), (31), and (40) for the case of Zernike functions on the other are all of a very similar nature with regard to convergence and accuracy matters. It therefore suffices to consider only the series in Eq. (14) and the series in Eq. (20). Next, from a simple comparison of the coefficients in the two series

$$\sum_{\ell=1}^{\infty}\frac{(\ell)_n z^{\ell}}{\Gamma^2\left(n+\frac{1}{2}\ell+1\right)} \text{ and } \sum_{\ell=1}^{\infty}\left(\frac{\left(-\frac{1}{2}\ell+1\right)_n}{\Gamma\left(n+\frac{1}{2}\ell+1\right)}\right)^2 z^{\ell}, \tag{B1}$$

it is seen that for either series $n=0$ is worst case. The series to be considered becomes for either case

$$S(z=-ika)=\sum_{\ell=1}^{\infty}\frac{z^{\ell}}{\Gamma^2\left(\frac{1}{2}\ell+1\right)}. \tag{B2}$$

From Stirling's formula $\Gamma(x+1)\approx e^{-x}x^{x+1/2}\sqrt{2\pi}$ it readily follows that the modulus of the terms t_ℓ in the series in Eq. (B2) is accurately estimated by

$$|t_\ell|\approx m_\ell=\frac{1}{\pi\ell}\left(\frac{2e|z|}{\ell}\right)^{\ell}. \tag{B3}$$

Hence, m_ℓ is of order unity and less from $\ell=2e|z|$ onwards. Next, setting $\ell=2e|z|+b$ with $0\leqslant b\leqslant \ell$, it follows that

$$m_\ell=\frac{1}{\pi\ell}\left(1-\frac{b}{\ell}\right)^{\ell}\leqslant\frac{1}{\pi\ell}e^{-b}=\frac{1}{\pi\ell}e^{2e|z|-\ell}. \tag{B4}$$

Hence, for integer $L\geqslant 2e|z|$ it holds that

$$\sum_{\ell=L}^{\infty}m_\ell\leqslant\frac{e^{2e|z|-L}}{\pi L(e-1)}. \tag{B5}$$

When, for instance, $L\geqslant 2e|z|+10$, the absolute accuracy by including in the series in Eq. (B2) the terms with $\ell=1,\ldots,L-1$ is at least 10^{-6}. Note that the true value of $S(z=-ika)$ is of order unity, see Fig. 2. However, loss-of-digits occurs in summing the series. The maximum value over $\ell=1,2,\ldots$ of m_ℓ in Eq. (B3) is assumed near $2|z|$ and is accurately given by $e^{2|z|}/2\pi|z|$. For instance, when $|z|\leqslant 12$ (which is the case in all figures) this maximum is $\leqslant 3.5\times 10^8$. As a result, when machine precision is 10^{-15}, an absolute accuracy of 10^{-6} is obtained in all cases when $0\leqslant ka\leqslant 12$ assuming that the series is truncated at an integer $L\geqslant 2eka+10$.

[1] D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000).
[2] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics* (Wiley, New York, 1982).
[3] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[4] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America through the American Institue of Physics, New York, 1989).
[5] H. Stenzel and O. Brosze, *Guide to Computation of Sound Phenomena (Leitfaden zur Berechnung von Schallvorgängen)*, 2nd ed. (Springer-Verlag, Berlin, 1958).
[6] M. Greenspan, "Piston radiator: Some extensions of the theory," J. Acoust. Soc. Am. **65**, 608–621 (1979).
[7] G. R. Harris, "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20 (1981).
[8] H. Stenzel, "On the acoustical radiation of membranes (published in German as Über die akustische Strahlung von Membranen)," Ann. Phys. **7**, 947–982 (1930).
[9] L. V. King, "On the acoustic radiation field of the piezo-electric oscillator and the effect of viscosity on transmission," Can. J. Res. **11**, 135–155 (1934).
[10] A. Schoch, "Contemplations on the sound field of piston membranes (published in German as Betrachtungen über das Schallfeld einer Kolbenmembran)," Akust. Z. **6**, 318–326 (1941).
[11] J. A. Archer-Hall, A. I. Bashter, and A. J. Hazelwood, "A means for computing the Kirchhoff surface integral for a disk radiator as a single

integral with fixed limits," J. Acoust. Soc. Am. **65**, 1568–1570 (1979).
[12]C. J. Bouwkamp, "A contribution to the theory of acoustic radiation," Philips Res. Rep. **1**, 251–277 (1946).
[13]F. Oberhettinger, "On transient solutions of the "baffled piston" problem," J. Res. Natl. Bur. Stand., Sect. B **65B**, 1–6 (1961).
[14]T. Hansen, "Probe-corrected near-field measurements on a truncated cylinder," J. Acoust. Soc. Am. **119**, 792–807 (2006).
[15]D. A. Hutchins, H. D. Mair, P. A. Puhach, and A. J. Osei, "Continuous-wave pressure fields of ultrasonic transducers," J. Acoust. Soc. Am. **80**, 1–12 (1986).
[16]R. J. McGough, T. V. Samulski, and J. F. Kelly, "An efficient grid sectoring method for calculations of the near field pressure generated by a circular piston," J. Acoust. Soc. Am. **115**, 1942–1954 (2004).
[17]J. F. Kelly and R. J. McGough, "An annular superposition integral for axisymmetric radiators," J. Acoust. Soc. Am. **121**, 759–765 (2007).
[18]T. D. Mast and F. Yu, "Simplified expansions for radiation from a baffled circular piston," J. Acoust. Soc. Am. **118**, 3457–3464 (2005).
[19]T. Mellow, "On the sound field of a resilient disk in an infinite baffle," J. Acoust. Soc. Am. **120**, 90–101 (2006).
[20]R. C. Wittmann and A. D. Yaghjian, "Spherical-wave expansions of piston-radiator fields," J. Acoust. Soc. Am. **90**, 1647–1655 (1991).
[21]R. M. Aarts and A. J.E.M. Janssen, "On-axis and far-field sound radiation from resilient flat and dome-shaped radiators," J. Acoust. Soc. Am. **125**, 1444–1455 (2009).
[22]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).
[23]T. Hasegawa, N. Inoue, and K. Matsuzawa, "A new rigorous expansion for the velocity potential of a circular piston source," J. Acoust. Soc. Am. **74**, 1044–1047 (1983).
[24]N. W. McLachlan, "The acoustic and inertia pressure at any point on a vibrating circular disk," Philosophical Magazine and Journal of Science **14**, 1012–1025 (1932).
[25]F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode (Diffraction theory of the knife-edge test and its improved version, the phase-contrast method)," Physica (Amsterdam) **1**, 689–704 (1934).
[26]J. W.S. Rayleigh, *The Theory of Sound* (Dover, New York, 1896), Vol. **2**.
[27]B. R.A. Nijboer, "The diffraction theory of aberrations," Ph.D. thesis, University of Groningen, The Netherlands, 1942.
[28]A. G. Warren, "A note on the acoustic pressure and velocity relations on a circular disc and in a circular orifice," Proc. Phys. Soc. London **40**, 296–299 (1928).
[29]R. M. Aarts and A. J.E.M. Janssen, "Approximation of the Struve function $\mathbf{H}_1(z)$ occurring in impedance calculations," J. Acoust. Soc. Am. **113**, 2635–2637 (2003).
[30]A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series, Volume 2: Special Functions*, (Gordon and Breach Science, New York, 1986).
[31]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge University Press, Cambridge, 1944).

Hybrid method for determining the parameters of condenser microphones from measured membrane velocities and numerical calculations[a)]

Salvador Barrera-Figueroa[b)] and Knud Rasmussen
Danish Primary Laboratory of Acoustics, Danish Fundamental Metrology Ltd., Matematiktorvet 307, 2800 Kongens Lyngby, Denmark

Finn Jacobsen
Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads 352, 2800 Kgs. Lyngby, Denmark

(Received 27 February 2009; revised 15 June 2009; accepted 20 July 2009)

Typically, numerical calculations of the pressure, free-field, and random-incidence response of a condenser microphone are carried out on the basis of an assumed displacement distribution of the diaphragm of the microphone; the conventional assumption is that the displacement follows a Bessel function. This assumption is probably valid at frequencies below the resonance frequency. However, at higher frequencies the movement of the membrane is heavily coupled with the damping of the air film between membrane and backplate and with resonances in the back chamber of the microphone. A solution to this problem is to measure the velocity distribution of the membrane by means of a non-contact method, such as laser vibrometry. The measured velocity distribution can be used together with a numerical formulation such as the boundary element method for estimating the microphone response and other parameters, e.g., the acoustic center. In this work, such a hybrid method is presented and examined. The velocity distributions of a number of condenser microphones have been determined using a laser vibrometer, and these measured velocity distributions have been used for estimating microphone responses and other parameters. The agreement with experimental data is generally good. The method can be used as an alternative for validating the parameters of the microphones determined by classical calibration techniques.
 [DOI: 10.1121/1.3203939]

PACS number(s): 43.38.Kb, 43.38.Bs [AJZ] Pages: 1788–1795

I. INTRODUCTION

The numerical calculation of pressure, free-field, and random-incidence responses of microphones has become a popular method for validating results obtained experimentally. Furthermore, numerical calculations are sometimes used to complement experimental results at frequencies where the experimental methods might yield unreliable results.[1–5] However, the numerical calculations are usually carried out under a number of assumptions that are not necessarily completely realistic. Several attempts to develop a complete coupled model of a condenser microphone numerically are described in the literature.[6–8] Unfortunately, determining the velocity of the membrane numerically has proven to be an elusive task. Whereas complex geometries and configurations can easily be simulated, other parameters such as the velocity distribution of the membrane of a microphone are usually assumed to have a well defined analytical form. However, experimental results indicate that the velocity of the membrane may have a quite different shape.

A few years ago, Behler and Vorländer proposed an alternative solution that consists of measuring the velocity of the membrane of the microphone using a non-contact method, laser vibrometry. The measured membrane velocity was used in determining the "monopole sensitivity" of the microphone (i.e., the pressure sensitivity) and also combined with a numerical model of the effect of the body of the microphone and used for determining the free-field sensitivity.[9] However, in the authors' words, "The scope [of their investigation was] not to achieve maximum accuracy (…) but to illustrate whether the method is worth to be checked for qualification as precision method or not," and thus the results were not compared with precision data.

This paper pursues the same idea and examines the possibility of using the measured velocity of the membrane of a microphone for determining a number of microphone parameters. The velocity of the membrane of laboratory standard (LS) microphones has been measured using a laser vibrometer, and this velocity has been used in a boundary element method (BEM) model of a microphone as the boundary condition at the membrane of the microphone. The acoustic center, the free-field correction, the pressure sensitivities, and directivity index of several different types of microphones have been determined from the calculated pressure on the membrane and the sound field surrounding the microphone.

[a)] Portions of this work were presented in "On determination of microphone response and other parameters by a hybrid experimental and numerical method," Proceedings of Acoustics'08, Paris, France, June 2008.

[b)] Author to whom correspondence should be addressed. Electronic mail: sbf@dfm.dtu.dk

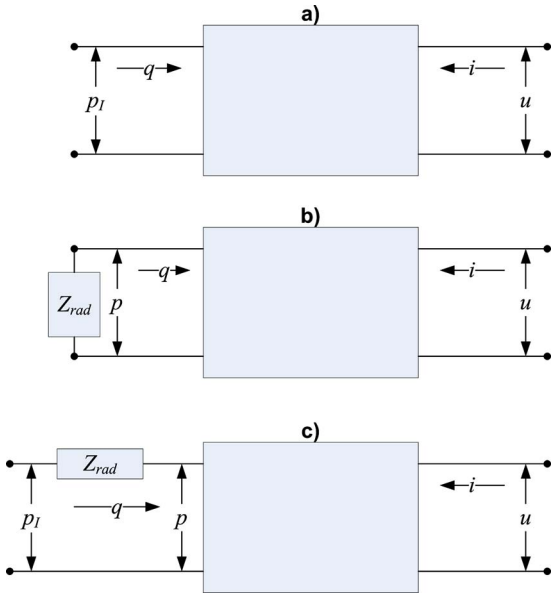


FIG. 1. (Color online) Network representation of a microphone. (a) Unloaded microphone, (b) microphone used as a sound source, and (c) microphone used as a receiver.

II. THEORETICAL BACKGROUND

A condenser microphone is a reciprocal transducer the behavior of which can be defined in terms of the equations of a four port electro-acoustic network. A graphic representation of the network of the microphone is shown in Fig. 1.

Because a condenser microphone is a reciprocal transducer, the open-circuit pressure sensitivity of the microphone, M_p, can be determined from the canonical equations of the network,[10–12]

$$M_p = \frac{u_{i=0}}{p} = -\frac{q_{p=0}}{i}, \tag{1}$$

where $u_{i=0}$ is the open-circuit voltage (no electric load), $q_{p=0}$ is the volume velocity under conditions of no acoustic load, p is the pressure on the acoustic terminal, and i is the current. If the microphone acts as a sound source, the ratio of the volume velocity to the current will be affected by the load of the radiation impedance,

$$-\frac{q}{i} = M_p \frac{Z_a}{Z_a + Z_{\text{rad}}}, \tag{2}$$

where Z_a is the acoustic impedance of the microphone and Z_{rad} is the radiation impedance.

A. Acoustic center

The concept of acoustic center has been widely used in the development and practical realization of free-field reciprocity calibration of microphones.[12–14] In this context, the microphone is substituted by a point source or a point receiver located at the position of the acoustic center. Thus, the accuracy of the estimated free-field sensitivity of a measurement microphone depends on the accuracy of the position of the acoustic center. In an international standard, the acoustic center of a microphone is defined as follows: "For a sound emitting transducer, for a sinusoidal signal of given frequency and for a specified direction and distance, the point from which the approximately spherical wavefronts, as observed in a small region around the observation point, appear to diverge."[12]

The acoustic center of LS microphones has been determined using different methodologies: (a) from the decay of the sound pressure with the distance when a condenser microphone is used as a transmitter and the sound pressure was measured using a probe microphone,[15] (b) from measurements of the transfer impedance between two microphones at different distances,[3] and (c) from measuring the decay of the sound pressure with the distance when the sound field is generated by a source of known acoustic center and the decay of the sound pressure is measured with the microphone under test.[16]

Method (a) provided values of the acoustic center that were in good agreement with the expected theoretical results, but there were significant variations due to reflections in the anechoic chamber. The acoustic centers determined using method (b) combined with a time-selective technique that removed the influence of reflections provided values of the acoustic center that were in good agreement with numerical BEM estimates. However, method (b) may not give reliable values at low frequencies because of the application of the time-selective procedure introduces some unwanted ripple at the extremes of the frequency range. Method (c) is designed to provide the acoustic center at low frequencies because the development is based on assumptions about the membrane of the source having a simple theoretical form.

The experimental values presented in Ref. 3 showed discrepancies at high frequencies with an estimate of the acoustic center determined numerically using BEM. It was concluded that such a discrepancy might be due to a possible difference between the simple analytical velocity distribution of the membrane of the microphone used in the BEM calculations and the actual velocity distribution of the membrane of the microphone. However, using the measured velocity distribution together with BEM implies that no theoretical guess of how the membrane moves is needed.

It can be assumed that the microphones can be regarded as axi-symmetric sources. Under this assumption, the acoustic center must be somewhere on the axis. If the amplitude of the sound pressure is plotted as a function of the distance, a straight line can be fitted over the region of concern. Thus, the position of the acoustic center, $x(k,r)$, can be determined using the expression

$$x(k,r) = r + |p(r)|/(\partial|p(r)|/\partial r), \tag{3}$$

where k is the wave number, r is the axial distance from the diaphragm of the microphone, $p(r)$ is the sound pressure as a function of distance, and the rate of change, $\partial|p(r)|/\partial r$, must be estimated by any available means, for example, by using least-squares fitting.[3]

B. Free-field correction

The free-field sensitivity of microphones can be determined directly for each microphone using either primary (reciprocity) or secondary (comparison) techniques. The ap-

plication of these techniques requires either expensive experimental facilities such as an anechoic room or analysis techniques that simulate free-field conditions. This makes it difficult to determine the free-field sensitivity. A common alternative is to determine the free-field sensitivity as the sum of the pressure sensitivity (in decibel) and a free-field correction.[17,18] The free-field correction, C_{ff}, is defined as the logarithmic ratio of the free-field sensitivity to the pressure sensitivity,

$$C_{\mathrm{ff}} = 10 \log\{|M_{\mathrm{ff}}|^2/|M_p|^2\}, \tag{4}$$

where M_{ff} is the free-field sensitivity and M_p is the pressure sensitivity of the microphone. The free-field correction can also be defined as the product of the free-field sensitivity (in linear units), the diffraction factor, and the load of the radiation impedance on the radiation impedance of the microphone.[12] Therefore, the free-field correction will mainly depend on the geometry and acoustic impedance of the microphone. Thus, it remains unchanged if the microphone does not suffer from any change in its geometry or impedance. Additionally, for a particular type of microphone, the geometry and impedance variations are small enough to assume that the free-field correction of a typical microphone of such a type is valid for all microphones of the same type within a stated uncertainty. In any case, the determination of the free-field correction involves measurement of the pressure and free-field sensitivities of the microphones.

Alternatively, the free-field correction can be determined using the expression[1]

$$C_{\mathrm{ff}} = 20 \log_{10}\left\{\int (p(r)/p_0)v(r)rdr/\int v(r)rdr\right\}, \tag{5}$$

where $p(r)$ is the pressure on the membrane as a function of the radius r, $v(r)$ is the velocity of the membrane as a function of r, and p_0 is the undisturbed incident pressure. The pressure on the membrane can be determined numerically when the velocity is known. The calculation is carried out using an iterative procedure that also involves estimating the acoustic impedance of the microphone by means of any available method.

C. Directivity index

The random-incidence sensitivity of a microphone can be determined as the sum of the pressure sensitivity and the random-incidence correction. The random-incidence response is the average of the free-field response of the microphone to incoherent plane waves coming from all directions, and thus it has similar properties as the free-field correction. The random-incidence correction can also be determined from the directivity index and the free-field response. In general, the directivity factor, Q, at the frequency f is defined as

$$Q(f) = \frac{4\pi|H(f,\theta_0,\phi_0)|^2}{\int_0^{2\pi}\int_0^{\pi}|H(f,\theta,\phi)|^2 \sin\theta d\theta d\phi}, \tag{6}$$

where $H(f,\theta,\phi)$ is the frequency response at the angles θ and ϕ. The index 0 indicates the axial direction.[19] Assuming that the microphone is axi-symmetric and substituting the integral by a discrete series, Eq. (6) simplifies to

$$Q(f) = \frac{2|H(f,\theta_0)|^2}{\sum_{n=1}^{\pi/\Delta\theta}|H(f,\theta)|^2 \sin\theta_n \Delta\theta}. \tag{7}$$

The directivity index, D, is the directivity factor expressed in logarithmic fashion, i.e.,

$$D = 10 \log Q. \tag{8}$$

III. EXPERIMENTAL SETUP

The velocity of the membranes of various microphones has been measured using a laser vibrometer, Polytech PDV-100. The microphone was used as a transmitter, and its membrane was driven using a reciprocity apparatus, Brüel & Kjær (B&K) type 5998. The voltage on the terminals of the reference impedance on the transmitter unit, B&K type ZE0796, and the output of the vibrometer was measured using a B&K "PULSE" analyzer. Figure 2 shows a block diagram of the measurement setup, and Fig. 3 shows a picture of the vibrometer and the microphone mounted on the positioning rig.

The signal used for driving the microphone was pseudo-random noise with a bandwidth of 25.6 kHz and 6400 spectral lines. The laser vibrometer can measure up to 24 kHz. Although several types of microphones were measured, only results for 1-in. and 0.5-in. LS microphones (LS1 and LS2, respectively) are presented in what follows.

IV. BEM MODELING

An open source implementation of BEM has been used in the calculations. Details of the formulation can be found elsewhere,[1] but summarizing, such a formulation relates a possible incident wave p^I and the pressure p and velocity v at the point Q on the surface S of a body to the pressure at the external point P,

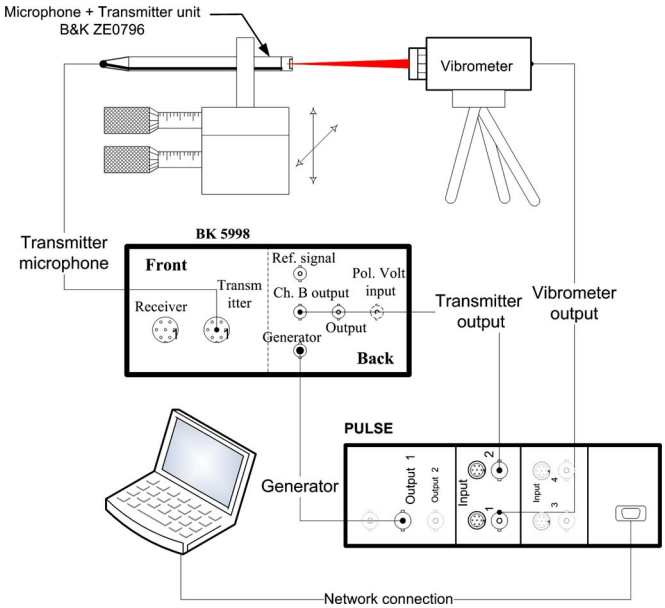


FIG. 2. (Color online) Block diagram of the measurement system.

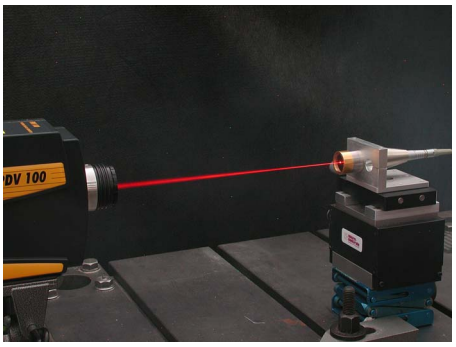


FIG. 3. (Color online) Measurement setup. The laser beam measured the velocity at a point on the membrane.

$$C(P)p(P) = \int_L \int_\theta \left(p(Q)\frac{\partial G(R)}{\partial n} + jkz_0 v(Q)G(R) \right) \times d\theta \rho(Q) dL(Q) + 4\pi p^I(P), \quad (9)$$

where $G(R)=e^{-jkR}/R$ is the free space Green's function, R is the distance between the points P and Q, z_0 is the characteristic impedance of air (ρc), and $C(P)$ is the solid angle seen from P. In the axi-symmetric case, the velocity and pressure are independent of the rotation angle, ϕ. However, it is possible to introduce non-axi-symmetric boundary conditions if the pressure and velocity are expanded in cosine series. In this case, only the axi-symmetrical expansion (zeroth term) has been used. The methodology proposed in this paper is to define $v(Q)$ in Eq. (9) as the measured velocity of the membrane instead of assuming an analytical form.

The geometry used in the BEM calculations is shown in Fig. 4. The ideal, semi-infinite rod was approximated by a cylindrical rod with a length of 60 cm and a hemispherical back-end. The termination of the rod inevitably gives rise to reflections, but because of the length of the rod the reflections can be expected that to have a negligible influence.

The BEM calculations were used for determining quantities such as the acoustic center, the free-field correction, and the directivity index. The acoustic center has an asymptotic behavior at low frequencies, and the directivity index and the free-field correction tend to zero at low frequencies. Therefore, there is no need for calculations, say, below 1 kHz for LS1 microphones. Moreover, the experimental estimates to be compared with the results from the hybrid method have a comparable lower frequency limit. On the other hand, the laser vibrometer can only measure up to 24 kHz, which sets the high frequency limit. Hence, the frequency range used in the calculations was from 1 to 24 kHz for LS1 microphones and from 2 to 24 kHz for LS2 microphones. The size of the smallest element in the axi-symmetric mesh was 2.5 and 1.5 mm for LS1 and LS2 microphones, respectively. Thus, there were at least six elements per wavelength at the highest frequency.

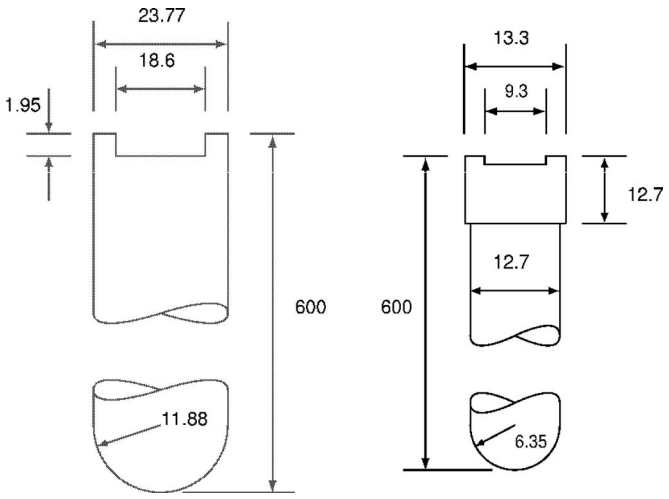


FIG. 4. Geometry of LS1 and LS2 microphones used in the simulations.

In order to avoid the non-uniqueness problem, a random CHIEF point was added in the interior of the geometry as described in Ref. 20, and the calculations were checked by determining the condition numbers of the BEM matrices and by repeating calculations with small frequency shifts.[21]

Depending on the quantity to be determined, the microphone acts as receiver or as a source. When the microphone acts as a source, the radiation problem is solved by assigning the measured velocity to the membrane of the microphone. In the scattering problem, the structural coupling between the membrane and the scattered sound field is solved using an iterative procedure.

V. RESULTS AND DISCUSSION

A. Movement of the membrane

Figures 5 and 6 show the measured velocity of the membrane of two different types of microphones at different frequencies. It can be seen in Fig. 5 that the shape of the movement of the membrane of an LS1 microphone is similar to a parabola at frequencies below 5 kHz. Above this frequency and around the resonance frequency, the shape deviates from the assumed parabola, and the deviations become more obvious the higher the frequency. From 14 kHz and upward, the center of the membrane flattens and no longer looks like a parabola nor like any other simple analytical shape. It is apparent that above 20 kHz, the center of the membrane does not move as much at a rim between the center and the fixed perimeter of the membrane. The velocity profiles are the result of the interaction between the membrane and the backplate of the microphone. The positions of the maxima coincide with the position of the holes and the recess on the backplate.

The movement of the membrane of an LS2 microphone shows a different behavior (see Fig. 6). It can be seen that the shape is more regular in the same frequency interval even around the resonance frequency (approximately 18 kHz). Only above the resonance frequency does the shape seem to flatten slightly. It can also be noticed that in both cases there is a phase delay at the center of the membrane with respect to the rim. This makes the movement of the membrane even more complex and difficult to emulate using simple analytical shapes. Thus, it seems to be difficult to make any *a priori* assumption of the movement of the membrane above the resonance frequency of the microphones.

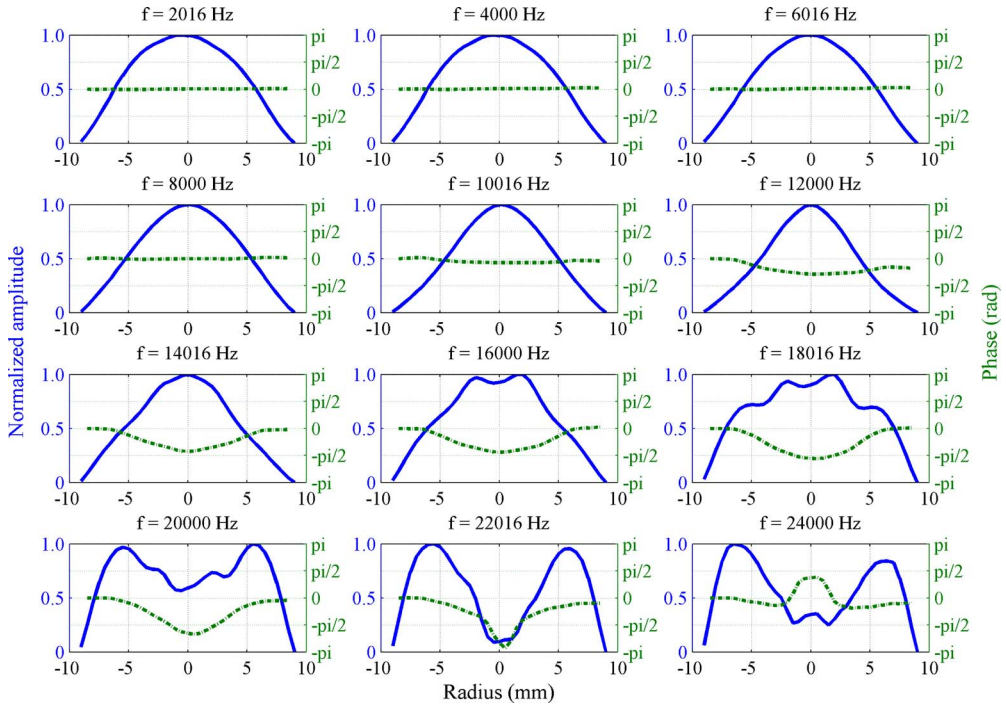


FIG. 5. (Color online) Velocity of the membrane of an LS1 microphone measured with a laser vibrometer at several frequencies. Solid line: normalized amplitude; dash-dotted line: phase.

B. Pressure sensitivity

There is no need for BEM calculations in determining the pressure sensitivity. Determining the ratio of the volume velocity to the current in Eq. (2) simply requires integrating the velocity over the membrane. Figures 7(a) and 7(b) show the normalized pressure sensitivity of an LS1 and an LS2 microphone compared with the experimental response obtained using the reciprocity technique.

At high frequencies, around and above the resonance, the difference between the responses determined by reciprocity calibration and by Eq. (2) is caused by the different load of the radiation impedance described by Eq. (2). In the vibrometer measurements, the radiation impedance is that of a microphone radiating to the open, whereas in the case of reciprocity calibration, the radiation impedance is the input impedance of a coupler terminated by a finite impedance at

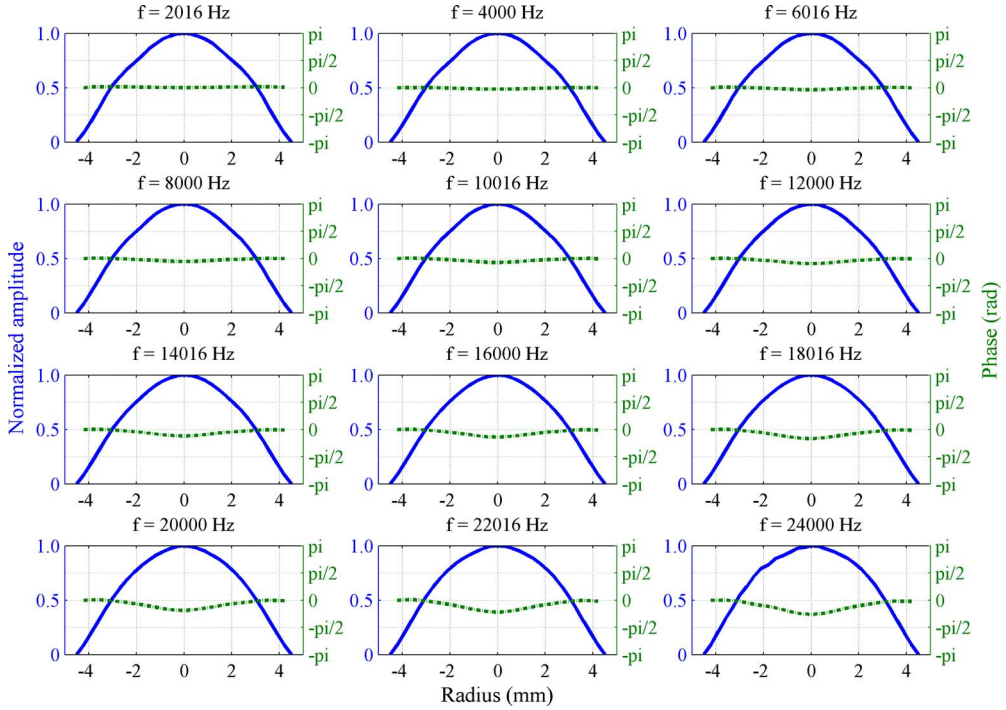


FIG. 6. (Color online) Velocity of the membrane of an LS2 microphone measured with a laser vibrometer at several frequencies. Solid line: normalized amplitude; dash-dotted line: phase.

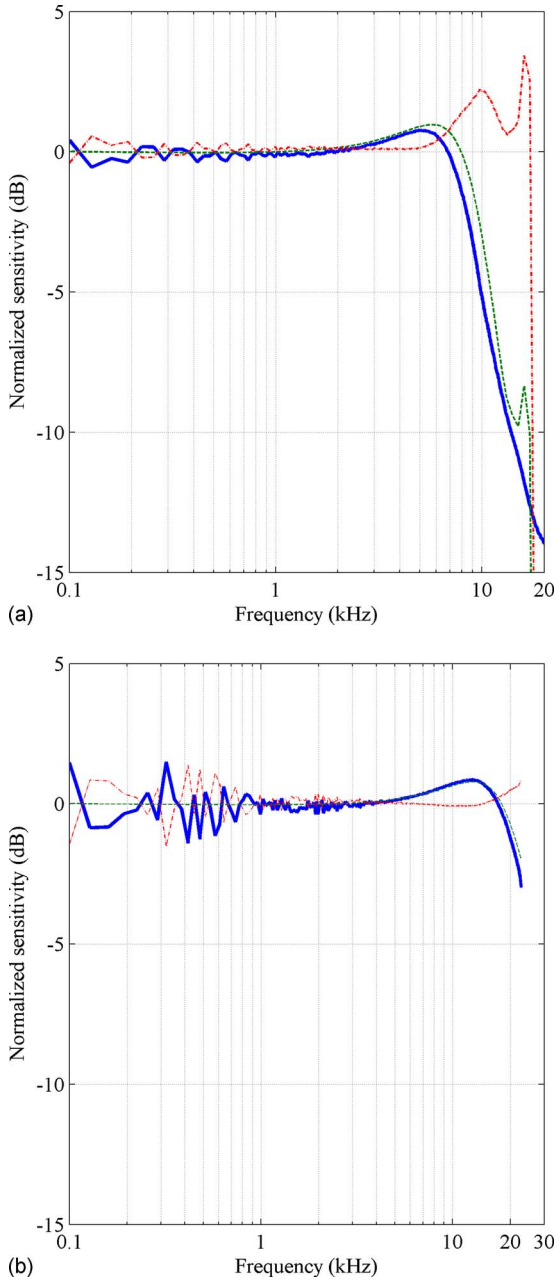


FIG. 7. (Color online) Normalized pressure response of (a) an LS1 microphone and (b) an LS2 microphone. Solid line: estimate from measurements of the velocity of the membrane; dashed line: estimate obtained from reciprocity calibration; dash-dotted line: difference between estimates.

the opposite end. The radiation impedance that loads the membrane in the vibrometer method is very similar to the radiation impedance that loads the membrane in electrostatic actuator measurements, and therefore a similar difference between reciprocity calibration and the hybrid method and between reciprocity calibration and actuator calibration can be expected. The difference seen in Figs. 7(a) and 7(b) agrees quite well with results observed in the literature for the difference between reciprocity and electrostatic actuator.[17] However, there is a small peak at about 16 kHz for the LS1 microphones that does not occur in the estimate from velocity measurements. This may be caused by the resonances of longitudinal modes in the couplers used in reciprocity measurements.

In the LS1, and in particular in the LS2 results, erratic sawtooth variations occur below 1 kHz. These variations may have been caused by insufficient averaging. Below 1 kHz, the microphones are very poor radiators of sound partly because the amplitude of the displacement of the membrane is very small, and the measurements may well have been affected by extraneous noise. It seems likely that an optimization of the measurement technique could improve the results.

C. Acoustic center

Figure 8 shows the acoustic center of an LS1 microphone determined from numerical BEM calculations using a parabolic function, a Bessel-like function, and the measured velocity distribution. The results are compared with data obtained experimentally from reciprocity measurements. It can be seen that the agreement between measured data and the BEM calculations using the measured velocity distribution is very good at any frequency above 2 kHz. The bump at 1.5 kHz in the estimate determined from measurements of the transfer function between two microphones is caused by the application of the time-selective procedure as discussed above. This suggests that the sound field calculated from the measured velocity distribution is more accurate than calculations based on any assumed distribution, in particular at high frequencies.

D. Free-field correction

Figure 9 shows the free-field correction of LS1 microphones determined with the hybrid method and from combined free-field and pressure reciprocity calibrations. The agreement between the reciprocity result and the estimate obtained with the hybrid method is not very good around and above the resonance frequency (about 8 kHz). The reason for this may be that in order to determine the correction an iterative BEM calculation procedure that requires knowledge of the acoustic impedance of the membrane of the micro-

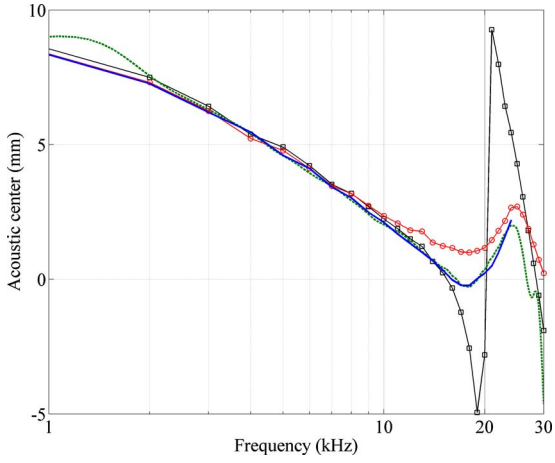


FIG. 8. (Color online) Acoustic center of LS1 microphones. Solid line: estimate from the hybrid method; dashed line: estimate determined from measurements of the transfer function between two microphones; line with square markers: numerical estimate assuming a Bessel-like movement; line with circular markers: numerical estimate obtained assuming a uniform velocity distribution.

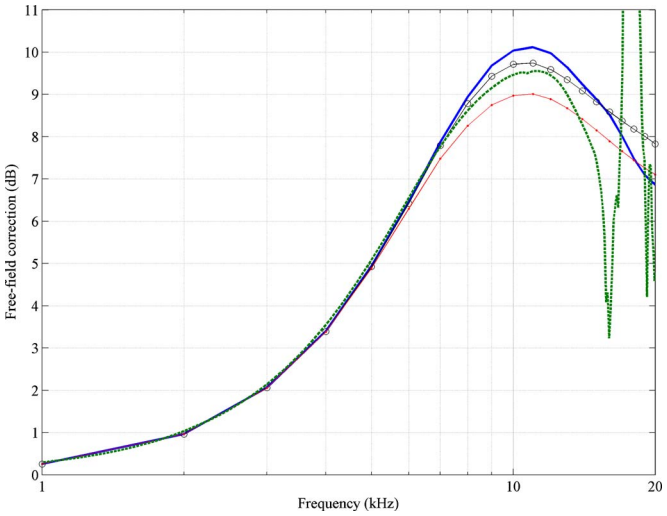


FIG. 9. (Color online) Free-field correction of LS1 microphones. Solid line: estimate from the hybrid method; dashed line: estimate determined from free-field and pressure reciprocity; line with circular markers: numerical estimate assuming parabolic movement; line with dot markers: numerical estimate assuming uniform movement.

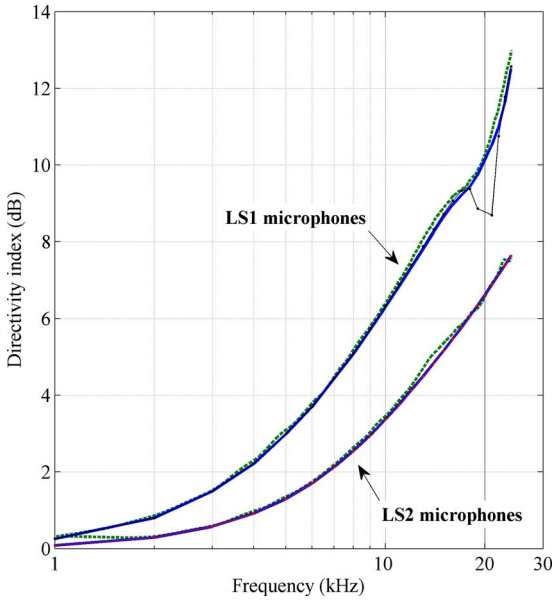


FIG. 10. (Color online) Directivity index of LS1 and LS2 microphones determined experimentally and with the hybrid method. The arrows point to the results for each type of microphone. Solid line: hybrid method; dashed line: estimate determined using the random-incidence technique; line with solid circular markers: numerical estimate assuming Bessel-like movement.

phone has been be used. In this case, a lumped-parameter approximation of the acoustic impedance was used.[1] This approximation is only accurate at frequencies below the resonance frequency, and therefore it is expected that the hybrid method should give less accurate results at frequencies near and above the resonance. Besides, as mentioned above, at frequencies above 15 kHz the pressure sensitivity determined from reciprocity, and thus the free-field correction, is no longer reliable due to longitudinal and radial resonances in the couplers.

E. Directivity index

Figure 10 shows the directivity index of LS1 and LS2 microphones. It is evident that the directivity index calculated from the measured velocity distribution follows the experimental estimate better than calculations based on an assumed distribution. This is particularly clear for the case of the LS1 microphone, in which the experimental index shows a change in slope at around 15 kHz. This behavior cannot be reproduced when using a Bessel movement in the simulations.

In the case of the LS2 microphones, the difference between the results of the experimental and the hybrid method coincide very well up to 24 kHz and with the numerical calculations using a Bessel-like movement. This is not unexpected because the resonance frequency of the microphone is about 22 kHz, and the actual movement of the membrane is "well-behaved" and still resembles a Bessel function; the same can be said of LS1 microphones at frequencies around and below the resonance frequency (about 8 kHz for LS1 microphones). However, nothing can be said for LS2 microphones at frequencies much higher than resonance because these frequencies are outside the measurement range of the laser vibrometer.

VI. CONCLUSIONS

The velocity distributions of the membranes of different types of microphones measured with a laser vibrometer have been found to be fairly complicated and demonstrate that no general assumption can be made for the behavior of all microphones. Preliminary results of the pressure sensitivity, the acoustic center, the free-field correction, and the directivity index obtained by a hybrid method that combines the measured velocity distributions with calculations with the BEM are, in general, in good agreement with experimental results obtained by traditional methods where the quantities are determined directly. It can therefore be concluded that the hybrid method can be used for validating new experimental setups. Furthermore, the hybrid method can be used in production environments to check the responses of a prototype microphone without the need of a complete calibration setup, and it is potentially useful for calibrating non-standard microphones. However, the hybrid method is not a substitute of an individual calibration of a particular transducer.

ACKNOWLEDGMENT

The investigation presented in this paper was made as a part of a project carried out with support from the Danish Agency for Technology and Production (FTP).

[1]P. M. Juhl, "A numerical investigation of standard condenser microphones," J. Sound Vib. **177**, 433–446 (1994).
[2]A. J. Zuckerwar, "Theoretical response of condenser microphones," J. Acoust. Soc. Am. **64**, 1278–1285 (1978).
[3]S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "The acoustic center of laboratory standard microphones," J. Acoust. Soc. Am. **120**, 2668–2675 (2006).
[4]S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "On experimental determination of the free-field correction of laboratory standard microphones at normal incidence," Metrologia **44**, 57–63 (2007).
[5]S. Barrera-Figueroa, K. Rasmussen, and F. Jacobsen, "On experimental

determination of the random-incidence response of microphones," J. Acoust. Soc. Am. **121**, 2628–2636 (2007).
[6]X. Bao and Y. Kagawa, "A simulation of condenser microphones in free field by boundary element approach," J. Sound Vib. **119**, 327–337 (1987).
[7]V. Cutanda-Henríquez, "Numerical transducer modeling," Ph.D. thesis, Ørsted-DTU, Technical University of Denmark (DTU), Denmark (2002).
[8]R. S. Grinnip III, "Advanced simulation of a condenser microphone capsule," J. Audio Eng. Soc. **54**, 157–166 (2006).
[9]G. Behler and M. Vorländer, "Reciprocal measurements on condenser microphones for quality control and absolute calibration," Acta Acust. **90**, 152–160 (2004).
[10]W. R. MacLean, "Absolute measurement of sound without a primary standard," J. Acoust. Soc. Am. **12**, 140–146 (1940).
[11]A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, New York, 1994), Chap. 4.10, pp. 199–203.
[12]"Measurement microphone—Part 3: Primary method for free-field calibration of laboratory standard microphones by the reciprocity technique," IEC International Standard 61094-3 (International Electrotechnical Commission, Geneve, Switzerland, 1995).
[13]W. Wathen-Dunn, "On the reciprocity free-field calibration of microphones," J. Acoust. Soc. Am. **21**, 542–546 (1949).
[14]I. Rudnick and M. N. Stein, "Reciprocity free-field calibration of microphones to 100 kc in air," J. Acoust. Soc. Am. **20**, 818–825 (1948).
[15]K. Rasmussen, "Acoustic centres of condenser microphones," Report No. 5, The Acoustics Laboratory, Technical University of Denmark, 1971.
[16]D. Rodrigues, J. N. Durocher, M. Bruneau, and A. M. Bruneau, "Free-field calibration of microphones: Theoretical and experimental determination of the acoustic center (A)," J. Acoust. Soc. Am. **123**, 3847 (2008).
[17]G. Rasmussen, "The free field and pressure calibration of condenser microphones using electrostatic actuator," Brüel & Kjær Technical Review **2**, 3–7 (1969).
[18]"Measurement microphones—Part 7: Values for the difference between free-field and pressure sensitivity levels of laboratory standard microphones," IEC Technical Specification 61094-7, 1st ed. (International Electrotechnical Commission, Geneve, Switzerland, 2006).
[19]L. L. Beranek, *Acoustical Measurements* (American Institute of Physics, Melville, NY, 1988).
[20]H. A. Schenck, "Improved integral formulation for acoustic radiation problems," J. Acoust. Soc. Am. **44**, 41–58 (1968).
[21]P. M. Juhl, F. Jacobsen, V. Cutanda, and S. Quiros-Alpera, "On the non-uniqueness problem in a 2-D half-space BEM formulation," in Proceedings of the Ninth International Congress on Sound and Vibration, Orlando, FL (2002).

Channel separation of crosstalk cancellation systems with mismatched and misaligned sound sources

Xiaojun Qiu[a)]
Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Jiangsu 210093, China

Bruno Masiero and Michael Vorländer
Institute of Technical Acoustics, RWTH Aachen University, D-52056 Aachen, Germany

(Received 7 October 2008; revised 25 July 2009; accepted 28 July 2009)

Loudspeakers in virtual sound imaging systems are usually modeled as omnidirectional monopole sources. These models are, however, only an approximation for the low frequency range. This paper presents an analytical model of crosstalk cancellation systems in a free field which takes into account the scattering and spatial characteristics of the sound sources. Based on the proposed model, the effects caused by the spatial characteristics of the sound source and its misalignments on the performance of the crosstalk cancellation system are studied numerically. It is found that although the factors such as the directivity of the sound sources and the distance between the sound sources and receiver affect the performance of the system to a certain extent, the channel separation of the crosstalk cancellation system, however, is most sensitive to the misalignment of the subtended angle of the sound sources. Therefore, if highly accurate binaural cues are required in practical applications, the type and characteristics of the playback sound sources, their locations, and orientations all should be considered carefully.
 [DOI: 10.1121/1.3206660]

PACS number(s): 43.38.Md, 43.38.Vk, 43.66.Pn [AJZ] Pages: 1796–1806

I. INTRODUCTION

During the past few decades, considerable attention has been paid to the area of sound field reproduction, raising many issues in acoustics and signal processing. Several approaches can be used to reproduce a required sound with loudspeakers, from which the most popular are the binaural techniques, ambisonics, and wave field synthesis.[1,2] If the sound is required to be reproduced at only two points in space (the ears of a listener), two loudspeakers can be used to generate a virtual sound image by pre-filtering the sound signal with a pair of causal linear filters.[2–6] If the sound arrival directions at the center point of a volume are of most interest, an ambisonics system with a small number of loudspeakers can be used, which is based on spherical harmonic decomposition of the acoustic field centered at the listener position.[7] For even larger listening area, the wave field synthesis concept based on the Kirchhoff–Helmholtz integral equation and the high order ambisonics technique can be used.[7–9] This paper concentrates on the particular binaural virtual sound imaging system that uses two loudspeakers together with the crosstalk cancellation (CTC) technique.

The head related transfer functions (HRTFs) describe the filtering of the sound from a specific location by the diffraction and reflection properties of the head, pinna, and torso of an individual before it reaches the individual's eardrums.[1–4] A synthesized binaural signal can be created by convolving a sound with the appropriate HRTFs for generating spatial audio effects by using headphones or loudspeakers. The difficulty with the loudspeaker binaural virtual sound imaging system is the crosstalk in the contralateral paths from the loudspeakers to the listener's ears, which destroys the three-dimensional (3D) cues of the binaural signal, and the CTC technique is the most commonly used way to solve the problem.

CTC systems (the systems that use CTC technique) have been studied for nearly 50 years, and an in-depth review of relevant literature can be found in recent book and papers.[1–4] It has been observed that a nearly perfect CTC (greater than 40 dB up to 5000 Hz) can be obtained for a matched CTC system where the playback HRTFs are the same as the setup HRTFs used to design the CTC filters. However, for a mismatched system, the channel separation drops significantly. It has been concluded by Akeroyd *et al.*[3] that although test trials with CTC systems are very impressive and provide compelling subjective test results on the angle perception, only a matched system can yield sufficiently accurate binaural cues required for spatial hearing experiments.

There are many factors that make the playback HRTFs mismatch the setup ones. For instance, the difference between HRTFs of different individuals whose pinnae, heads, and torsos differ in size and shape, the misalignment of the listener's head position and orientation, the misalignment of position and orientation of playback sound sources, the variation in the spatial and temporal characteristics of the playback sound sources, and the variation in the playback acoustics environment caused by wall reflections. The loss in performance caused by the difference between the setup HRTFs and the playback HRTFs of different individuals has been studied by Akeroyd *et al.*[3] and it turned out that the average channel separation can decrease by up to 20 dB in

[a)]Author to whom correspondence should be addressed. Electronic mail: xjqiu@nju.edu.cn

some cases. The study of the robustness to head misalignment by Takeushi *et al.*[5] shows that a system with loudspeakers that are positioned closer to each other is more robust to the misalignment of the listener's head. Rose *et al.*[6] further analyzed this fact in terms of the off-axis asymmetric listener locations and found that the width of tolerable lateral head translations for asymmetric locations is comparable to that for the symmetric case. A recent study carried out by Bai *et al.*[4] shows that, despite the fact that using a small source span angle of 10° results in a large relative sweet spot, using a larger source span of 60° or even 120° is more desirable in practical applications. This can produce a larger absolute sweet spot where the channel separation performance is guaranteed.

In addition to these robustness studies which focused mainly on the listener, there are also several studies on the influence of playback environments. Ward[10] developed a closed-form expression to predict the performance of a CTC system in a reverberant environment. Subjective tests carried out by Lentz show that adding reflecting walls to the listening environment decreases the performance of the CTC system significantly.[11] In some practical applications, such as spatial hearing experiments, where highly accurate binaural cues are required, it becomes evident that the scattering and reflections from the loudspeaker cabinets have to be taken into account.[3] However, little is known about effects of mismatched playback sound sources on the CTC system. In the design of the CTC filters, the setup HRTFs used are usually measured in prior with an ideal small sound source which radiates equally in all direction. But the sound sources used in playback are usually not the same ones, and their time and spatial characteristics might be different. These mismatches might deteriorate the performance of a CTC system.

In sometime applications, for example, in a cave automatic virtual environment (CAVE), the directivity of the sound sources needs to be considered. A CAVE is an immersive virtual reality environment where video projectors are directed to three, four, five, or six of the walls of a room-sized cube.[11] In a CAVE, a user can walk around to have a proper view of 3D objects from different directions by wearing special glasses, so a dynamic CTC system has to be used to provide a valid CTC filter set for each position the user might be to generate proper virtual sound images. The term "dynamic" here means updating the CTC filter dynamically (adaptively) with the listener position change detected by a head tracker. Under this situation, the relative angle between the loudspeakers' orientation and the listener is not fixed and it might not be easy to align the orientation of the playback sound sources with the moving listener adequately in practice, so the directivity of the sound sources might need to be taken into account to achieve the required performance.[11]

All these facts motivate the current research to investigate the significance of the effects caused by the spatial characteristics and misalignment of playback sound sources. Loudspeakers used in CTC systems are usually considered as monopoles in existing models, but this model is only an approximation in the low frequency range. This paper develops first of all an analytical model of CTC systems in a free field, which takes into account the scattering and spatial characteristics of the sound sources. Then, based on the proposed model, the effects of the spatial characteristics and misalignments of sound sources on the performance of CTC systems are studied.

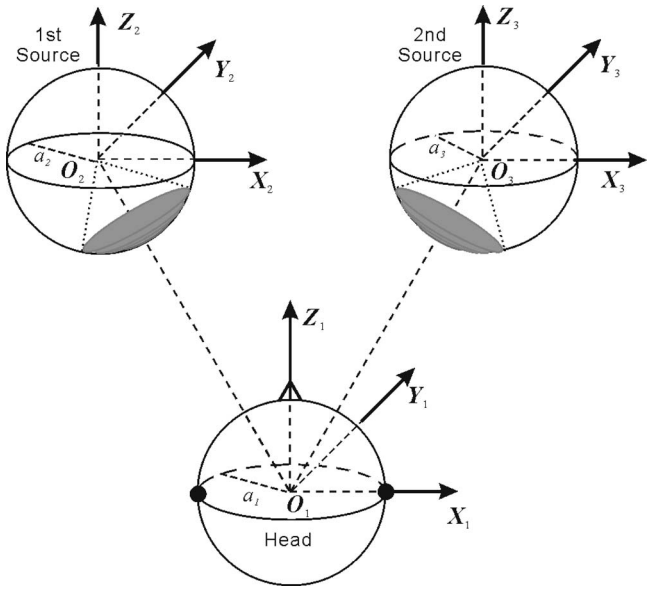


FIG. 1. The free field geometrical arrangements of the CTC systems with two vibrating cap sound sources on two rigid spheres and two receiving points on a third rigid sphere.

II. ANALYTICAL MODEL

A CTC system is modeled with three rigid spheres with a radius of a_i $(i=1,2,3)$, as shown in Fig. 1. Each sphere has its own spherical coordinate system referred to as O_i $(i=1,2,3)$. The coordinate systems O_2 and O_3 are obtained by a translational movement of the original spherical center from $\mathbf{r}_{o1}=(0,0,0)$ to $\mathbf{r}_{o2}=(d_{o2},\theta_{o2},\phi_{o2})$ or $\mathbf{r}_{o3}=(d_{o3},\theta_{o3},\phi_{o3})$ in the coordinate system O_1. The same receiving point is called $\mathbf{r}_{1r}=(r_{1r},\theta_{1r},\phi_{1r})$ in the coordinate system O_1 or $\mathbf{r}_{2r}=(r_{2r},\theta_{2r},\phi_{2r})$ in the coordinate system O_2 or $\mathbf{r}_{3r}=(r_{3r},\theta_{3r},\phi_{3r})$ in the coordinate system O_3.

Sphere 1 is used to model the head of a listener and Spheres 2 and 3 are used to approximate the loudspeaker cabinets. The model for the sound source used here consists of a rigid sphere with a radially vibrating cap (gray part shown in Fig. 1) defined by a certain solid angle. It is based on the assumption that the diffraction effects caused by a sphere and a cube with similar dimensions are similar if the dimensions of the diffracting bodies are considerably smaller than the acoustic wavelength.[12] Despite the fact that the error caused by this idealization increases in higher frequency, it can be used to show the general effects of the loudspeaker's directional proprieties and diffractions on the performance of the CTC systems and provides some insights into the basic mechanisms.

The simplest Z axis symmetrical case is considered first to calculate the spherical harmonic coefficients of a cap source in its own coordinate system. The velocity distribution of a cap source located at the North Pole on the surface of such a sphere is assumed as

$$U_o(\theta,\phi)=\begin{cases}u_0, & 0\le\theta\le\theta_s, 0\le\phi<2\pi\\ 0, & \theta_s<\theta\le\pi, 0\le\phi<2\pi,\end{cases} \tag{1}$$

where θ_s is half of the cap source span angle, and the time-dependent factor $e^{j\omega t}$ is suppressed throughout the analysis. The velocity distribution of this radially vibrating polar cap can be expanded in terms of spherical harmonics as[13]

$$U_o(\theta,\phi)=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}U_{lm}^{o}Y_{lm}(\theta,\phi), \tag{2}$$

where the spherical harmonics functions are defined as

$$Y_{lm}(\theta,\phi)=\sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\,\theta)e^{jm\phi}, \tag{3}$$

$P_l^m(\cos\,\theta)$ is the *associated Legendre function* of degree l and order m evaluated at $\cos\,\theta$, and the spherical harmonic coefficients are given by

$$U_{lm}^{o}=\begin{cases}\dfrac{u_0}{2}\sqrt{\dfrac{4\pi}{2l+1}}[P_{l-1}^0(\cos\,\theta_s)-P_{l+1}^0(\cos\,\theta_s)], & m=0\\ 0, & m\neq 0.\end{cases} \tag{4}$$

For $l=0$, $P_{-1}^0(x)=1$.

After rotating the center of the vibrating polar cap to a new position with an elevation angle of θ_D and an azimuth angle of ϕ_D, the spherical harmonic coefficients of the obtained source velocity distribution on the sphere surface in the original coordinate can be written as[14–17]

$$U_{lm}^{D}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_D,\phi_D), \tag{5}$$

where superscript "*" denotes the complex conjugate.

The center of the first source on the surface of Sphere 2 is assumed to be (θ_L,ϕ_L) in its own coordinate system O_2 and that of the second source on the surface of Sphere 3 be (θ_R,ϕ_R) in its coordinate system O_3; then the spherical harmonic coefficients of the velocity distribution of each source in its own coordinate systems are

$$U_{lm}^{L}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_L,\phi_L), \tag{6}$$

$$U_{lm}^{R}=\sqrt{\frac{4\pi}{2l+1}}U_{l0}^{o}Y_{lm}^{*}(\theta_R,\phi_R). \tag{7}$$

The total sound field produced by the sources and the three spheres consists of three parts: the radiated and scattered sound fields from source 1 (Sphere 2) and source 2 (Sphere 3), and the scattered field from the listener's head (Sphere 1). It can be expressed in each coordinate system as

$$\begin{aligned}p_t(\mathbf{r}_{1r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}),\\ p_t(\mathbf{r}_{2r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}),\\ p_t(\mathbf{r}_{3r})&=p_{s_1}(\mathbf{r}_{1r})+p_{s_2}(\mathbf{r}_{2r})+p_{s_3}(\mathbf{r}_{3r}).\end{aligned} \tag{8}$$

The sound field (radiated and/or scattered) from each sphere can be conveniently expressed in its own coordinate system as

$$\begin{aligned}p_{s1}(\mathbf{r}_{1r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}C_{lm}h_l(kr_{1r})Y_{lm}(\theta_{1r},\phi_{1r}),\\ p_{s2}(\mathbf{r}_{2r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}D_{lm}h_l(kr_{2r})Y_{lm}(\theta_{2r},\phi_{2r}),\\ p_{s3}(\mathbf{r}_{3r})&=\sum_{l=0}^{\infty}\sum_{m=-l}^{l}E_{lm}h_l(kr_{3r})Y_{lm}(\theta_{3r},\phi_{3r}),\end{aligned} \tag{9}$$

where C_{lm}, D_m, and E_m are unknown spherical harmonic coefficients to be determined by applying the boundary conditions on the surfaces of the spheres. $k=\omega/c_0$ is the wave number and c_0 is the sound speed. $h_l(x)=j_l(x)-jn_l(x)$ is the *spherical Hankel function* of order l, $j_l(x)$ is the *spherical Bessel function* of order l, and $n_l(x)$ is the *spherical Neumann function* of order l.

The difficulty in solving the above equations is that there are functions and variables in three different coordinate systems. This can be simplified by using the translational addition theorem,[18–22] which expresses the sound pressure of the same point in space, originally represented on the coordinate system O_j in terms of the coordinate system O_i by

$$\begin{aligned}&h_l(kr_{jr})Y_{lm}(\theta_{jr},\phi_{jr})\\ &\quad=\sum_{p=0}^{\infty}\sum_{q=-p}^{p}Q_{pq}^{lm}(\mathbf{r}_{oij})j_p(kr_{ir})Y_{pq}(\theta_{ir},\phi_{ir}),\\ &\quad i,j=1,2,3,\end{aligned} \tag{10}$$

where $\mathbf{r}_{oij}=\mathbf{r}_{oj}-\mathbf{r}_{oi}=(r_{oij},\theta_{oij},\phi_{oij})$.

$$\begin{aligned}Q_{pq}^{lm}(\mathbf{r}_{oij})=&\sum_{n=|l-p|:2}^{l+p}4\pi(-j)^{p+n-l}h_n(kr_{oij})\\ &\times Y_{n,m-q}(\theta_{oij},\phi_{oij})g(m,l,-q,p,n),\end{aligned} \tag{11}$$

with $n=l+p,l+p-2,\ldots,|l-p|$, and

$$\begin{aligned}&g(m,l,-q,p,n)\\ &\quad=(-1)^m\sqrt{\frac{(2l+1)(2p+1)(2n+1)}{4\pi}}\\ &\qquad\times\begin{pmatrix}l & p & n\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}l & p & n\\ m & -q & -m+q\end{pmatrix}\end{aligned} \tag{12}$$

is related to the Gaunt coefficients, which are expressed as products of the Wigner $3j$ symbols, as shown in Eq. (12).[18,19,21,22]

Substituting Eqs. (9) and (10) into Eq. (8) and using the boundary condition $j\rho_0\omega v_n(\mathbf{r})=-\partial p_t(\mathbf{r})/\partial r$ at the surface of each rigid sphere in its own coordinate system (where ρ_0 is the density of the medium) yield

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}h_l'(ka_1)Y_{lm}(\theta_{1r},\phi_{1r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o12})j_p'(ka_1)Y_{pq}(\theta_{1r},\phi_{1r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o13})j_p'(ka_1)Y_{pq}(\theta_{1r},\phi_{1r}) = 0,$$

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o21})j_p'(ka_2)Y_{pq}(\theta_{2r},\phi_{2r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}h_l'(ka_2)Y_{lm}(\theta_{2r},\phi_{2r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o23})j_p'(ka_2)Y_{pq}(\theta_{2r},\phi_{2r}) = -j\rho_o c_o\sum_{l=0}^{\infty}\sum_{m=-l}^{l} U_{lm}^{L}Y_{lm}(\theta_{2r},\phi_{2r}),$$

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} C_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o31})j_p'(ka_3)Y_{pq}(\theta_{3r},\phi_{3r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} D_{lm}\sum_{p=0}^{\infty}\sum_{q=-p}^{p} Q_{pq}^{lm}(\mathbf{r}_{o32})j_p'(ka_3)Y_{pq}(\theta_{3r},\phi_{3r}) + \sum_{l=0}^{\infty}\sum_{m=-l}^{l} E_{lm}h_l'(ka_3)Y_{lm}(\theta_{3r},\phi_{3r}) = -j\rho_o c_o\sum_{l=0}^{\infty}\sum_{m=-l}^{l} U_{lm}^{R}Y_{lm}(\theta_{3r},\phi_{3r}). \quad (13)$$

The prime denotes the differentiation with respect to the argument of the function. Truncating the number of summations to L (depending on the calculation precision required) and equating the coefficients of $Y_{lm}(\theta_{1r},\phi_{1r})$, $Y_{lm}(\theta_{2r},\phi_{2r})$, and $Y_{lm}(\theta_{3r},\phi_{3r})$ in Eq. (13), the following coupled linear complex equations can be obtained:

$$C_{lm}h_l'(ka_1) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq}Q_{lm}^{pq}(\mathbf{r}_{o12})j_l'(ka_1) + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq}Q_{lm}^{pq}(\mathbf{r}_{o13})j_l'(ka_1) = 0,$$

$$\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq}Q_{lm}^{pq}(\mathbf{r}_{o21})j_l'(ka_2) + D_{lm}h_l'(ka_2) + \sum_{p=0}^{L}\sum_{q=-p}^{p} E_{pq}Q_{lm}^{pq}(\mathbf{r}_{o23})j_l'(ka_2) = -j\rho_o c_o U_{lm}^{L},$$

$$\sum_{p=0}^{L}\sum_{q=-p}^{p} C_{pq}Q_{lm}^{pq}(\mathbf{r}_{o31})j_l'(ka_3) + \sum_{p=0}^{L}\sum_{q=-p}^{p} D_{pq}Q_{lm}^{pq}(\mathbf{r}_{o32})j_l'(ka_3) + E_{lm}h_l'(ka_3) = -j\rho_o c_o U_{lm}^{R}. \quad (14)$$

For all l and $m(l=0,1,2,\ldots,L;m=-L,-L+1,\ldots,0,1,2,\ldots,L)$, the equations above can be expressed in a more compact form by using matrix notations as

$$\begin{bmatrix} \mathbf{S}_1 & \mathbf{Q}_{12} & \mathbf{Q}_{13} \\ \mathbf{Q}_{21} & \mathbf{S}_2 & \mathbf{Q}_{23} \\ \mathbf{Q}_{31} & \mathbf{Q}_{32} & \mathbf{S}_3 \end{bmatrix}\begin{bmatrix} \mathbf{C} \\ \mathbf{D} \\ \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix}, \quad (15)$$

where the lmth element of the $(L+1)^2$ row vectors $\mathbf{A_1}$, $\mathbf{A_2}$, and $\mathbf{A_3}$ are 0, $-j\rho_o c_o U_{lm}^{L}$, and $-j\rho_o c_o U_{lm}^{R}$, respectively. The three groups of $(L+1)^2$ row unknown vectors are

$$\mathbf{C} = [C_{0,0} \quad C_{1,-1} \quad C_{1,0} \quad C_{1,1} \quad \cdots \quad C_{l,m} \quad \cdots \quad C_{L,L}]^T,$$

$$\mathbf{D} = [D_{0,0} \quad D_{1,-1} \quad D_{1,0} \quad D_{1,1} \quad \cdots \quad D_{l,m} \quad \cdots \quad D_{L,L}]^T,$$

$$\mathbf{E} = [E_{0,0} \quad E_{1,-1} \quad E_{1,0} \quad E_{1,1} \quad \cdots \quad E_{l,m} \quad \cdots \quad E_{L,L}]^T, \quad (16)$$

and the three groups of $(L+1)^2\times(L+1)^2$ diagonal complex matrices are

$$\mathbf{S}_1 = \mathrm{diag}\{s_{0,0}^{(1)} \quad s_{1,-1}^{(1)} \quad s_{1,0}^{(1)} \quad s_{1,1}^{(1)} \quad \cdots \quad s_{l,m}^{(1)} \quad \cdots \quad s_{L,L}^{(1)}\},$$

$$s_{l,m}^{(1)} = h_l'(ka_1), \quad \forall\, m,$$

$$\mathbf{S}_2 = \mathrm{diag}\{s_{0,0}^{(2)} \quad s_{1,-1}^{(2)} \quad s_{1,0}^{(2)} \quad s_{1,1}^{(2)} \quad \cdots \quad s_{l,m}^{(2)} \quad \cdots \quad s_{L,L}^{(2)}\},$$

$$s_{l,m}^{(2)} = h_l'(ka_2), \quad \forall\, m,$$

$$\mathbf{S}_3 = \mathrm{diag}\{s_{0,0}^{(3)} \quad s_{1,-1}^{(3)} \quad s_{1,0}^{(3)} \quad \cdots \quad s_{1,1}^{(3)} \quad \cdots \quad s_{L,L}^{(3)}\},$$

$$s_{l,m}^{(3)} = h_l'(ka_3), \quad \forall\, m \quad (17)$$

The element of $(L+1)^2\times(L+1)^2$ complex matrix $\mathbf{Q}_{ij}$ is

$$\xi_{lm,pq}^{ij} = Q_{lm}^{pq}(\mathbf{r}_{oij})j_l'(ka_i), \quad (18)$$

where the row index is lm and the column index is pq, in the orders of

$$lm = \{(0,0)(1,-1)(1,0)(1,1)\cdots(l,m)\cdots(L,L)\},$$

$$pq = \{(0,0)(1,-1)(1,0)(1,1)\cdots(l,m)\cdots(L,L)\}. \quad (19)$$

Having obtained the $3(L+1)^2$ unknown spherical harmonic coefficients C_{lm}, D_{lm}, and E_{lm}, the total sound field can be calculated by using any equation in Eq. (8). For example,

$$p_t(\mathbf{r}_{1r}) = \sum_{l=0}^{L}\sum_{m=-l}^{l} C_{lm} h_l(kr_{1r}) Y_{lm}(\theta_{1r}, \phi_{1r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} D_{lm} h_l(kr_{2r}) Y_{lm}(\theta_{2r}, \phi_{2r}) + \sum_{l=0}^{L}\sum_{m=-l}^{l} E_{lm} h_l(kr_{3r}) Y_{lm}(\theta_{3r}, \phi_{3r}). \quad (20)$$

It is often convenient to designate the complex acoustic pressure p as the output and the complex source volume acceleration $j\omega\rho_0 q/4\pi$ as the input.[2] For the radially vibrating cap source, $q=2\pi a_i^2(1-\cos\theta_s)u_0$. The relation between the source input signal and the receiving output signal or the frequency response from one sound source to the receiving point can be obtained the same way as that for Eq. (20). Nevertheless, the given source spherical harmonic coefficients used now are the followings instead of Eqs. (6) and (7),

$$U_{lm}^{L} = \sqrt{\frac{4\pi}{2l+1}} \frac{U_{l0}^{o} Y_{lm}^{*}(\theta_L, \phi_L)}{j\omega\rho_0 a_2^2(1-\cos\theta_s)u_0/2}. \quad (21)$$

$$U_{lm}^{R} = \sqrt{\frac{4\pi}{2l+1}} \frac{U_{l0}^{o} Y_{lm}^{*}(\theta_R, \phi_R)}{j\omega\rho_0 a_3^2(1-\cos\theta_s)u_0/2}. \quad (22)$$

Obviously, this model can be used to model similar systems where the radii of the spheres are different or when the vibrating cap sound sources are located at different positions on the sphere with different apertures. The model can also be extended to more sophisticated settings with multiple listeners and loudspeakers.[23]

III. NUMERICAL RESULTS AND DISCUSSION

A. Sound source characteristics

The sound sources used in the simulations include a single point source, a baffled piston, a single radially vibrating cap on a rigid sphere, and a radially vibrating cap on a rigid sphere with another rigid sphere aside. The single receiving point is on the central axis line of the sound source with a distance of 1.0 m from the source acoustic center. If the source is a vibrating cap on the surface of a sphere, the distance is that between the center of the cap on the surface of the sphere (not the center of the sphere) and the receiving point. The radius of the baffled piston is 0.1 m and the radius of source spheres is $a_1=a_2=0.1$ m. Half of the cap source span angle, θ_s, is 30°, which is about the size of a 4 in. loudspeaker. For settings with two source cases where another sphere is introduced, the two sources subtend angles of 60° and 20° at the receiving position, and the distance between the centers of the two spheres are 1.1 and 0.382 m, respectively. Even though the "stereo dipole" configurations[2] require the two sources to be closely spaced with only a 10° subtended angle, the size of the source sphere in this paper is too large for a 10° subtend angle, so that a subtended angle of 20° is used in the simulations. The locations of the vibrating cap center on the surface of the left sphere (Sphere 2) are ($\theta_L=150°$, $\phi_L=0°$) and ($\theta_L=170°$, $\phi_L=0°$) in coordinate system O_2, respectively, for the 20° and 60° subtended angle configurations, and the locations of the vibrating cap on the surface of right sphere (Sphere 3) are ($\theta_R=150°$, $\phi_R=180°$) and ($\theta_R=170°$, $\phi_R=180°$) in coordinate system O_3.

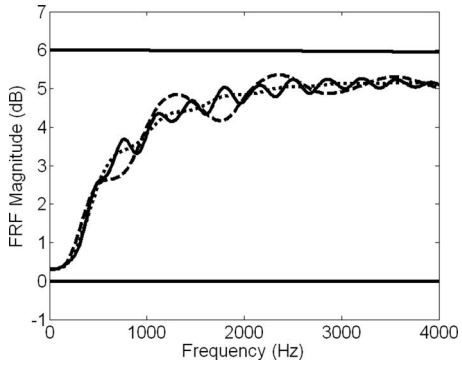


FIG. 2. Magnitude of the calculated FRFs for various sound source configurations. The lower solid line with 0 dB magnitude is that of the point source, the upper solid line with a magnitude of 6 dB is that of a baffled piston, the dotted curve is that of a single cap source on a rigid sphere, and the middle solid and dashed curves are for the two sphere sound sources with 60° and 20° subtended angles, respectively.

In a series of preliminary calculations, it was detected that using a spherical harmonics expansion degree of $L=10$ provides sufficient iteration precision, with the errors smaller than 1% for the current setup in the frequency range up to 4 kHz. Higher expansion degrees should be used for larger ratio of the size of the sphere to the wavelength. As this acoustical model can only approximate the characteristics of a loudspeaker up to a certain frequency range with the wavelength larger than the source size, no attempt has been made to obtain results at higher frequencies. Nevertheless, the range up to 4 kHz covers the basic and most important frequency range in fields such as communication and sound reproduction.

The frequency response functions (FRFs) are calculated for all above mentioned configurations with a source input of the same volume acceleration at 257 evenly distributed points between 0 and 4096 Hz, and the results are presented in Fig. 2. The lower solid line with 0 dB magnitude is the FRF of the point source, and the upper solid line with a magnitude of 6 dB is that of the baffled piston. The dotted curve is the FRF of a single cap source on a rigid sphere, and finally the middle solid and dashed curves are that of the two sphere cases. The second source on the right sphere is not active and the sphere is assumed to be rigid for the configurations with two sound sources, so that the frequency response is just that of one source. As expected, the sources of a vibrating cap on a rigid sphere do not have a flat frequency response (in comparison to an ideal point source). The vibrating cap on a rigid sphere radiates like a point source in the low frequency range. In the high frequency range, it almost behaves like a baffled piston and the pressure it radiates is nearly two times higher, which is caused by the scattering of the rigid sphere where the cap sound source is located.

From the FRF curves of the configurations with two rigid sphere sound sources, it can be seen that the scattering from a neighboring rigid sphere adds extra fluctuations on the original FRF curves. The magnitude and frequency of the

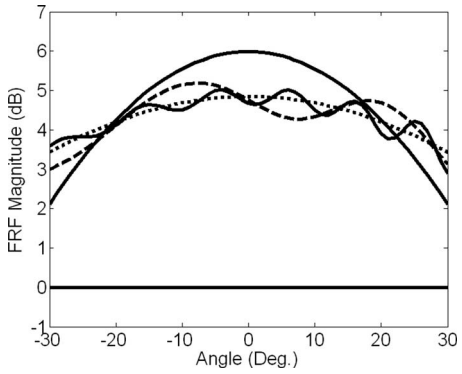


FIG. 3. Magnitude of the calculated FRFs at 2000 Hz when the receiving point swings from −30° (left) to 30° (right) for the configurations of Fig. 2. The distance between the source and the receiving point remains at 1.0 m. The lower solid line is that for the point source, the upper solid curve is that for the baffled piston, the dotted curve is that for the single cap source on a rigid sphere, and the middle solid and dashed curves are that for the two sphere sound sources with 60° and 20° subtended angles, respectively.

fluctuations depend on the distance between the two spheres, and the fluctuation magnitude becomes larger for nearer distance. The magnitude variation, for example, caused by the neighboring rigid sphere scattering can be as large as about 1 dB around 2000 Hz. It also can be observed that the magnitude of the fluctuation decreases at higher frequencies. This is because the cap source behaves more directional at higher frequencies, so less sound is radiated to and scattered from the neighboring sphere.

Figure 3 illustrates the directional behavior of the simulated systems, which depicts the calculated transfer functions at a specific frequency of 2000 Hz. At this frequency, the receiving point swings from −30° (left) to 30° (right) (with the distance between the source and the receiving point remaining at 1.0 m). 0° is the angle for the central axis line. The lower solid line of 0 dB is for the monopole source. The upper solid curve is that of the baffled position, the dotted curve is for one rigid sphere sound source, and the middle solid and dashed curves are for the two rigid sphere sound sources that subtend angles of 60° and 20°, respectively. As expected, the vibrating cap on a rigid sphere presents a directional behavior and the magnitude of the FRF at 30° can be nearly 2 dB lower than that at 0°. The scattering from the neighboring rigid sphere makes the sound pressure spatial distribution more complicated. This might result in a large difference in the excess sound pressure level between two ears of a CTC system.

B. CTC systems with non-omnidirectional sound sources

Assume that $\mathbf{C}$ is the 2×2 plant transfer function matrix in a CTC system, $\mathbf{V}=[v_1,v_2]^T$ is the source input signal vector, $\mathbf{H}$ is the 2×2 CTC matrix being designed, and $\mathbf{U}=[u_1,u_2]^T$ is the recorded/synthesized binaural signal vector. The system output $\mathbf{W}=[w_1,w_2]^T$ can be expressed as

$$\mathbf{W}=\mathbf{CV}=\mathbf{CHU}. \tag{23}$$

In order to ensure the reproduction of the binaural signals at the receiving points, $\mathbf{W}=\mathbf{U}e^{-j\omega\Delta}$, where Δ is a delay to ensure causality, it is necessary that $\mathbf{CH}=\mathbf{I}e^{-j\omega\Delta}$; thus $\mathbf{H}=\mathbf{C}^{-1}e^{-j\omega\Delta}$. However, the matrix could be ill-conditioned in some situations, so the following approximate solution is often used:[2–5]

$$\mathbf{H}=[\hat{\mathbf{C}}^{\mathrm{H}}\hat{\mathbf{C}}+\beta\mathbf{I}]^{-1}\hat{\mathbf{C}}^{\mathrm{H}}e^{-j\omega\Delta}, \tag{24}$$

where $\hat{\mathbf{C}}$ is the estimation of $\mathbf{C}$, and for the matched case, $\hat{\mathbf{C}}=\mathbf{C}$. In this paper, $\hat{\mathbf{C}}$ is also referred to as the setup plant transfer functions or setup HRTFs, while $\mathbf{C}$ is referred to as the playback plant transfer functions or HRTFs. β is a regularization parameter which constraints the energy of the source input signals and prevents the singularity of $\hat{\mathbf{C}}^{\mathrm{H}}\hat{\mathbf{C}}$ from saturating the filter gains. The performance of a CTC system is described by the performance matrix below:

$$\mathbf{P}=\mathbf{CH}, \tag{25}$$

where $\mathbf{P}$ should be a unit matrix under ideal CTC and the elements of the matrix should be $P_{11}=P_{22}=1$ and $P_{12}=P_{21}=0$. Channel separation is used here to illustrate the performance variation with frequency, which is defined as the ratio of the ipsilateral ear response to the contralateral ear response from the left (or right) recorded/synthesized binaural signal source:[4,5,24]

$$J_L(\omega)=20\log_{10}(|P_{11}(\omega)/P_{21}(\omega)|),$$

$$J_R(\omega)=20\log_{10}(|P_{22}(\omega)/P_{12}(\omega)|). \tag{26}$$

For the left-right symmetrical configurations, the right to left channel separation J_R (for the right recorded/synthesized signal) and the left to right channel separation J_L (for the left recorded/synthesized signal) are equal. The channel separations are thus interchangeably and the term "channel separation" is used to refer to both.

The commonly used models for studies dealing with CTC systems are the free field model and the spherical head model, which both assume that the playback sound sources are ideal point sources. The model shown in Fig. 1 provides the basis for this paper and is used to study non-ideal sound sources and their interaction. Here, the distance between the source center and the center of the listener head is 1.0 m, and the two sound sources subtend an angle of 60° at the listener position, the radius of source spheres is $a_1=a_2=0.1$ m, and the radius of receiver sphere (head) is $a_3=0.09$ m.

To emphasize the effects of the sound source directivity on the CTC performance, the spherical head model with two ideal point sources is used first of all by assigning different gains at different directions on its playback plant transfer functions. More complicated situations involving orientation changes of the playback sound sources are studied in Secs. III C and III D. Figure 4 shows the simulation results of the corresponding CTC systems. The dotted line corresponds to the ideal omnidirectional playback sound sources and the solid lines correspond to the playback sound sources without omnidirectional directivities. Note that a channel separation greater than 50 dB is clipped at 50 dB throughout the paper.

In the above simulations, the setup plant transfer functions obtained from the spherical head model with two ideal point sources are used to design the CTC filters. While cre-

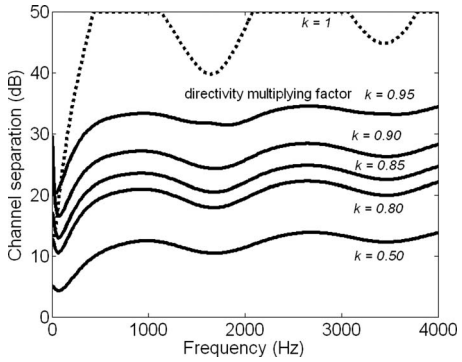


FIG. 4. Channel separations of the CTC system using playback plant transfer functions with different directivities. The setup plant transfer functions are obtained with the point sources of the spherical head model. The playback sound sources are the same as the setup ones (dotted line) or the ones that radiate less in the crosstalk path direction with the transfer function magnitude k (k=0.95, 0.90, 0.85, 0.80, and 0.50) times of that of the original (solid curves, the curve with the highest channel separation of above 30 dB is that for 0.95, then 0.90, 0.85, 0.80, and the lowest curve is that for 0.50).

ating the filters, the regularization parameter β is adjusted to make the channel separation just greater than 40 dB above 500 Hz under the matched situation throughout the paper. In this simulation, β is 0.012. The directivities of the playback plant transfer function are simulated by directly multiplying small values of 0.95, 0.90, 0.85 0.80, and 0.5 on the crosstalk path transfer functions from the same spherical head model with two ideal point sources, which corresponds to differences between the signals at two ears of about 0.4, 0.9, 1.4, 1.9, and 6 dB of excess sound pressure level. This simple approximation might be too rough for modeling the frequency-dependent directivities of real loudspeakers, yet it shows clearly that a small difference in the magnitude between two responses at two ears reduce the channel separation significantly. For example, it can be seen from the figure that if the excess sound pressure level difference between two ears caused by the source directivity is about 2 dB, and the channel separation drops significantly from above 40 dB to less than 20 dB.

C. Dynamic CTC systems with non-ideal sound sources

Figure 5 shows an example of a dynamic CTC system where a listener moves to a location which is not at the center of a symmetrical loudspeaker setup.[11] This situation

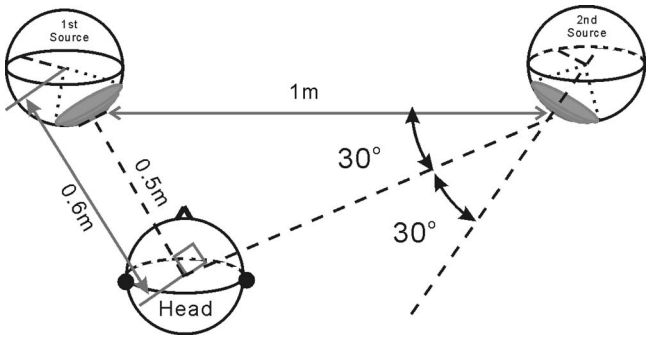


FIG. 5. A dynamic CTC system with the listener being located at an asymmetrical position.

can often be met in a CAVE-like environment where the user can walk around freely inside. At the stage of designing the binaural virtual sound imaging system for a CAVE, the plant transfer functions (or HRTFs) from a database are used as the setup ones to design a set of CTC filter for each potential listener position and direction in the CAVE. It should be emphasized that these plant transfer functions (or HRTFs) in the database were usually measured in prior with an ideal omnidirectional sound source for sound from different distances and directions. However, in practical applications in a CAVE, the loudspeaker is rarely ideal and usually radiates differently in different directions, and it might also not be realistic sometimes for the loudspeakers to be rotated in realtime according to the listener's position.

In the example shown in Fig. 5, the setup plant transfer functions are obtained with ideal point sources, the playback sound sources are the vibrating cap sources, their orientations were fixed in prior, and the listener at the present position is not on the central axis line of the right source. The distance between the two source centers is 1.0 m and the distances from the listener's head to the left and the right sources are 0.5 and 0.866 m, respectively. In the coordinate system O_1, the center of the listener head is $\mathbf{r}_{o1}=(0,0,0)$, the center of the left source sphere is $\mathbf{r}_{o2}=(0.6,30°,180°)$, and that of the right source sphere is $\mathbf{r}_{o3}=\mathbf{r}_{o2}+(1.1,90°,0°)$. The location of the vibrating cap center on the surface of the left sphere is (θ_L=150°, ϕ_L=0°) in its own coordinate system O_2, and the location of the vibrating cap on the surface of the right sphere is (θ_R=150°, ϕ_R=180°) in its own coordinate system O_3.

The left to right and the right to left channel separations of the CTC system, which are based on different playback plant transfer functions, are shown in Fig. 6. The regularization parameter β is 0.035. The dotted curves show the channel separation when the playback sound sources are the same as the setup ideal point sources, while the solid curves are for vibrating cap playback sources in the original pre-fixed orientation. The dashed curves show the channel separation, while the orientation of the right source of Fig. 5 is adaptively rotated to face the listener so that the listener is on the central axis line of the sound sources. This kind of rotation is sometimes hard to be implemented in practice; however, its simulation results are included in the figure to show the effects of the source orientation on the channel separation.

For this configuration, the center of the right source sphere is at $\mathbf{r}_{o3}=(0.966,60°,0°)$ and the location of the vibrating cap on its surface is (θ_R=120°, ϕ_R=180°) in its own coordinate system O_3. By comparing the dashed and dotted curves, it can be observed that the mismatch between the spatial and spectrum characteristics of the point sources and cap sources reduces the channel separation from over 40 dB to around 30 dB. By comparing the dashed and solid curves, it can be seen that the channel separation will be further reduced by a few decibels for the pre-fixed source orientations, especially in higher frequency range due to the higher directivity of the cap source. Unfortunately, it is often hard in practice to have the orientation of the right source rotated in realtime to face the listener so that the listener is on the original pre-fixed direction or central axis line of it. It can

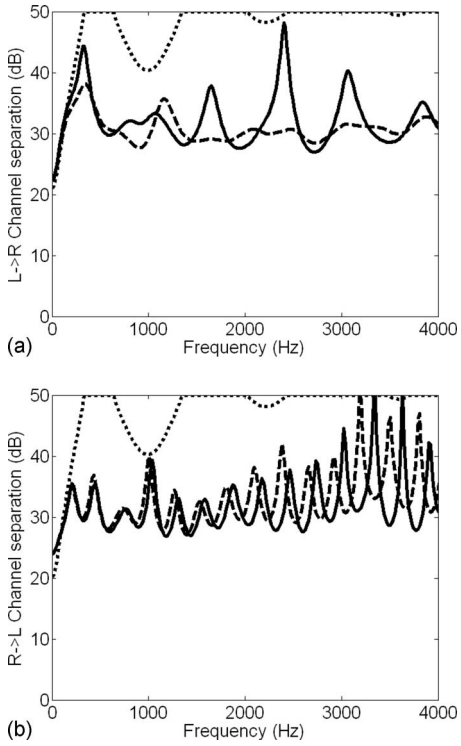


FIG. 6. Channel separations of a dynamic CTC system using different playback sound sources. The setup plant transfer functions are obtained with the point sources instead of the vibrating cap sources in Fig. 5. The playback sound sources are the same as the setup ones (dotted curve), or are the vibrating cap sources with the right source being rotated to face the listener head (dashed curve), or are the vibrating cap sources in original pre-fixed orientation (solid curve). (a) The left to right channel separation. (b) The right to left channel separation.

also be found that the right to left channel separation fluctuates more frequently than the left to right one, and this is because the distance between the listener's head and the right source sphere is larger than that with the left one.

The magnitude difference of the frequency responses between the present and the original pre-fixed orientations can be larger than 3 dB at 4000 Hz due to the directional properties of the source. Nevertheless, the performance loss is much smaller than that shown in Fig. 4, which illustrates that a small alteration of only 0.9 dB in terms of the difference in excess sound pressure level between two ears caused by the source directivity can reduce the channel separation from more than 40 to about 25 dB. The smaller performance loss for the dynamic CTC system can be explained by the fact, that even though the magnitude difference of the frequency responses in the present and the original pre-fixed directions caused by the source directional properties is larger than 3 dB at 4000 Hz, the change in the excess sound pressure level difference between two ears caused by the change in source orientation is much smaller than 3 dB as the two ear subtended angle seen from the source is much smaller than the value of the change in the source orientation angle (30° here). With the increase in the distance between the listener and the source, the two ears' subtended angle seen from the source becomes narrower, and this weakens the influence of the source directional properties on the channel separation of the dynamic CTC systems.

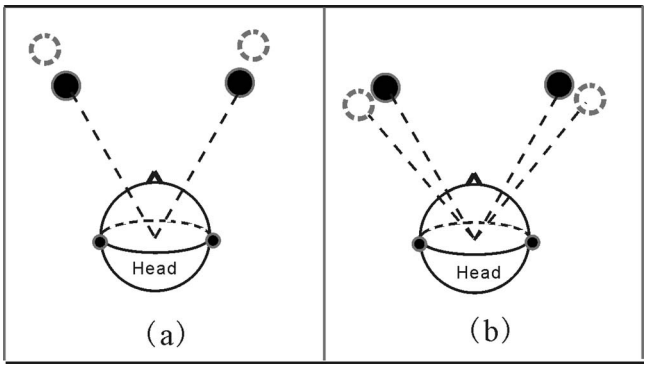


FIG. 7. The configurations for the CTC system with misaligned ideal point sound sources: (a) sources and the listener distance misalignment; (b) sources' subtended angle misalignment.

D. CTC systems with misaligned sound sources

In order to highlight the basic mechanisms, the impact of misalignments of the source-listener distance and the source subtended angle are illustrated by using the spherical head model with two ideal point sources, as shown in Fig. 7, while the effect of source orientation misalignment is illustrated by the cap sources. The channel separations of the CTC system with playback sound sources at different distance are depicted in Fig. 8, where the setup plant transfer functions are obtained when the distance between each of the sources and the listener is 1.0 m and the two sound sources subtend an angle of 60°. The playback functions are the same as the setup functions or obtained by only changing the distance from each of the sources to the listener to 1.1, 1.2, 1.3, 1.4, and 1.5 m along the original source-listener lines, as shown in Fig. 7(a). It can be observed from the figure that a small change in the distance along the original source-listener lines does not have a significant impact on the channel separations. Even if the distance is increased by 50% to 1.5 m, the average channel separation is still greater than 30 dB.

Figure 9 shows the channel separations of the CTC sys-

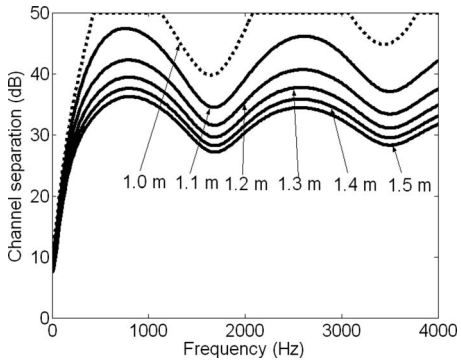


FIG. 8. Channel separations of the CTC system with distance misaligned playback sound sources. The setup plant transfer functions are obtained with the spherical head model with two ideal point sources when the distance between each of the sources and the listener is 1.0 m and the two sound sources subtend an angle of 60°. The playback ones are the same as the setup ones (dotted curve) or obtained by changing the distance from each of the sources to the listener to 1.1, 1.2, 1.3, 1.4, and 1.5 m (solid curves, the curve with the highest channel separation is that for 1.1 m, then 1.2 m, 1.3 m, 1.4 m, and the lowest curve is that for 1.5 m).

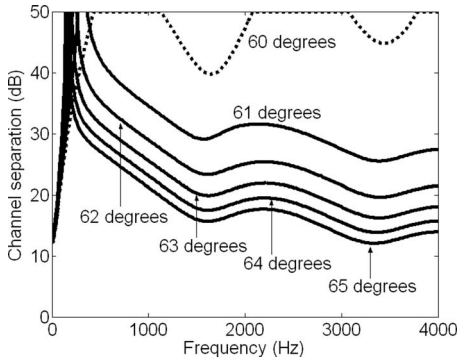


FIG. 9. Channel separations of the CTC system with subtended angle misaligned playback sound sources. The setup plant transfer functions are the same as that in Fig. 7. The playback ones are the same as the setup ones (dotted curve) or obtained by increasing the two source subtended angle from 60° with 1°, 2°, 3°, 4°, and 5° (solid curves, the curve with the highest channel separation is that for 1°, then 2°, 3°, 4°, and the lowest curve is that for 5°).

tem with the playback sound sources at different subtended angles as the configurations shown in Fig. 7(b). The setup plant transfer functions are the same as that used for Fig. 8. The playback functions are the same as the setup functions or obtained by increasing the two source subtended angle from 60° by 1°, 2°, 3°, 4°, and 5°, while the distance from the source to listener remains at 1.0 m. The figure underlines that changing the sources' subtended angle has a larger influence on the channel separation than changing the distance. The average channel separation drops to nearly 30 dB for a 1° change and to nearly 15 dB for a 5° change. It appears that this kind of change causes larger differences between the transfer functions between two ears and the source than those caused by changing the distance.

The loudspeaker box center (for example, the rigid sphere center in the model instead of the cap center) is sometimes improperly treated as the acoustic center of the sound source in practical applications. If the source orientation misalignment is referred to this box center, the channel separation drops significantly from more than 40 dB to about 20 dB with a misalignment of only 10°. Detailed analyses of preliminary simulations show that the main reason for this performance loss is the change in the sound source subtended angle seen from the listener, which is caused by the rotation of the sound source whose physical center (box or rigid sphere center) is not the acoustic center.[25] The performance loss caused by the misalignment of the subtended angle of the sources has already been shown in Fig. 9. In the following simulations, as shown in the configurations of Fig. 10, the relative positions (angle and the distance) between the acoustic center of the sound sources (the center of the sound cap) and the listener remain unchanged and only the orientations of the cap sources are changed to focus on the influence of the directional properties of the source.

Before the orientation of the cap source changes, the center of the listener head is $\mathbf{r}_{o1}=(0,0,0)$, the center of the left source sphere is $\mathbf{r}_{o2}=(1.1,30°,180°)$, and that of the right source sphere is $\mathbf{r}_{o3}=(1.1,30°,0°)$ in the coordinate system O_1 of Fig. 1. The location of the vibrating cap center on the surface of the left sphere is $(\theta_L=150°,\ \phi_L=0°)$ in its own coordinate system O_2 and that of the right vibrating cap is $(\theta_R=150°,\ \phi_R=180°)$ in its own coordinate system O_3. After rotating the orientation of the cap source α degree outside (positive value) or inside (negative value) with the center of cap source, the center of the left source sphere is $\mathbf{r}_{o2}=(1.0,30°,180°)+(0.1,30°-\alpha,180°)$ and that of the right source sphere is $\mathbf{r}_{o3}=(1.0,30°,0°)+(0.1,30°-\alpha,0°)$. The location of the vibrating cap center on the surface of the left sphere is $(\theta_L=150°+\alpha,\ \phi_L=0°)$ in its own coordinate system O_2 and the location of the vibrating cap on the surface of right sphere is $(\theta_R=150°+\alpha,\ \phi_R=180°)$ in its own coordinate system O_3. In the simulation results shown here, α was set as 15°.

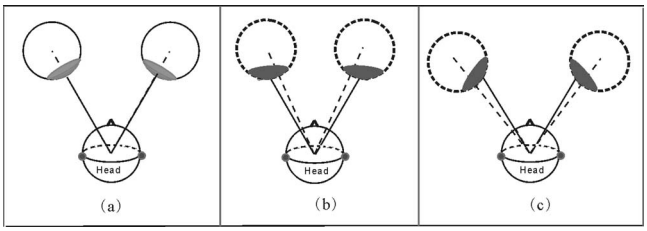


FIG. 10. The configurations for the CTC system with orientation misaligned directional sound sources. (a) The listener is on the central axis line of the sound sources, (b) the orientation of the cap sources is rotated 15° outside with the center of cap source, and (c) the orientation of the cap sources is rotated 15° inside with the center of cap source.

The channel separations of the CTC system caused by a change in the orientation of the directional source are shown in Fig. 11 where the regularization parameter β is 0.03. It becomes evident that the channel separation decreases significantly from more than 40 dB to about 25 dB, and the channel separation is larger for higher frequency because of the stronger directional properties of the source at high frequencies. It is also interesting to note that there are some frequencies (peaks in the curves) where the channel separation does not drop so much. These frequencies correspond to the impact of the scattering from the neighboring sphere, where the frequency responses differences at the two ears that can be observed for different configurations might be similar due to the scattering. By comparing the dashed curve and the solid curve, it can be found that the channel separa-

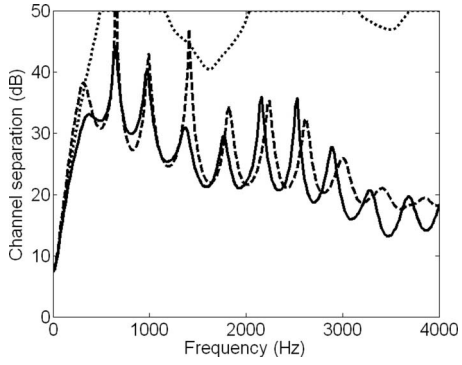


FIG. 11. Channel separations of the CTC system caused by the orientation change in the directional sources for the 60° source span system. The setup plant transfer functions are obtained with the original orientation. The playback ones are the same as the setup ones (dotted curve) or obtained by rotating the orientation of the cap sources 15° outside (dashed curve) or inside (solid curve) with the center of cap source.

tion is lower when the playback sound sources are rotated inside (solid curve). This might be explained by the larger scattering from the neighboring sphere due to the directivity of the cap source at high frequencies when the playback sound sources are rotated inside.

IV. DISCUSSIONS

By comparing the results of Fig. 11 with Fig. 9, where the two sources subtended angle changes by only a few degrees, it can be concluded that the influence on the channel separation of the CTC system, caused by the change in the directional source orientation, is smaller than the influence of the misalignment of the two sources subtended angle. Similar simulations were also carried out for the 20° source span system and when the source orientation rotates to different degrees. As the results are similar, they are not presented in this paper. All these results show that a slight change in the distance between the sources and the listener (still on the source-listener axis line) causes only a minor performance loss. The directional properties of the playback sound sources are responsible for a slight performance loss, especially at high frequencies, while the scattering from the neighboring sphere sometimes increases and sometimes decreases the channel separation in the middle frequency range. However, the most sensitive source misalignment is triggered by changing the subtended angle of the playback sound sources seen from the listener.

Extensive experiments have been carried out in the semi-anechoic chamber at the Institute of Technical Acoustics of the RWTH Aachen University to investigate the effects of using different playback sound sources, the channel separations of the dynamic CTC system with different source orientations, and the channel separations of the CTC system with misaligned sound sources.[26] The trends that could be detected from those experiment results are in concordance with that of the numerical simulations presented in this paper, so they are not given in this paper for conciseness. These results clearly show that the playback sound sources also play an important role in CTCs. Even though the different types of sources, the spatial characteristics (or directivity in the far field) of the sources, and positions of the sound sources all have a certain influence on the performance of the system, the channel separation of the CTC is most sensitive to the subtended angle of the playback sound sources. If highly accurate binaural cues are required in practical applications, the type and characteristics of the playback sound sources, their locations, and orientations all should be considered carefully.

For a CTC system, the factors that make the playback HRTFs mismatch the setup ones include the variation in individual HRTFs, the misalignment of the listener's head position and orientation, the misalignment of position and orientation of playback sound sources, the variation in the spatial and temporal characteristics of the playback sound sources, and the variation in the playback acoustics environment. The channel separation can decrease more than 15 dB just due to the difference between two individuals' HRTFs, and that caused by the position and orientation misalignment of the listener's head and/or the playback sound sources can be even larger depending on the degree of misalignment. It has also been found that the variations in the playback sound sources and the playback acoustics environment both can decrease the performance of the CTC system significantly. From all results obtained in this paper and the earlier studies by others on the robustness of the CTC systems, it can be deduced that even though any kind of playback and setup transfer function mismatches can reduce the channel separation performance of the CTC system to a certain extent, the CTC system is more sensitive to the mismatch that makes the change in propagation time from a source to one ear be in opposite direction to that to the other ear (reverse time response structures at two ears).[4–6,11]

For binaural virtual sound imaging systems that use two loudspeakers together with the CTC techniques, it has been proved that the system can only deliver accurate binaural cues when the setup and playback plant transfer functions are closely matched.[3] Two kinds of approaches can be used to address this problem: the dynamic (adaptive) CTC system and the robust design. The robust design means to optimize the physical properties (for example, locations of the sound sources) of the CTC systems so that its performance is not sensitive to the mismatch.[4,5] To implement either approach, the relation between the performance change and the basic parameters of the system need to be understood. From previous research and the findings in this paper, it is clear that a small change in the magnitude and phase difference between the transfer functions of two ears might reduce the performance of a CTC system significantly because of the coherent canceling between two waves. In the high frequency range, a small misalignment in space can cause a large change in phase, so that the robustness of the CTC system decreases in the high frequency range. Fortunately, human heads and loudspeaker boxes usually show more directionality in the high frequency range, and a hybrid system that uses both CTC and the inherent natural channel separation capabilities of sound sources and human head might offer more robustness and better channel separation performance.[4,24]

V. CONCLUSIONS

In the field of virtual acoustic imaging studies, there are several analytical models that are usually used. In the models such as the free field model and the spherical head model, the sound sources involved are treated as omnidirectional point sources. Based on spherical harmonics decomposition and the translational addition theorem, an analytical model has been developed in this paper that can take the scattering and spatial characteristics of the sound sources into consideration. The model is universal and could be useful when it comes to modeling similar systems with different sizes of spheres and with different positions and apertures of the cap sound sources. The model can also be extended and thus used for more sophisticated applications with multiple listeners and loudspeakers.

The effects of the spatial characteristics of a sound source and a misalignment on the performance of the CTC systems were studied based on the proposed model. It was

found that despite the fact that the different types of sources, the spatial characteristics of the sources, and the positions of the sound sources all have a certain influence on the system performance, the channel separation of the CTC is most sensitive to the misalignment of the subtended angle of the playback sound sources. The influence of the spatial characteristics of the source on the channel separation can be diminished somehow by placing the sources as far away from the listener as possible, so that the subtended angle from the source to two ears of the listener becomes smaller.

Future work can be extending the proposed model to investigate the influence of large reflective walls on the CTC system quantitatively and to investigate more sophisticated settings with multiple listeners and loudspeakers in a CAVE-like environment or an ordinary room.

ACKNOWLEDGMENTS

X.Q. is a research fellow of the Alexander von Humboldt Foundation and this research project is partially supported by Project No. 10674068 of NSFC. B.M. is a scholarship holder from CNPq, Brazil.

[1]M. Vorländer, *Auralization* (Springer-Verlag, Berlin, 2008), pp. 279–298.
[2]P. A. Nelson, M. Park, T. Takeuchi, and F. M. Fazi, "Binaural hearing and systems for sound reproduction (A)," J. Acoust. Soc. Am. **123**, 3469 (2008); in Proceedings of Acoustics'08 Conference, Paris, France (2008), pp. 3531–3536.
[3]M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," J. Acoust. Soc. Am. **121**, 1056–1069 (2007).
[4]M. R. Bai and C. Lee, "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," J. Acoust. Soc. Am. **120**, 1976–1989 (2006).
[5]T. Takeuchi, P. A. Nelson, and H. Hamada, "Robustness to head misalignment of virtual sound imaging systems," J. Acoust. Soc. Am. **109**, 958–971 (2001).
[6]J. Rose, P. A. Nelson, B. Rafaely, and T. Takeuchi, "Sweet spot size of virtual sound imaging systems at asymmetric listener locations," J. Acoust. Soc. Am. **112**, 1992–2002 (2002).
[7]J. Daniel, R. Nicol, and S. Moreau, "Further investigations of high order ambisonics and wavefield synthesis for holophonic sound imaging," in AES 114th Convention (2003), Paper No. 5788.
[8]A. J. Berkhout, D. de Vries, and P. Vogel, "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778 (1993).
[9]P. Gauthier and A. Berry, "Adaptive wave field synthesis for active sound field reproduction: Experimental results," J. Acoust. Soc. Am. **123**, 1991–2002 (2008).
[10]D. B. Ward, "On the performance of acoustic crosstalk cancellation in a reverberant environment," J. Acoust. Soc. Am. **110**, 1195–1198 (2001).
[11]T. Lentz, "Binaural technology for virtual reality," Ph.D. thesis, RWTH Aachen University, Aachen, Germany (2007).
[12]J. Garcia-Bonito and S. J. Elliott, "Local active control of diffracted diffuse sound fields," J. Acoust. Soc. Am. **98**, 1017–1024 (1995).
[13]Z. Lin, J. Lu, C. Shen, X. Qiu, and B. Xu, "Active control of radiation from a piston set in a rigid sphere," J. Acoust. Soc. Am. **115**, 2954–2963 (2004).
[14]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, London, 1999), pp. 183–232.
[15]M. Pollow, "Variable directivity of dodecahedron loudspeakers," Ph.D. thesis, RWTH Aachen University, Aachen, Germany (2007).
[16]D. Healy, D. Rockmore, P. Kostelec, and S. Moore, "FFTs for the 2-sphere-improvements and variations," J. Fourier Anal. Appl. **9**, 341–385 (2003).
[17]F. Zotter and R. Höldrich, "Modeling radiation synthesis with spherical loudspeaker arrays," in Proceedings of 19th International Congress on Acoustics, Madrid, Spain (2007).
[18]W. C. Chew, "Recurrence relations for three-dimensional scalar addition theorem," J. Electromagn. Waves Appl. **6**, 133–142 (1992).
[19]G. C. Gaunaurd, H. Huang, and H. C. Strifors, "Acoustic scattering by a pair of spheres," J. Acoust. Soc. Am. **98**, 495–507 (1995).
[20]H. Huang and G. C. Gaunaurd, "Acoustic scattering of a plane wave by two spherical elastic shells above the coincidence frequency," J. Acoust. Soc. Am. **101**, 2659–2668 (1997).
[21]N. A. Gumerov and R. Duraiswami, "Computation of scattering from N spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701 (2002).
[22]K. M. Li, W. K. Lui, and G. H. Frommer, "The diffraction of sound by an impedance sphere in the vicinity of a ground surface," J. Acoust. Soc. Am. **115**, 42–56 (2004).
[23]B. Masiero and X. Qiu, "Two listeners crosstalk cancellation system modelled by four point sources and two rigid spheres," Acta. Acust. Acust. **95**, 379–385 (2009).
[24]W. G. Gardner, "3-D audio using loudspeakers," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, (1997).
[25]X. Qiu, M. Vorländer, and B. Masiero, "Effects of practical loudspeaker characteristics on virtual acoustic imaging systems (A)," J. Acoust. Soc. Am. **123**, 3202 (2008); in Proceedings of Acoustics'08 Conference, Paris, France (2008), pp. 1719–1724.
[26]X. Qiu, B. Masiero, and M. Vorländer, "Experimental study on channel separation of crosstalk cancellation system with mismatched sound sources," in Proceedings of the Tenth Western Pacific Acoustics Conference, Beijing, China (2009).

Steady state and time-dependent energy equilibration in two-dimensional random elastic slabs

Yanyi Wan, Ying Wu, and Zhao-Qing Zhang[a)]
Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

(Received 29 December 2008; revised 24 July 2009; accepted 27 July 2009)

The static and dynamic transport properties of elastic wave propagation through two-dimensional random slabs without internal reflection were studied at two different scattering parameters: one for Rayleigh scattering and the other for Rayleigh–Gans scattering. The spatial distribution and temporal evolution of shear ($s-$) and compressional ($p-$) wave energy densities inside the slabs were calculated by solving the radiative transfer equation and the generalized diffusion equation (GDE). The comparison of their results can determine the region of validity of the GDE. The process of energy equilibration between the two wave modes was demonstrated explicitly as well as the process of diffusion. The depth inside a slab that is needed to reach energy equilibration or diffusive behavior is found to be dependent on source polarization. The results also show that the bulk equilibration ratio can be found inside a sample only when the sample is sufficiently thick. Deviations of the equilibration ratio from its bulk value are found near the output surface due to the absence of in-flow energy flux. The behavior of the deviations is sensitive to the scattering parameter but independent of source polarization.
 [DOI: 10.1121/1.3206581]

PACS number(s): 43.40.Fz, 43.40.Hb, 43.20.Fn [RLW] Pages: 1807–1816

I. INTRODUCTION

The study of wave propagation through random media has been an active field of investigation in the past few decades.[1] One important topic is the multiple scattering of elastic waves in complex solid media, particularly in relation to high-frequency seismic wave propagation[2] and the ultrasonic characterization of polycrystalline materials.[3–7] Unlike other classical waves, e.g., electromagnetic waves and acoustic waves, elastic waves have two modes of propagation, i.e., shear ($s-$) waves and compressional ($p-$) waves. After a sufficient number of random scatterings, the elastic waves come to a state of energy equilibration in the diffusive regime, which states that the ratio of the s-wave energy density to the p-wave energy density converges to a universal number independent of the fluctuations that cause the scattering as well as the polarization of the source.[4,8,9] According to the principle of equipartition, this number is believed to be the ratio of densities of states (DOSs) of the two modes in an unbounded medium[8] or, equivalently, $(v_p/v_s)^2$ for two-dimensional (2D) systems and $2v_p^3/(v_s^3)$ for three-dimensional (3D) systems, where v_p and v_s are, respectively, the compressional and shear wave speeds of the medium. The process of energy equilibration and equipartition has been seen previously near the surface of semi-infinite 3D media in both static and dynamic studies using radiative transfer equation (RTE).[5–7,10] In another study using Monte-Carlo simulation in a bulk 3D medium,[11] it has been found that the time scale and length scale to reach equipartition heavily depend on the scattering parameters. The equilibration time obtained from a p-wave source can be twice as long as compared with an s-wave source. Recently, energy equilibration has been observed in seismic coda waves created by earthquakes.[12] In a dynamic study, it has been found that due to the presence of a preferential absorption of one of the two modes, the equilibration ratio is shifted in favor of the mode that is less absorbed.[13] Previous study also showed that, when the energy equilibration ratio of an unbounded medium is reached, the propagation of the elastic wave energy can be described by a scalar diffusion equation with the total wave energy density as the only scalar quantity.[14] However, the equilibration ratio found in a bulk or semi-infinite system may not necessarily apply to a finite-sized random sample. For example, for a thin slab, the equilibration ratio can be different from its bulk value due to insufficient number of scatterings. Near the output boundary, the absence of in-flow energy flux from outside can also modify the equilibration ratio.[15]

In this work, by using both the RTE and the generalized diffusion equation (GDE), we studied theoretically various static and dynamic properties of elastic wave propagation through 2D random slabs without internal reflections. In particular, we are interested in the processes of energy equilibration and diffusion. Unlike the GDE, the RTE is valid for thin samples. However, the GDE is easier to solve for thick samples. In this work, we ignored the effects due to absorption and wave interference. In this case, the RTE can be obtained from the Bethe–Salpeter equation.[5] The comparison between the results obtained from the RTE and the GDE can tell us the region of validity of the GDE. In both approaches, elastic waves are expressed in terms of the energy densities for s-waves (E_s) and p-waves (E_p), separately.[14] Here we have assumed our systems are weakly-scattered media so

a)Author to whom correspondence should be addressed. Electronic mail: phzzhang@ust.hk

that the interference energy density can be ignored after taking configurational average.[16] For some strongly-scattered media, nonzero interference energy has been reported.[15] For the simplicity of our study, we have also assumed a matched boundary condition so that internal reflections due to the sample boundaries do not exist. The random systems considered here consist of randomly placed epoxy cylinders in an iron host. Both the cases of Rayleigh and Rayleigh–Gans scatterings were considered.[11] Different from previous studies in bulk[11] or semi-infinite media,[5–7,10] a slab geometry allows us to study energy equilibration near the output surface, which is independent of the source polarization when the sample is sufficiently thick.

In the study of static transport, we have observed the phenomenon of energy equilibration, with the energy density ratio converges to the value predicted by the DOS argument when the sample is thick enough. The depth inside the slab that is needed to reach within 1% of the bulk equilibration ratio is larger for a *p*-wave incidence. We have also found that the depth required for diffusion to set in is also larger for a *p*-wave incidence, and diffusion set in after energy equilibration has stabilized. The above results are true for both Rayleigh and Rayleigh–Gans scatterings. The fact that it takes more scatterings for the *p*-wave incidence to stabilize energy equilibration and reach a diffusive behavior than the *s*-wave incidence is because the *p*-wave energy is a minority component at equilibration and, therefore, more scatterings are required to convert the *p*-wave energy into the *s*-wave energy. Near the output boundary, energy equilibration ratio is found to be different from its bulk value due to the absence of backward scatterings. However, the behavior of the deviations depends on the scattering parameter. In addition, we have found a uniform extrapolation length for both *s*- and *p*-waves, and explained the result in terms of a weighted sum of the transport mean free paths of the two wave modes.

In the study of the dynamic transport, we are interested in the energy equilibration behavior near the output boundary. The equilibration ratio reached at long times is also found to be different from its bulk value. Moving into the sample, this ratio converges toward the bulk value only when the sample thickness is increased. These results are consistent with those found in the static transport. However, much larger sample thickness is needed to reach bulk equilibration ratio. Finally, through examining the long time decay rate of the energy density as a function of sample thickness, we have numerically confirmed the existence of the effective diffusion constant obtained from the GDE and a unified extrapolation length given as the weighted sum of the transport mean free paths of the two elastic wave modes.

This paper is organized as follows. In Sec. II, we summarize our theoretical framework by presenting the RTE and the GDE for the 2D elastic system. In Sec. III, we describe the physical system and outline the numerical calculations. The results for the static and dynamic transports are presented in Secs. IV and V, respectively. A summary of our main conclusions is given in Sec. VI.

II. THEORY

The 3D RTE for elastic waves and its diffusion approximation have been extensively studied in the past.[4–7,10,14,17] In the matrix notation, the 2D counterpart of the RTE has the form

$$\partial_t \underline{I}(\mathbf{r},t,\hat{s}) + \underline{\underline{V}} \cdot [\hat{s} \cdot \nabla_r \underline{I}(\mathbf{r},t,\hat{s})] + \underline{\underline{\tau}}^{-1} \cdot \underline{I}(\mathbf{r},t,\hat{s}) = \frac{1}{2\pi}\int_{2\pi} \underline{\underline{V}} \cdot \underline{\underline{P}}(\hat{s},\hat{s}') \cdot \underline{I}(\mathbf{r},t,\hat{s}')d\hat{s}', \tag{1}$$

where $\underline{I}(\mathbf{r},t,\hat{s})$ is the Stokes vector at position $\mathbf{r}$ at time t and pointing along direction $\hat{s}$. Both $\mathbf{r}$ and $\hat{s}$ are vectors on the 2D plane. $\underline{\underline{\tau}}$ and $\underline{\underline{V}}$ denote, respectively, the mean free time and wave velocity matrix, and have the forms

$$\underline{\underline{\tau}} = \begin{pmatrix} \tau_p & 0 \\ 0 & \tau_s \end{pmatrix} \quad \text{and} \quad \underline{\underline{V}} = \begin{pmatrix} v_p & 0 \\ 0 & v_s, \end{pmatrix}, \tag{2}$$

where the subscripts p and s denote for p and s waves, respectively. The Stokes vector in 2D has two components, representing the energy density current of p and s waves, respectively, i.e.,

$$\underline{I} \equiv \begin{pmatrix} I_p \\ I_s \end{pmatrix}. \tag{3}$$

Here we can define the scattering mean free paths l_p and l_s by

$$l_p^{-1} = (\tau_p v_p)^{-1} = \frac{1}{2\pi}\int_{2\pi} [P_{11}(\hat{s},\hat{s}') + P_{21}(\hat{s},\hat{s}')]d\hat{s}', \tag{4}$$

$$l_s^{-1} = (\tau_s v_s)^{-1} = \frac{1}{2\pi}\int_{2\pi} [P_{12}(\hat{s},\hat{s}') + P_{22}(\hat{s},\hat{s}')]d\hat{s}', \tag{5}$$

where the Mueller matrix, $\underline{\underline{P}}(\hat{s},\hat{s}')$, governs the scattering of the Stokes vector from directions $\hat{s}'$ to direction $\hat{s}$. For a medium containing discrete scatters, the Mueller matrix is related to the density of scatters, η, and the scattering differential cross-sections of a single scatterer through

$$\underline{\underline{P}}(\hat{s},\hat{s}') = \eta \begin{pmatrix} \frac{d\sigma}{d\theta}(p\hat{s}' \to p\hat{s}) & \frac{d\sigma}{d\theta}(s\hat{s}' \to p\hat{s}) \\ \frac{d\sigma}{d\theta}(p\hat{s}' \to s\hat{s}) & \frac{d\sigma}{d\theta}(s\hat{s}' \to s\hat{s}) \end{pmatrix}. \tag{6}$$

From the Stokes vector, we can obtain the local energy density, $E_{p,s}(\mathbf{r},t)$, and the local current density, $\vec{J}_{p,s}(\mathbf{r},t)$, through

$$E_{p,s}(\mathbf{r},t) = \int_{2\pi} I_{p,s}(\mathbf{r},t,\hat{s})/v_{p,s}d\hat{s}, \tag{7}$$

and

$$\vec{J}_{p,s}(\mathbf{r},t) = \int_{2\pi} \hat{s} I_{p,s}(\mathbf{r},t,\hat{s})d\hat{s}. \tag{8}$$

The RTE, i.e., Eq. (1), can be simplified by using diffusion approximation in the long time and long distance limit, which, for 2D systems, reads

$$I_{p,s}(\mathbf{r},t,\hat{s}) = \frac{1}{2\pi} v_{p,s} E_{p,s}(\mathbf{r},t) + \frac{1}{\pi} \vec{J}_{p,s}(\mathbf{r},t) \cdot \hat{s}. \tag{9}$$

Following the procedures given in Ref. 14, we obtain the following GDE:

$$\partial_t \begin{pmatrix} E_p(\mathbf{r},t) \\ E_s(\mathbf{r},t) \end{pmatrix} - \underline{\underline{D}} \cdot \triangle \begin{pmatrix} E_p(\mathbf{r},t) \\ E_s(\mathbf{r},t) \end{pmatrix} = - \underline{\underline{J}} \cdot \begin{pmatrix} E_p(\mathbf{r},t) \\ E_s(\mathbf{r},t) \end{pmatrix}, \tag{10}$$

where $\underline{\underline{D}}$ and $\underline{\underline{J}}$ are, respectively, the diffusion and collision matrices given by

$$\underline{\underline{D}} = \frac{1}{2N} \begin{pmatrix} v_p(\sigma_{ss} + \sigma_{sp} - \sigma_{ss}^*) & v_s \sigma_{sp}^* \\ v_p \sigma_{ps}^* & v_s(\sigma_{pp} + \sigma_{ps} - \sigma_{pp}^*) \end{pmatrix} \tag{11}$$

and

$$\underline{\underline{J}} = \frac{\eta}{v_p} \sigma_{ps} \begin{pmatrix} v_p^2 & -v_s^2 \\ -v_p^2 & v_s^2 \end{pmatrix}, \tag{12}$$

with $N = \eta[(\sigma_{ss}+\sigma_{sp}-\sigma_{ss}^*)(\sigma_{pp}+\sigma_{ps}-\sigma_{pp}^*)-\sigma_{sp}^*\sigma_{ps}^*]$. In the derivation, we have used the reciprocity relation for the total scattering cross-section $\sigma_{sp}/\sigma_{ps} = v_s/v_p$.[18] σ_{nm}^* is defined as $\sigma_{nm}^* = \int_{2\pi} (\hat{s}\cdot\hat{s}')(d\sigma/d\theta)(n\hat{s}' \rightarrow m\hat{s}) d\hat{s}$. The transport mean free paths can now be defined from the diffusion matrix, $\underline{\underline{D}}$, as

$$l_p^* = 2(D_{11} + D_{21})/v_p, \tag{13}$$

$$l_s^* = 2(D_{22} + D_{12})/v_s. \tag{14}$$

By diagonalizing the collision matrix $\underline{\underline{J}}$ in Eq. (10), we obtain

$$\partial_t \tilde{\underline{E}}(\mathbf{r},t) - \underline{\underline{\tilde{D}}} \cdot \Delta \tilde{\underline{E}}(\mathbf{r},t) = - \underline{\underline{\tilde{J}}} \cdot \tilde{\underline{E}}(\mathbf{r},t), \tag{15}$$

where the vector $\tilde{\underline{E}}(\mathbf{r},t)$ and the matrices $\underline{\underline{D}}$ and $\underline{\underline{J}}$ in the new basis become

$$\begin{pmatrix} \tilde{E}_1 \\ \tilde{E}_2 \end{pmatrix} = \frac{v_p^2}{v_s^2 + v_p^2} \begin{pmatrix} E_p + E_s \\ E_p - \frac{v_s^2}{v_p^2} E_s \end{pmatrix}, \tag{16}$$

$$\underline{\underline{\tilde{J}}} = \begin{pmatrix} 0 & 0 \\ 0 & 1/\tau_{eq} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \eta\sigma_{ps}(v_p^2 + v_s^2)/v_p \end{pmatrix}, \tag{17}$$

and

$$\underline{\underline{\tilde{D}}} = \frac{1}{\frac{v_s^2}{v_p^2} + 1} \begin{pmatrix} \frac{v_s^2}{v_p^2}(D_{11} + D_{21}) + (D_{12} + D_{22}) & D_{11} + D_{21} - D_{12} - D_{22} \\ \frac{v_s^2}{v_p^2} D_{11} - \frac{v_s^4}{v_p^4} D_{21} + D_{12} - \frac{v_s^2}{v_p^2} D_{22} & D_{11} - \frac{v_s^2}{v_p^2} D_{21} - D_{12} + \frac{v_s^2}{v_p^2} D_{22} \end{pmatrix}, \tag{18}$$

respectively. When the bulk equilibration ratio is established, the lower component of $\tilde{\underline{E}}(\mathbf{r},t)$ becomes zero since $E_s/E_p = v_p^2/v_s^2$. In this case, Eq. (15) is reduced to the following scalar diffusion equation for the total energy density $E_{\text{total}} = E_p + E_s$:

$$\partial_t E_{\text{total}}(\mathbf{r},t) - D_e \Delta E_{\text{total}}(\mathbf{r},t) = 0, \tag{19}$$

where D_e is the diffusion constant at energy equilibration and is given by

$$D_e = \frac{1}{1 + v_p^2/v_s^2}(D_{11} + D_{21}) + \frac{1}{1 + v_s^2/v_p^2}(D_{12} + D_{22}). \tag{20}$$

In the above derivations, two assumptions have been made. First, from the RTE, i.e., Eq. (1), to the GDE, i.e., Eq. (15), we have assumed that the p- and s- wave energy densities are each linear in $\hat{s}$, i.e., Eq. (9) is satisfied. This diffusion-type assumption is valid only after a sufficient number of scatterings so that the source information is completely lost. Thus, the reduction in the RTE to the GDE does not hold in thin samples. Second, when the GDE becomes valid in a thick sample, it will reduce to the DE, i.e., Eq. (19), when the bulk equilibration ratio has established. It will be shown later that the latter assumption is violated near the output boundary due to the absence of backward scatterings from outside the sample. However, in most practical situations, such as the systems considered in this work or in previous study, e.g., Ref. 21, the number of scatterings needed to establish bulk energy equilibration ratio in a thick sample is less than that needed for the diffusion to set in. In other words, if we move deeper into a thick sample from the input surface, the bulk equilibration ratio will stabilize first. A larger depth is required for Eq. (9) to hold. When Eq. (9) holds, the GDE will also give the result of the scalar DE because the bulk equilibration ratio has already established. However, when the output boundary is approached, the DE will become invalid again due to deviations from bulk equilibration ratio. In this case, we will show later that the GDE can give a reasonably good description of the energy density distributions.

III. NUMERICAL CALCULATIONS

In our study, we used a realistic system reported in Ref. 15. In that paper, the 2D elastic random medium was obtained by randomizing a 2D phononic crystal consisting of a square array of epoxy cylinders embedded in an iron host. The cylinders' axes are parallel to the y-axis. The radius of the cylinder is $r_s = 0.15a$, where a is the lattice constant of the phononic crystal. The s- and p-wave speeds in iron are $v_s = 3226.7$ m/s and $v_p = 5825.2$ m/s, respectively. Thus, the bulk equilibration ratio of this system is $E_s/E_p = v_p^2/v_s^2 = 3.26$. The wave speeds in epoxy are $v_{ss} = 1138.4$ m/s for

s-wave and $v_{ps}=2569.5$ m/s for p-wave. Our calculations were done at two frequencies with scattering parameters $k_s r_s=0.19$ and 2.26, where $k_s=2\pi/\lambda_s$ represents the wave vector of shear wave in the iron host. Both frequencies are located away from resonances so that the dwell time can be ignored.[15] The former one is in the Rayleigh regime, while the latter one is in the Rayleigh–Gans regime.[11] We calculated various scattering cross-sections for each case and the results are $\sigma_{ss}=0.0108a$, $\sigma_{sp}=0.00121a$, $\sigma_{ps}=0.00218a$, $\sigma_{pp}=0.00212a$, $\sigma^*_{ss}=-0.00284a$, $\sigma^*_{sp}=-0.000494a$, $\sigma^*_{ps}=-0.000893a$, and $\sigma^*_{pp}=-0.000878a$ for the case of $k_s r_s=0.19$ and $\sigma_{ss}=0.648a$, $\sigma_{sp}=0.141a$, $\sigma_{ps}=0.255a$, $\sigma_{pp}=0.402a$, $\sigma^*_{ss}=0.345a$, $\sigma^*_{sp}=0.00415a$, $\sigma^*_{ps}=0.00749a$, and $\sigma^*_{pp}=0.0597a$ for the case of $k_s r_s=2.26$. From these numbers, it is easy to see that more energy is scattered in the backward direction for the case of Rayleigh scattering, as all $\sigma^*_{\alpha\beta}$ are negative, whereas for the case of Rayleigh–Gans scattering, more energy is scattered in the forward direction. The scattering cross-sections are about two orders of magnitude smaller than the case of Rayleigh–Gans scattering. The above single-scattering properties in two different regimes are qualitatively similar to those considered in previous 3D studies.[5,11]

We would like to point out that, in the case of strong-scattering, the above system can have a nonzero averaged interference energy density $\langle E_i\rangle$ across the sample due to sharp impedance mismatch at epoxy boundaries. It was reported in Ref. 15 that the value $\langle E_i\rangle$ is about 10% of the total energy density when $\eta=1/a^2$ at $k_s r_s=2.26$. In this work, we ignored the interference energy by scaling down the number density from $\eta=1/a^2$ to $\eta=1/(s^2a^2)$ with $s>1$ as $\langle E_i\rangle \propto 1/s^2$ in the limit of large s. This is equivalent to increase the lattice constant of the original system from a to sa, while keep the values of $k_s r_s$ and r_s unchanged. From Sec. II, it is easy to see that the scaling down of number density is equivalent to scaling up all the mean free paths by a factor s^2, while keeping their relative magnitudes unchanged. Thus, if we take a as the length unit of the scatterers, it is convenient to introduce another length unit for mean free paths, i.e., $b\equiv 1/(\eta a)=s^2 a$. For the two cases considered here, the following mean free paths are obtained: $l_s=83.18b$, $l_p=232.53b$, $l^*_s=61.20b$, and $l^*_p=182.54b$ for $k_s r_s=0.19$ and $l_s=1.27b$, $l_p=1.52b$, $l^*_s=2.26b$, and $l^*_p=1.81b$ for $k_s r_s=2.26$. Here, $l_{s,p}$ and $l^*_{s,p}$ correspond to, respectively, the scattering and transport mean free paths.

Consider a plane-wave normally incident ($\theta=0$) on the front surface ($z=0$) of the slab of thickness L. For the static transport, we used time-independent incident energy and dropped all the time dependence in Eqs. (1) and (10). For dynamic transport, we used a delta-function incident energy in time. As noted earlier, we simplified our investigation by assuming non-reflecting boundaries so as to neglect the complication from the internal reflections.

The equivalent integral equation to Eq. (1) in the static case has the following form:

$$\underline{I}(z,\theta)=\underline{I}_0\delta(\theta-0)\begin{pmatrix} e^{-z/l_p} & 0 \\ 0 & e^{-z/l_s}\end{pmatrix} + \int_0^{2\pi}\int_0^z \frac{dx}{\cos\theta}\begin{pmatrix} e^{-(z-x)/l_p\cos\theta} & 0 \\ 0 & e^{-(z-x)/l_s\cos\theta}\end{pmatrix} \times \underline{\underline{P}}(\theta,\theta')\underline{I}(x,\theta')d\theta', \tag{21}$$

for the forward energy current with $0\le\theta<\pi/2$ and $3\pi/2<\theta<2\pi$, and

$$\underline{I}(z,\theta)=\int_L^z \frac{dx}{\cos\theta}\begin{pmatrix} e^{-(z-x)/l_p\cos\theta} & 0 \\ 0 & e^{-(z-x)/l_s\cos\theta}\end{pmatrix} \times\left[\int_0^{2\pi}\underline{\underline{P}}(\theta,\theta')\underline{I}(x,\theta')d\theta'\right], \tag{22}$$

for the backward energy current with $\pi/2<\theta<3\pi/2$. In obtaining the integral equations, we have incorporated the following boundary conditions:

$$\underline{I}(0,\theta)=\underline{I}_0\delta(\theta-0),\quad 0\le\theta<\frac{\pi}{2}\quad\text{and}\quad\frac{3\pi}{2}<\theta<2\pi,$$

$$\underline{I}(L,\theta)=0,\quad\frac{\pi}{2}<\theta<\frac{3\pi}{2}. \tag{23}$$

These integral equations can be readily solved by numerical iteration. For the dynamic transport, we worked in the frequency space.[19] With the substitutions $l_p^{-1}\to l_p^{-1}+i\omega/v_p$, and $l_s^{-1}\to l_s^{-1}+i\omega/v_s$, where ω is the frequency, the time-dependent equations become identical to the previous time-independent equations. The equations can be solved similarly through numerical iteration. The temporal results can then be retrieved through the Fourier transform:

$$\begin{pmatrix} I_p(z,\theta,t)\\ I_s(z,\theta,t)\end{pmatrix}=\int_{-\infty}^{\infty}\begin{pmatrix} I_p(z,\theta,\omega)\\ I_s(z,\theta,\omega)\end{pmatrix}e^{i\omega t}d\omega. \tag{24}$$

To solve the GDE, we used a line source close to the input surface of the slab at a penetration depth, z_p, inside the sample. We specify this energy source as $Q(z,t)=\delta(z-z_p)E_0(t)$. The value of z_p is determined by averaging a distribution of penetration depth weighted by the exponential attenuation of the coherent intensity.[20] For simplicity, we used the scattering mean free path of the incident wave as the penetration depth, i.e., $z_p=l_p$ or l_s depending on the polarization of the incident wave. We supplement the GDE with the following boundary conditions:

$$\begin{pmatrix} v_p & 0\\ 0 & v_s\end{pmatrix}\begin{pmatrix} E_p\\ E_s\end{pmatrix}\pm\frac{\pi}{2}\underline{\underline{D}}\frac{\partial}{\partial z}\begin{pmatrix} E_p\\ E_s\end{pmatrix}=0, \tag{25}$$

where the "+" sign is for $z=L$ surface, and the "−" sign is for $z=0$ surface. These equations are obtained from the flux conservation method mentioned in Ref. 14 for the non-reflecting boundaries.

In the static case, Eq. (10), with a source term, can be written as

$$-\underline{\underline{D}}\cdot\frac{d^2}{dz^2}\begin{pmatrix}E_p(z)\\E_s(z)\end{pmatrix}=-\underline{\underline{J}}\cdot\begin{pmatrix}E_p(z)\\E_s(z)\end{pmatrix}+\underline{E}_0\delta(z-z_p),\qquad(26)$$

where $\underline{E}_0=\binom{1}{0}$ for $z_p=l_p$ or $\underline{E}_0=\binom{0}{1}$ for $z_p=l_s$. The equation above can be solved analytically inside $[0,z_p]$ and $[z_p,L]$, respectively. The general solution has the form

$$\begin{pmatrix}E_p\\E_s\end{pmatrix}=(u_0+u_1z)\begin{pmatrix}1/v_p^2\\1/v_s^2\end{pmatrix}+[u_2e^{-K_mz}+u_3e^{K_mz}]\underline{\underline{D}}^{-1}\cdot\begin{pmatrix}1\\-1\end{pmatrix},\qquad(27)$$

where u_0, u_1, u_2, and u_3 are real constants to be determined, and $K_m=\sqrt{D_e/(\det\underline{\underline{D}}\cdot\tau_{\rm eq})}$, with $\tau_{\rm eq}$ given by $[\eta\sigma_{ps}(v_p^2+v_s^2)/v_p]^{-1}$. The determination of the eight constants, with four for the interval $[0,z_p]$ and four for the interval $[z_p,L]$, is obtained by two vector equations at $z=z_p$:

$$\underline{E}|_{z\to z_p^-}=\underline{E}|_{z\to z_p^+}\qquad(28)$$

and

$$\frac{d}{dz}\underline{E}|_{z\to z_p^-}-\frac{d}{dz}\underline{E}|_{z\to z_p^+}=\underline{\underline{D}}^{-1}\cdot\underline{E}_0,\qquad(29)$$

together with the boundary conditions given by Eq. (25).

In the dynamic case, the GDE, in the frequency space, has the form

$$-\underline{\underline{D}}\cdot\frac{d^2}{dz^2}\begin{pmatrix}E_{\omega,p}(z)\\E_{\omega,s}(z)\end{pmatrix}=-[\underline{\underline{J}}+i\omega\underline{\underline{I}}]\cdot\begin{pmatrix}E_{\omega,p}(z)\\E_{\omega,s}(z)\end{pmatrix}+\underline{E}_0\delta(z-z_p).\qquad(30)$$

The general solution in the regions $[0,z_p]$ and $[z_p,L]$ now has the form

$$\begin{pmatrix}E_{\omega,p}(z)\\E_{\omega,s}(z)\end{pmatrix}=(u_1+u_2i)e^{i\lambda_1z}\underline{v}_1+(u_3+u_4i)e^{-i\lambda_1z}\underline{v}_1+(u_5+u_6i)e^{i\lambda_2z}\underline{v}_2+(u_7+u_8i)e^{-i\lambda_2z}\underline{v}_2,\qquad(31)$$

where λ_1^2 and λ_2^2 are the two nondegenerate complex eigenvalues of the matrix $\underline{\underline{D}}^{-1}\cdot[\underline{\underline{J}}+i\omega\underline{\underline{I}}]$, and u_1 to u_8 are arbitrary real coefficients. The exact values of these coefficients for the specific solution are determined again by the conditions at $z=z_p$ and the boundary conditions at $z=0$ and $z=L$. The time-dependent results were then calculated by using Eq. (24).

IV. RESULTS FOR STATIC ENERGY TRANSPORT

We first discuss the case of Rayleigh–Gans scattering, i.e., $k_sr_s=2.26$. In Figs. 1(a) and 1(b), we show the static p-wave and s-wave energy density distributions, $E_p(z)$ and $E_s(z)$, inside a slab of thickness $L=7.9l_s$ for p- and s-wave incidences, respectively. In each figure, the open squares and diamonds denote, respectively, the energy density distributions for p- and s-waves obtained from the RTE. The dash-dotted and dotted curves are the results obtained from the GDE. The corresponding results for $L=15.8l_s$ are shown in Figs. 2(a) and 2(b). We noted that due to our use of the energy line source there exist un-physical kinks at $z=l_{p,s}$ in the diffusion results. This requires us to study the region away from the line source for comparing the results of the GDE and that of the RTE. For the p-wave incidence, we found from Fig. 1(a) that although good agreement is reached for $E_p(z)$ when $z\geq3.9l_s$, discrepancy persists in $E_s(z)$ for the entire range of z. However, from Fig. 2(a), we found good agreements for both $E_p(z)$ and $E_s(z)$ when $z>7.9l_s$. Similarly, for an s-wave incidence, both Figs. 1(b) and 2(b) indicate that diffusion theory is valid when $z>3.9l_s$. The larger sample size is needed for the p-wave incidence because that p-wave energy is the minor component at equilibration, so more scatterings are needed to redistribute the energy into the two wave modes from the incident p-wave energy.

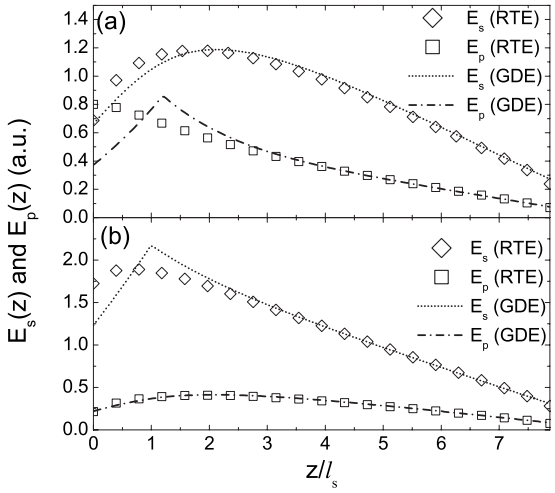


FIG. 1. The static energy density distributions in a sample with thickness $L=7.9l_s$. (a) The results are calculated from both the RTE and the GDE with pure p-wave incidence. ($E_s(z)$: s-wave energy density; $E_p(z)$: p-wave energy density.) (b) The results obtained with pure s-wave incidence.

To investigate energy equilibration, we calculated the ratio of the s-wave energy density to the p-wave energy density and compared it with the ratio $E_s/E_p=v_p^2/v_s^2$ predicted for an unbounded medium. In Figs. 3(a) and 3(b), we plotted the function $E_s(z)/E_p(z)$ inside the slab for $L=7.9l_s$ and $15.8l_s$, respectively. In each figure, the results of both s-wave

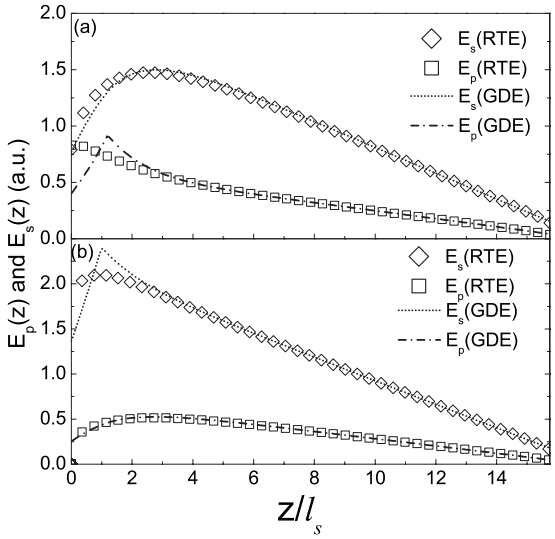


FIG. 2. Same as in Fig. 1, but with $L=15.8l_s$.

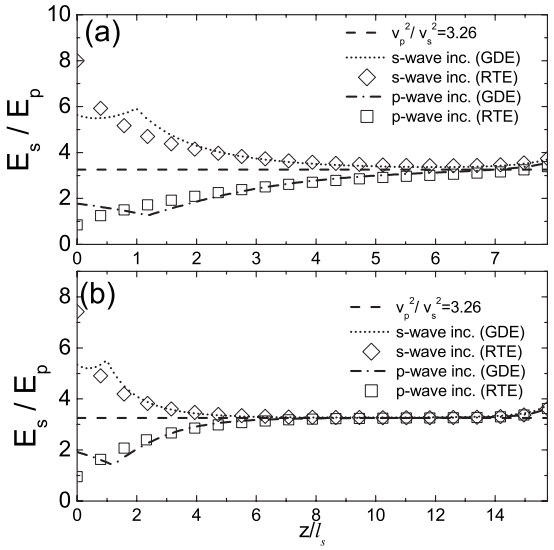


FIG. 3. (a) The energy density ratio in sample with thickness $L=7.9l_s$. The figure is obtained from the results in Figs. 1(a) and 1(b), for which the incident wave is p- and s-waves, respectively. The long-dashed line shows the bulk equilibration ratio. (b) same as (a) but with $L=15.8l_s$

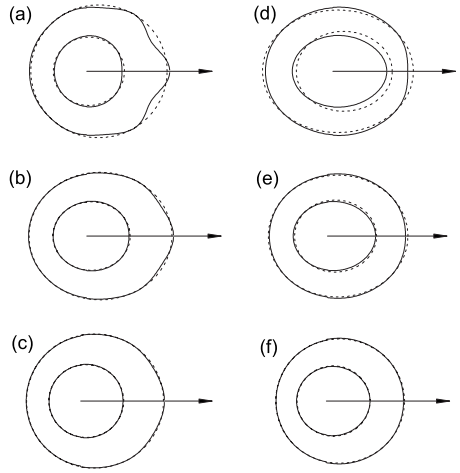


FIG. 4. Angular distributions of $I_{p,s}$ at $k_s r_s=2.26$. Inner circles represent I_p and outer circles are for I_s. The solid and dashed curves are obtained form RTE and GDE, respectively. (a)–(c) are results for s-wave incidence at depths of $5l_s$, $7l_s$, and $8l_s$, respectively. (d)–(f) are results for p-wave incidence at depths of $5l_s$, $7l_s$, and $9l_s$, respectively. The arrow in each figure shows the forward scattering direction, i.e., $\theta=0$.

and p-wave incidences are shown. The long-dashed horizontal line shows the bulk equilibration ratio, i.e., $v_p^2/v_s^2=3.26$. Figure 3(a) shows that complete energy equilibration is not reached inside the thin slab with thickness $L=7.9l_s$, whereas Fig. 3(b) shows that the bulk ratio is reached and stabilize when $L=15.8l_s$. However, from a more careful examination of this figure, we found that the s-wave incidence gives saturated ratios of 3.421 and 3.266, respectively, for $L=7.9l_s$ and $L=15.8l_s$, which are about 5% and 0.2% above the bulk ratio. Further examination on Fig. 3(b) reveals that the depth needed for reaching within 1% of its bulk equilibration ratio is about $6.5l_s$ for the s-wave incidence and $8.7l_s$ for the p-wave incidence. Again, we found that more scatterings are needed for the p-wave incidence to reach the bulk equilibration ratio. This is consistent with the previous finding of Ref. 11 in a dynamic study of 3D system. However, the above numbers seem to suggest that energy equilibration is stabilized after diffusion has set in. This is in contrast with another previous dynamic study in a bulk medium which showed that energy equilibration is stabilized before diffusion has set in.[21] In that work, the comparison was made on the angular distribution of energy density current, $I_{s,p}(z,\theta)$, rather than its average, $E_{s,p}(z)$. To make a similar comparison, we also plot in Fig. 4 the functions $I_{s,p}(z,\theta)$ obtained from both the RTE and the GDE at some depths inside the sample for both the s- and p-wave incidences. From this figure, it is found that the distance for diffusion to set in is about $8l_s$ for the s-wave incidence and $9l_s$ for the p-wave incidence. These depths are slightly longer than those needed for the stabilization of energy equilibration, and is in consistent with the finding of Ref. 21.

Figure 4 also implies that Eq. (15) of the GDE is valid only when $z>8l_s$ for the s-wave incidence and $z>9l_s$ for the p-wave incidence. In the region where the GDE is valid, we would expect that the scalar DE of Eq. (19) is also valid due to the establishment of the bulk equilibration ratio. However, the DE becomes invalid near the output boundary where deviations from the bulk equilibration ratio are found, as can be seen from Fig. 3. Near the output boundary, the equilibration ratio increases above its bulk value when output boundary is approached. In other words, the s-wave energy density tends to outweigh p-wave energy density at the boundary. However, this up-turn phenomenon is not universal. We will show later that the opposite is true for the case of $k_s r_s=0.19$. The origin of this phenomenon will be discussed at the end of this section. It should be pointed out that energy equipartition, in principle, cannot be fully completed in a finite-sized sample due to the presence of small diffusive currents, $\vec{J}_{p,s}$, appearing in the last term of Eq. (9) that makes $I_{p,s}$ slightly deviate from being isotropic.[21] This is clearly seen in Fig. 4, in which the angular distribution is not isotropic at $z=9l_s$, at which the energy equilibration has already stabilized. This is different from the case of semi-infinite media studied previously,[5–7] in which $I_{p,s}$ can be isotropic due to the absence of diffusive currents.

For scalar waves, when the wave energy becomes diffusive after a sufficient number of random scatterings, an extrapolation length, z_e, is normally introduced to describe the boundary condition of the corresponding diffusion equation. This length is defined and determined from the linear decay of the diffusive energy density inside the slab toward its boundary. In the case of a non-reflecting boundary, z_e has been determined from the Milne solution and found to be $0.7104l^*$ for 3D systems.[22] In 2D, the solution of the Bethe–Salpeter equation gives $z_e=0.82l^*$.[23] However, by using flux conservation at the sample boundary, one obtains $z_e=2l^*/3$ for 3D systems and $z_e=\pi l^*/4$ for 2D systems.[24,25] Here l^* is the transport mean free path. For the case of elastic waves studied here, due to mode-mode scatterings, energy distribution of either s waves or p waves does not decay linearly as can be seen from the last term in Eq. (27). However, when L is sufficiently large and the bulk equilibration ratio is approached, the first term on the right-hand side of Eq. (27), which gives the linear decay of the energy densities of both

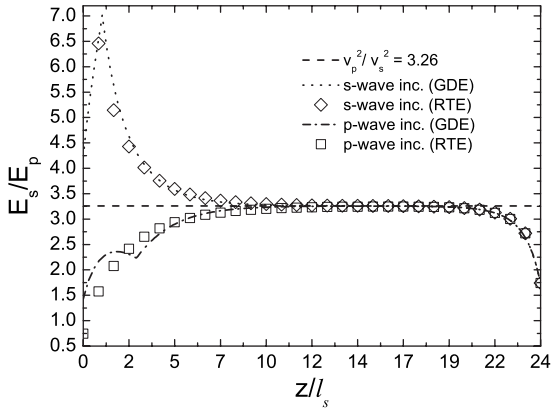


FIG. 5. (a) The energy density ratio in sample with thickness $L=24.0l_s$. The figure is the same as Fig. 3 while the sample is changed in which $k_s r_s = 0.19$.

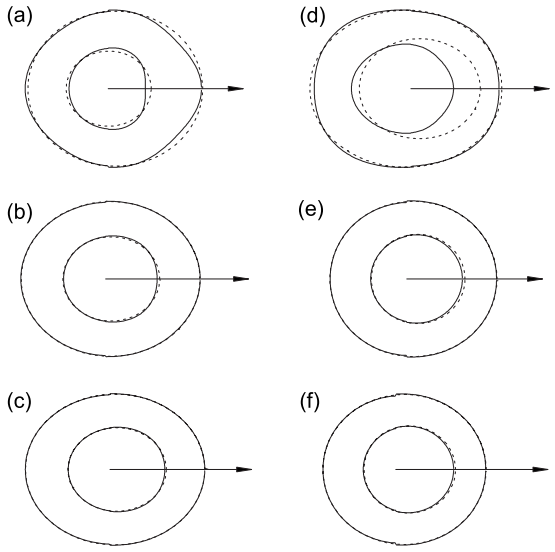


FIG. 6. The same as Fig. 4, but for the case of $k_s r_s = 0.19$. (a)–(c) are results for s-wave incidence at depths of $5l_s$, $9l_s$, and $11l_s$, respectively (d)–(f) are results for p-wave incidence at depths of $5l_s$, $16l_s$, and $18l_s$, respectively. The arrow in each figure shows the forward scattering direction, i.e., $\theta=0$.

wave modes and their bulk equilibration ratio, dominates the entire energy distribution and the second term is negligibly small except near the boundary, as shown in Fig. 3(b). Thus, if we ignore the deviation near the output boundary, we can still obtain a unified extrapolation length for both wave modes from the relation $z_e = -u_0/u_1 - L$, where u_0 and u_1 are constants in Eq. (27). For samples of thicknesses $L=15.8l_s$, $23.6l_s$, and $39.4l_s$, we found $z_e = 1.3l_s$, independent of sample thickness. It is interesting to relate the value $z_e = 1.3l_s$ found here to the weighted sum of the transport mean free path of the two wave modes, i.e.,

$$l^*_{\text{eff}} = \frac{1}{1+v_p^2/v_s^2} l^*_p + \frac{1}{1+v_s^2/v_p^2} l^*_s, \tag{32}$$

which gives the value $1.7l_s$. With this definition of the effective mean free path, we have $z_e \cong (\pi/4) l^*_{\text{eff}}$, which is consistent with the result found for 2D scalar waves.[25]

For the case of Rayleigh scattering, i.e., $k_s r_s = 0.19$, we plot the ratio of s-wave energy to p-wave energy in a slab with a thickness of $L=24.0l_s$ in Fig. 5. In this case, the p-wave incidence gives a saturated ratio of 3.248, which is about 0.4% below the bulk ratio. The depth required for reaching within 1% of the equilibration ratio is $9.1l_s$ for the s-wave incidence and $10.2l_s$ for the p-wave incidence. These numbers are larger than the corresponding numbers found for the case of Rayleigh–Gans scattering and, therefore, depend on the scattering parameter. Similar conclusion has been found in Ref. 11. In Fig. 6, we also plot the functions $I_{s,p}(z,\theta)$ obtained from both the RTE and the GDE at certain depths. From this figure, we find the depths for diffusion to set in, or the GDE to become valid, are about $11l_s$ and $18l_s$ for the s-wave and p-wave incidences, respectively. These numbers are again consistent with the findings of Ref. 21. Here we also find the depth required for diffusion to set in for the p-wave incidence is about twice that required for s-wave incidence. This is different from the case of Rayleigh–Gans scattering, where two depths are comparable to each other. The difference, we believe, is due to a larger ratio of l_p/l_s for the case of $k_s r_s = 0.19$, which is 2.8, than the case of $k_s r_s = 2.26$, which is 1.2. This is, again, consistent with the previous conclusion drawn for the case of Rayleigh–Gans scattering that more scattering is needed for p-wave incidence to reach diffusion.

One interesting phenomenon exhibited in Fig. 5 is that there is a decrease in the energy ratio when the output boundary is reached. This is different from the previous case of $k_s r_s = 2.26$, which exhibits a slight up-turn in the energy ratio near the output boundary. These deviations from the bulk equilibration ratio arise from the incomplete scatterings at the output boundary. This can be understood from the following qualitative analysis of the forward and backward scattered energy densities. We first enlarge the previous samples by doubling their thicknesses. Near the middle of the new samples, the ratios of E_s/E_p is close to 3.26. Then, we estimated, at the middle points of the samples, the ratios of s-wave energy to p-wave energy contributed by the backward current coming from the right-half of the sample. The ratio, to the first order approximation, can be estimated as $[E_s(\sigma_{ss} - \sigma^*_{ss}) + E_p(\sigma_{ps} - \sigma^*_{ps})]/[E_s(\sigma_{sp} - \sigma^*_{sp}) + E_p(\sigma_{pp} - \sigma^*_{pp})]$, which gives the values of 1.64 and 5.57 for $k_s r_s = 2.26$ and 0.19, respectively. The former number is smaller than the bulk ratio 3.26, while the latter number is larger. This implies that for the case of $k_s r_s = 2.26$ the s-wave energy density has less contribution from the backward current than the p-wave. Since in our study we used the non-reflection boundary condition, there is no backward current at the sample boundary, which is equivalent to moving away the right-half of the double-sized sample. Therefore, the ratio of s-wave energy to p-wave energy increases accordingly. On the contrary, for the case of $k_s r_s = 0.19$, the s-wave energy density has more contribution from the backward current than p-wave. Thus, the cutoff of the backward current gives rise to a decrease in the ratio of the s-wave energy to p-wave energy at the slab boundary. Here we do not consider mismatched boundary condition. Internal reflections caused by a mismatched boundary can provide backward currents which tend to restore the bulk equilibration ratio. It has been reported that the

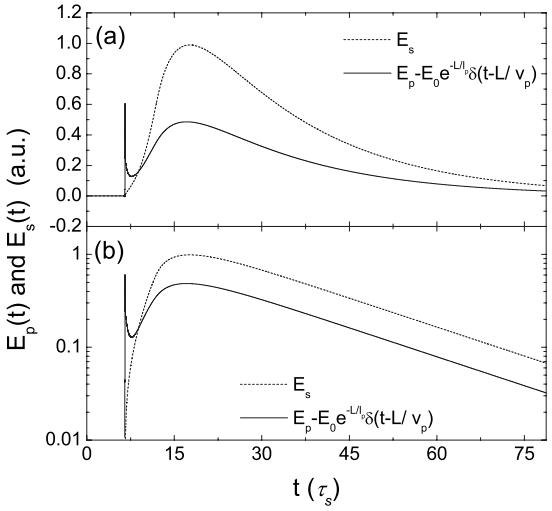


FIG. 7. (a) The time-resolved transmitted energy densities at the output surface of a slab with $L=11.8l_s$. The results are calculated by RTE with pure p-wave incidence. $E_s(t)$ for s-wave energy density and $E_p(t)$ for p-wave energy density. (b) Re-plot of (a) with the vertical axis changed to log-scale. Time unit is the mean free time of the s-wave.

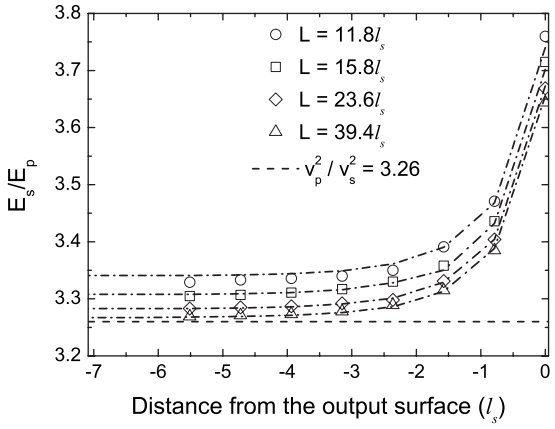


FIG. 8. Dynamic equilibration ratio inside slabs with different thicknesses, obtained by solving the RTE. The horizontal axis shows the positions inside the slab as measured by their distance from the output surface. The dot-dashed lines are the results obtained by solving the GDE for corresponding sample sizes. The analytical equilibration ratio is presented by the long-dashed line.

presence of a fluid-solid interface in a semi-infinite 3D medium can increase the approach to the diffusive limit and bulk energy equilibration near the interface.[7] However, such an interface can also produce a nonzero interference energy in a distance of a few shear wavelengths near the interface as has been observed in seismic coda waves.[12]

V. RESULTS FOR DYNAMIC ENERGY TRANSPORT

To study the dynamic energy transport, we used a delta-function pulse incident onto one surface of a slab of thickness L at time $t=0$. Here, we, again, focus our discussion on the case of $k_s r_s=2.26$. By solving the time-dependent RTE, we can obtain a time-resolved energy density, $E_{s,p}(t)$, at any position along the slab. Here we focus on various time-resolved energy densities close to the output surface. A typical result for the p-wave incidence with $11.8l_s$ is shown in Fig. 7(a), in which the time unit is the mean free time, $\tau_s = l_s/v_s$, of the s-wave. The coherent energy component $E_0 e^{-L/l_p}\delta(t-L/v_p)$ has been subtracted, so the figure shows only the energy densities of the scattered waves. We observed that the p-wave energy density exhibits a sharp peak and followed by a diffusive tail. The sharp peak represents the wave energy that transports through the sample ballistically and is formed by the forward-scattered energy from a few scatterings. As these waves propagate almost straightly toward the output boundary, the transport time is close to that of the coherent beam, i.e., $t_{\text{peak}}=L/v_p$. This ballistic peak decays exponentially with L. The diffusive tail represents the wave energy that arrives at the output surface after many scatterings inside the sample. According to the conventional diffusion theory, the transport time of the diffusive peak is proportional to L^2. The results for pure s-wave incidence have similar features.

Beyond the diffusive peak, both s-wave and p-wave energy densities exhibit exponential decay in time, in agreement with the conventional diffusion theory. This exponential decay is also associated with the energy equilibration between s-waves and p-waves, i.e., the ratio of the two energy densities reaches a constant value. These two phenomena can be easily demonstrated by re-plotting the energy densities in log-scale, as shown in Fig. 7(b). The two parallel straight lines at later times for the s-wave and p-wave energy densities indicate that energy equilibration has been reached. Similar to the static case shown in Fig. 3, the equilibration ratio found here does not depend on the source polarization and is higher than the bulk ratio 3.26 near the output boundary. The discrepancy can partly be attributed to the same boundary condition arguments as discussed in Sec. IV for the static transport. When we move our measurement position away from the output surface inside the sample, the influence of the boundary decreases, and the equilibration ratio decreases gradually and saturates to a constant value, which is about 3.33, when the distance from the output surface is about $5.5l_s$. This is shown by open circles in Fig. 8. It is interesting to notice that the saturated ratio is still 2% above the bulk ratio. The reason, we believe, is the existence of the other boundary. This can be checked by using a larger L. The results of $L=15.8l_s, 23.6l_s$, and $39.4l_s$ are also plotted in Fig. 8. The saturation ratios indeed approach the bulk value when L is increased. It reaches to about 0.2% of the bulk value when $L=39.4l_s$, which is about two to three times the size required for the static equilibration ratio to reach about 0.2% of the bulk equilibration ratio, as shown in Fig. 3(b). The reason for a larger size to reach the same degree of equilibration in the dynamic case is probably because of the difference in the static and dynamic spatial energy distributions. In the static case, the energy distribution is peaked close to the input surface, as shown in Figs. 1 and 2, whereas in the dynamic case, the peak of the energy distribution at long time is always at the middle of the sample. Similar results are found from the GDE, which are shown by dash-dotted lines in Fig. 8. Excellent agreements between the RTE and GDE are evident at long times when $L>11.8l_s$. We have fitted the saturated equilibration ratio obtained from the RTE

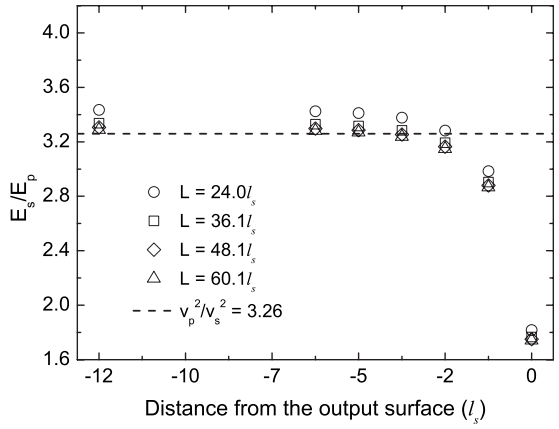


FIG. 9. Dynamic equilibration ratio inside slabs with different thicknesses for $k_s r_s = 0.19$. The results are obtained by solving the GDE. The horizontal axis shows the positions inside the slab as measured by their distance from the output surface.

by a function $y = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ with $x = 1/L$ and found $a_0 = 3.253$, which is consistent with the bulk ratio of 3.259 obtained in an unbounded medium.

We also studied the exponential decay of the total energy densities in the long time regime by using the result of scalar diffusion theory, i.e., by assuming $E(t) \propto e^{-\alpha t}$ with a decay rate

$$\alpha = \frac{D\pi^2}{(L + 2z_e)^2}, \tag{33}$$

where D is the diffusion constant and z_e is the extrapolation length.[26] By fitting Eq. (33) to the calculated decay rates at $L/l_s = 11.8$, 15.8, 23.6, and 39.4, we obtained both the extrapolation length, $z_e = 1.3 l_s$, and the effective diffusion constant, $D = 1.0 l_s v_s$. These values are consistent with the value obtained previously in the static transport and Eq. (20), i.e., $D_e = 0.98 l_s v_s$, at energy equilibration.

For the case of $k_s r_s = 0.19$, we plot in Fig. 9 the dynamic equilibration ratio inside slabs with different sample thicknesses as a function of the distance away from the output surface. These results are obtained from the GDE. Similar to the static case shown in Fig. 5, Fig. 9 exhibits a down-turn behavior in the equilibration ratios near the output boundary due to the absence of backward scatterings. Away from the output boundary, the equilibration ratio increases and overshoots the bulk equilibration ratio before it saturates. Such an overshot behavior is not seen in the case of $k_s r_s = 2.26$. The reason for the overshot is not clear. The saturated ratios approach the bulk value when the sample becomes thicker. The ratio reaches within 1% of the bulk value when the sample thickness is as large as $60.1 l_s$, which is, again, larger than the size required for the static case.

VI. CONCLUSIONS

In summary, we have studied the static and dynamic energy transports of 2D elastic waves through random slabs by using both the RTE and the GDE. The GDE can be reduced to a scalar DE with an effective diffusion constant when the bulk energy equilibration ratio is established. By comparing the results obtained from the RTE and the GDE, we are able to determine the region of validity of the GDE, which is much easier to solve than the RTE. Both Rayleigh and Rayleigh–Gans scatterings were considered. We have demonstrated the processes of energy equilibration and diffusion. We have shown that more scatterings are needed to establish a diffusive behavior than to stabilize the bulk energy equilibration ratio in a thick sample, independent of source polarization. This implies that whenever the GDE is valid after diffusion has set in, it also gives the result of the scalar DE. However, the DE becomes invalid near the output boundary, where deviations from the bulk equilibration ratio always occur due to the absence of backward scattering. Such deviations are independent of source polarization but are sensitive to the single-scattering properties of the random media.

We have also showed that a deeper depth is needed to establish a diffusive behavior and energy equilibration when the incident wave mode belongs to the minority component at equilibration. The precise depth depends on the scattering properties of a single scatterer and can be quite different between Rayleigh and Rayleigh–Gans scatterings. The bulk equilibration ratio can only be established inside the sample when the sample thickness is sufficiently large. And much larger sample thickness is required for the dynamic equilibration ratio to reach its bulk value. Finally, we have obtained a valid expression for the extrapolation length for the diffusion equation in elastic waves when the bulk equilibration ratio is reached inside a sample. Although the results obtained here are for 2D systems, we expect that their qualitative features will also hold for 3D systems.

ACKNOWLEDGMENT

The work was supported by Hong Kong RGC Grant No. 604703.

[1] P. Sheng, *Introduction to Wave Scattering, Localization, and Mesoscopic Phenomena* (Academic, San Diego, 1995).
[2] H. Sato and M. Fehler, *Seismic Wave Propagation and Scattering in the Heterogeneous Earth* (Springer, New York, 1998).
[3] S. Hirsekorn, "The scattering of ultrasonic waves by multiphase polycrystals," J. Acoust. Soc. Am. **83**, 1231–1242 (1988).
[4] R. L. Weaver, "Diffusivity of ultrasound in polycrystals," J. Mech. Phys. Solids **38**, 55–86 (1990).
[5] J. A. Turner and R. L. Weaver, "Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **96**, 3675–3683 (1994).
[6] J. A. Turner and R. L. Weaver, "Time dependence of multiply scattered diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **97**, 2639–2644 (1995).
[7] J. A. Turner and R. L. Weaver, "Ultrasonic radiative transfer in ploycrystalline media: Effects of a fluid-solid interface," J. Acoust. Soc. Am. **98**, 2801–2808 (1995).
[8] R. L. Weaver, "On diffuse waves in solid media," J. Acoust. Soc. Am. **71**, 1608–1609 (1982).
[9] G. C. Papanicolaou, L. V. Ryzhik, and J. B. Keller, "Stability of the P-to-S energy ratio in the diffusive regime," Bull. Seismol. Soc. Am. **86**, 1107–1115 (1996).
[10] J. A. Turner and R. L. Weaver, "Radiative transfer of ultrasound," J. Acoust. Soc. Am. **96**, 3654–3674 (1994).
[11] L. Margerin, M. Campillo, and B. A. Tiggenlen, "Monte Carlo simulation of multiple scattering of elastic waves," J. Geophys. Res. **105**, 7873–7892 (2000).
[12] R. Hennino, N. P. 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).

[13]L. Margerin, B. A. van Tiggelen, and M. Campillo, Bull. Seismol. Soc. Am. **91**, 624 (2001).
[14]N. P. Trégourès and B. A. van Tiggelen, "Generalized diffusion equation for multiple scattered elastic waves," Waves Random Media **12**, 21–38 (2002).
[15]Y. Wu, Y. Lai, Y. Wan, and Z. Q. Zhang, "Wave propagation in strongly scattered random elastic media: Energy equilibration and crossover from ballistic to diffusive behavior," Phys. Rev. B **77**, 125125 (2008).
[16]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[17]L. V. Ryzhik, G. C. Papanicolaou, and J. B. Keller, "Transport equations for elastic and other waves in random media," Wave Motion **24**, 327–370 (1996).
[18]V. Varatharajulu, "Reciprocity relations and forward amplitude theorems for elastic waves," J. Math. Phys. **18**, 537–543 (1997).
[19]R. Elaloufi, R. Carminati, and J. Greffet, "Time-dependent transport through scattering media: From radiative transfer to diffusion," J. Opt. A, Pure Appl. Opt. **4**, S103–S108 (2002).
[20]J. G. Rivas, R. Sprik, A. Lagendijk, L. D. Noordam, and C. W. Rella, "Static and dynamic transport of light close to the Anderson localization transition," Phys. Rev. E **63**, 046613 (2001).
[21]A. Paul, M. Campillo, L. Margerin, E. Larose, and A. Derode, "Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves," J. Geophys. Res. **110**, B08302 (2005).
[22]G. Placzek and W. Seidel, "Milne's problem in transport theory," Phys. Rev. **72**, 550–555 (1947).
[23]Y. Lai, S. K. Cheung, and Z. Q. Zhang, "Wave transport in two-dimensional random media: The ballistic to diffusive transition and the extrapolation length," Phys. Rev. E **72**, 036606 (2005).
[24]J. X. Zhu, D. J. Pine, and D. A. Weitz, "Internal reflection of diffusive light in random media," Phys. Rev. A **44**, 3948–3959 (1991).
[25]M. Haney and R. Snieder, "Breakdown of wave diffusion in 2D due to Loops," Phys. Rev. Lett. **91**, 093902 (2003).
[26]For example, G. H. Watson, Jr., P. A. Fleury, and S. L. McCall, "Searching for photon localization in the time domain," Phys. Rev. Lett. **58**, 945–948 (1987).

On the correlation of non-isotropically distributed ballistic scalar diffuse waves

Richard Weaver
Department of Physics, University of Illinois at Urbana-Champaign, 1110 W. Green, Urbana, Illinois 61801

Berenice Froment and Michel Campillo
Laboratoire de Géophysique Interne et Tectonophysique, Université Joseph Fourier, BP53 38041, Grenoble France

(Received 30 January 2009; revised 8 July 2009; accepted 10 July 2009)

Theorems indicating that a fully equipartitioned random wave field will have correlations equivalent to the Green's function that would be obtained in an active measurement are now legion. Studies with seismic waves, ocean acoustics, and laboratory ultrasound have confirmed them. So motivated, seismologists have evaluated apparent seismic travel times in correlations of ambient seismic noise and tomographically constructed impressive maps of seismic wave velocity. Inasmuch as the random seismic waves used in these evaluations are usually not fully equipartitioned, it seems right to ask why it works so well, or even if the results are trustworthy. The error, in apparent travel time, due to non-isotropic specific intensity is evaluated here in a limit of large receiver-receiver separation and for the case in which the source of the noise is in the far field of both receivers. It is shown that the effect is small, even for cases in which one might have considered the anisotropy to be significant, and even for station pairs separated by as little as one or two wavelengths. A formula is derived that permits estimations of error and corrections to apparent travel time. It is successfully compared to errors seen in synthetic waveforms.
 [DOI: 10.1121/1.3203359]

PACS number(s): 43.40.Ph, 43.50.Yw, 43.60.Cg [RAS] Pages: 1817–1826

I. INTRODUCTION

Much recent research has focused on the correlations of seismic, ocean acoustic, and laboratory ultrasonic noise. Theorems indicating that a fully equipartitioned noise field will have correlations $C(\tau)$ essentially equivalent to the Green's function $G(\tau)$ that would be obtained in an active measurement are now legion.[1–4] These have been supported by laboratory experiments[5–10] and analysis of ocean acoustic and seismic field data.[11–13] The identity promises to facilitate the acquisition of acoustic information without the use of a controlled source. Applications in seismology and exploration geophysics where sources are earthquakes or thumper trucks or explosives are especially interesting. The conditions on acoustic noise such that its correlations will indeed converge to the Green's function, and the rate and quality of that convergence, remain active areas for inquiry. Many questions relate to the robustness of the identity for the case of imperfectly diffuse noise fields, the effect of only partial equipartition, and the effect of dissipation. Questions as to how one might compensate in the case of imperfectly diffuse noise are also arising.[14] Reviews may be found in the special June 2006 issue of *Geophysics*.[15,16] A tangential literature has entertained sundry generalizations of the basic identity, for example, to media lacking time-reversal invariance,[17,18] to dissipative media[19] and to the diffusion equation,[20] and to structural acoustics and discrete media.[21,22]

It has long been recognized that field-field correlations of narrow-band diffuse waves have a universal *local* short range structure equal to a Bessel function of order zero. Rollwage *et al.*[23] showed that diffuse fields with wavenumber k in a shallow water tank have correlations $\langle\psi(\mathbf{x})\psi(\mathbf{y})\rangle \sim J_o(\mathbf{k}|\mathbf{x}-\mathbf{y}|)$. It is well known that the ensemble average (over different realizations of a multiple scattering medium) of field-field correlations in three dimensions is a spherical Bessel function of order zero as attenuated by scattering $\langle\psi(\mathbf{x})\psi(\mathbf{y})\rangle \sim J_o(k|\mathbf{x}-\mathbf{y}|)\exp\{-|\mathbf{x}-\mathbf{y}|/\text{meanfreepath}\}$. The microtremor survey method[24,25] suggested by Aki in 1957 and in wide use in seismology today is based on this notion of local correlations being essentially Bessel functions. In all this work it seems to have been little recognized that the Bessel-character of the short range local correlations has an extension to long ranges and to the time domain. It transpires that such fields have correlations equal to the *imaginary part of Green's function*[21] in turn equal to a Bessel function only at short range in an unbounded continuous homogeneous scalar medium. Field-field correlations are thus in general richer than simple Bessel functions; they depend on the type of wave, and include effects from the structure and geometry of the medium.

Proof of the identity between correlations and Green's function varies with definition of a diffuse field. In finite bodies it is convenient to take a modal perspective, in which a diffuse field has uncorrelated normal mode amplitudes but with equal expected energies. This definition is commonly adopted in room acoustics and structural vibration[26,27] and in thermal physics.

In the so-called ballistic case, with few scatterers, the proofs are especially simple and intuitive.[2,4] It is this case that pertains to travel time tomography and attracts the most attention in seismology.[16,28,29] Seismologists have con-

structed high resolution maps of Rayleigh and Love wave velocities from tomographic analyses of travel times seen in correlations of ambient seismic noise.[28,29]

An imperfectly equipartitioned diffuse field precludes confident application of field correlation methods, or so one imagines. Ambient seismic noise as used in Refs. 28 and 29 at frequencies below 1 Hz, usually has its origin in ocean storms and thus has a preferred direction. Seismic coda, as used in Ref. 30 has a degree of isotropy that develops slowly as the coda ages. Laboratory experiments[5–10] with ultrasound can be designed to better conform to the demands of the theorems but, except for the case of thermal noise,[5,31,32] even their correlations fail to perfectly match conventional waveforms obtained actively. Nevertheless noise correlations continue to provide useful information. They have been used to detect changes in material properties,[33,34] and most strikingly have been used in spite of the imperfect equipartition to generate high resolution maps of seismic velocity.[28,29] These maps appear to be robust, but doubts remain. How reliable are they? Why do they appear so robust in spite of the anisotropy of the diffuse field upon which they are based? The familiar condition that the noise field be fully equipartitioned in order to recover Green's function is perhaps over restrictive for the purpose of estimating travel times.[2]

Snieder[2] showed that, in an asymptotic limit in which the two receivers are separated by a distance long compared to a wavelength, and when the distribution of diffuse ballistic intensity is smooth, albeit not necessarily isotropic, the correlation is the Green's function; one need not have a fully equipartitioned isotropic noise field. There were in that discussion, however, no indications as to the errors that might follow from a nonsmooth intensity distribution, or a less than fully asymptotic receiver separation. Recently Malargia and Castellaro[35] argued that the apparent robustness is due to the probability density function for the direction of an incident plane wave corresponding to a probability density function for apparent seismic velocity that is strongly peaked at the actual seismic velocity. In Sec. II, we readdress the derivation of the relation between G and C, but for a non-isotropic distribution of incident incoherent intensity. Attention is confined to the two dimensional case, as the case of three dimensions is both analytically simpler and of less practical importance. Sections III–VIII analyze the effect of that nonisotropy on practical estimates of travel time at finite receiver separations. It is shown that a travel time estimate is corrupted only slightly at realistic values for these parameters. An expression for the travel time correction is derived.

II. FIELD CORRELATIONS IN A NONISOTROPIC DISTRIBUTION OF DIFFUSE PLANE WAVE INTENSITY

Consider two receivers as in Fig. 1: one at origin, and the other a distance x from the origin. We distribute incoherent impulsive sources $s(\theta)$ over an annular region of (large) radius R around the receivers. With $\langle s\rangle=0$, and $\langle s(\theta)s^*(\theta')\rangle = B(\theta)\delta(\theta-\theta')$.

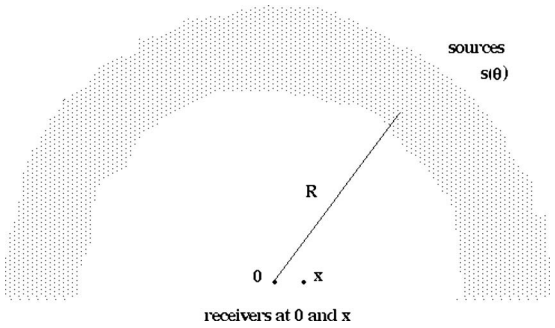


FIG. 1. Incoherent sources distributed at large distance from a pair of receivers.

These give rise to a field $\psi(\mathbf{r},t)$ whose Fourier transform

$$\tilde{\psi}(\mathbf{r},\omega) = \int \psi(\mathbf{r},t)\exp(-i\omega t)dt$$

is a superposition of cylindrical waves from the many sources s.

$$\tilde{\psi}(x,\omega) = \int s(\theta)d\theta \exp(i\omega R - i\omega x\cos\theta)/\sqrt{Ri(\omega-i\varepsilon)}. \tag{1}$$

The angle θ is defined relative to the strike line connecting the receivers. Here and below integrals over θ are over a full 2π and integrals over t and ω are from $-\infty$ to ∞, unless otherwise noted. The usual $-i\varepsilon$ has been inserted to analytically continue ω to its complex plane; it resolves ambiguities at real ω, and the choice of sign assures causality.[36] The cylindrical waves have been written in a form valid for asymptotically large ωR. Thus the present derivation is restricted to noise whose sources are in the far field. The receivers at 0 and x are not required to be in each other's far fields; indeed, attention is chiefly directed to the case in which they are separated by distances of order one wavelength. Units are used such that wave speed is unity.

The field has correlation defined by

$$C(\mathbf{r},\mathbf{r}',\tau) = \int \psi(\mathbf{r}',t)\psi(\mathbf{r},t+\tau)dt. \tag{2}$$

By the Wieiner–Khinchin cross-correlation theorem, C is the inverse Fourier transform of $\langle\tilde{\psi}^*\tilde{\psi}\rangle$

$$C(\mathbf{r},\mathbf{r}',\tau) = \frac{1}{2\pi}\int d\omega\langle\tilde{\psi}(\mathbf{r}',\omega)^*\tilde{\psi}(\mathbf{r},\omega)\rangle\exp(i\omega\tau). \tag{3}$$

The time derivative of the correlation between receivers at $\mathbf{r}=0$ and $\mathbf{r}'=\mathbf{i}x$ is

$$C'_{0,x}(\tau) \equiv \partial_\tau \int \psi(0,t)\psi(x,t+\tau)dt,$$

$$\tilde{C}'_{0,x}(\omega) \equiv i\omega \int C_{0,x}(\tau)\exp(-i\omega\tau)d\tau$$

$$= i\omega\langle \tilde{\psi}(0,\omega)^* \tilde{\psi}(x,\omega)\rangle \tag{4}$$

The relevant expectation is

$$\langle \tilde{\psi}(0)^* \tilde{\psi}(x)\rangle = \left\langle \int s(\theta')d\theta' \times \int s(\theta)d\theta \exp(-i\omega x\cos\theta)/R\sqrt{\omega^2+\varepsilon^2} \right\rangle$$

$$= \int B(\theta)d\theta \exp(-i\omega x\cos\theta)/R\sqrt{\omega^2+\varepsilon^2}. \tag{5}$$

We now return to the time domain, first multiplying by $i\omega$ (to impose the τ derivative needed for equivalence to G) and also inserting a rescaling factor $-R/4\pi$ for simplification of the final expressions.

$$C'_{0,x}(\tau) = \frac{-1}{4\pi}\int B(\theta)d\theta i\,\mathrm{sgn}(\omega) \times \exp(-i\omega x\cos\theta)\exp(i\omega\tau)d\omega. \tag{6}$$

The factor $\mathrm{sign}(\omega)$ arises from $\lim_{\varepsilon\to 0+} \omega/\sqrt{\varepsilon^2+\omega^2}$. $B(\theta)$ is now written in a Fourier series (cosines only, by symmetry, as the receiver correlation does not distinguish between positive and negative θ)

$$B(\theta) = B_0 + B_1\cos\theta + B_2\cos 2\theta + \cdots. \tag{7}$$

The odd harmonics could be removed by choosing to consider only a lapse time-symmetrized version of C.

The integration over θ is found in Abramowitz and Stegun 9.1.21 of Ref. 37. Thus

$$C'_{0,x}(\tau) = -\frac{1}{2}\sum_q (-i)^q B_q \int_{-\infty}^{\infty} i\,\mathrm{sgn}(\omega)\exp(i\omega\tau)J_q(\omega x)d\omega. \tag{8}$$

The evenness or oddness of J corresponds to that of q and so one may write

$$\int i\,\mathrm{sgn}(\omega)\exp(i\omega\tau)J_q(\omega x)d\omega = \begin{cases} 2i\int_0^\infty J_q(\omega x)\cos(\omega\tau)d\omega & \text{for } q \text{ odd} \\ -2\int_0^\infty J_q(\omega x)\sin(\omega\tau)d\omega & \text{for } q \text{ even.} \end{cases} \tag{9}$$

Thus,

$$C'_{0,x}(\tau) = \sum_{\text{even } q} (-1)^{q/2} B_q \int_0^\infty \sin(\omega\tau)J_q(\omega x)d\omega + \sum_{\text{odd } q} (-1)^{(q+1)/2} B_q \int_0^\infty \cos(\omega\tau)J_q(\omega x)d\omega. \tag{10}$$

These integrals are found in Abramowitz and Stegun 11.4.37 and 38 of Ref. 37. We take $x \geq 0$ without loss of generality, and $\tau \geq 0$ by recognizing that the expressions below can be evaluated for negative τ by replacing it with $|\tau|$ and changing the signs of the terms B_q for which q is even.

The expressions differ depending on whether or not $x > \tau$.

$$C'_{0,x}(\tau) = [-B_1/x - 2B_2\tau/x^2 + B_3(1-4\tau^2/x^2)/x + \cdots], \quad x > \tau$$

$$= \frac{1}{\sqrt{\tau^2-x^2}}[B_0 + xB_1/(\tau+\tau\sqrt{\tau^2-x^2}) + x^2B_2/(\tau+\sqrt{\tau^2-x^2})^2 + x^3B_3/(\tau+\sqrt{\tau^2-x^2})^3 + \cdots], \quad x < \tau. \tag{11}$$

In the special case of isotropy, where $B_q=0$ except for $q=0$, one recovers the well known result that C' is equal to the time-symmetrized Green's function, where G is

$$G = \begin{cases} 0 & \text{for } |x| > |\tau| \\ \mathrm{sgn}(\tau)/\sqrt{\tau^2-x^2} & \text{for } |x| < |\tau|. \end{cases} \tag{12}$$

These expressions have been evaluated numerically for a few choices $\{B_q\}$. Two of these are plotted in Fig. 2. A few points are evident: Nonisotropy leads to C' having support at times before the arrival of the Green's function; the waveform includes not just the trivial anticausal part at negative τ, but also a noncausal part at $|\tau|<|x|$. One also observes that every plot has a singularity at the arrival time $\tau=\pm x$. In most cases the singularity is of the form $1/\sqrt{(\tau^2-x^2)}$, but in some cases, notably Fig. 1(d) where $B(\pi)$ and $B(0)$ are close to zero, the singularities at the arrival times are less severe. Nevertheless, identification of arrival time is not difficult in any of these waveforms; lack of isotropy does not degrade estimation of wavespeed in these broad-band signals.

III. BAND-LIMITED CORRELATIONS

In practice, correlation waveforms have finite bandwidth. To address practical questions, we must therefore convolve the above waveforms with the time domain version of the square of the spectrum, a necessarily even function of time. In this case, Eq.(6) is modified:

$$C'_{0,x}(\tau) = \frac{-1}{4\pi}\int B(\theta)d\theta i\,\mathrm{sgn}(\omega) \times \exp(-i\omega x\cos\theta)\exp(i\omega\tau)|\tilde{a}(\omega)|^2 d\omega \tag{13}$$

If B is expanded in a Fourier series, Eq. (7), one has

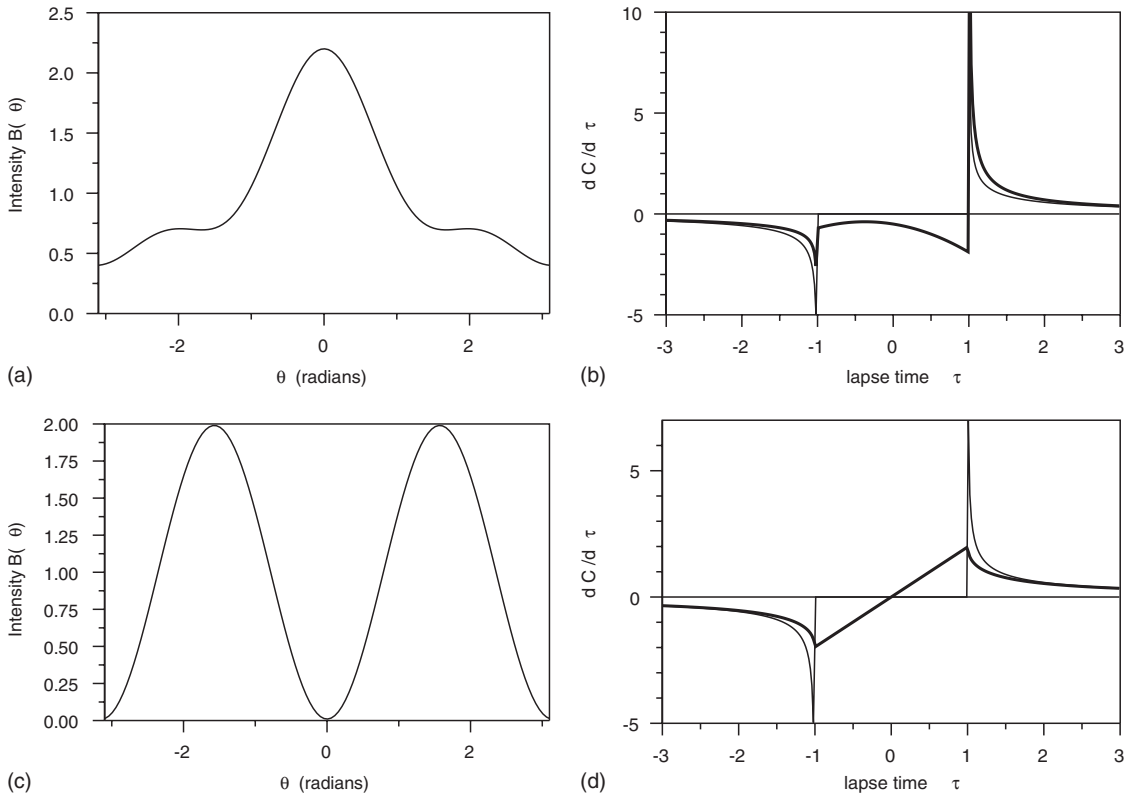


FIG. 2. Two examples of non-isotropic $B(\theta)$ [(a) and (c)] and the corresponding broadband field correlation [(b) and (d)] for distance $x=1$. Light line is Green's function, and the bold line is the correlation. The two cases correspond, respectively, to $\{B_q\}=\{1,0.7,0.3,0.2,0,0,0,\ldots\}$ in (a) and (b), and $\{1,0,-0.99,0,0,\ldots\}$ in (c) and (d).

$$C'_{0,x}(\tau) = -\frac{1}{2}\sum_q (-i)^q B_q \int_{-\infty}^{\infty} i\,\mathrm{sgn}(\omega)\exp(i\omega\tau)J_q(\omega x) \times |\tilde{a}(\omega)|^2 d\omega. \tag{14}$$

The new factor $|a|^2$ is the spectral power density of the noise, assumed here to have been independent of direction θ. It is necessarily even and real. It is useful to take it in the form of a (nonunique) sum of real contributions centered on positive and negative frequencies $\pm\omega_o$:

$$|\tilde{a}(\omega)|^2 = \tilde{A}(\omega-\omega_0) + \tilde{A}(\omega+\omega_0). \tag{15}$$

In particular, it can be useful to take these contributions to be Gaussian

$$\tilde{A}(\varpi) = \exp(-\varpi^2 T^2) \tag{16}$$

and to be real with chief support near $\varpi=0$. Its inverse Fourier transform is

$$\frac{1}{2\pi}\int |\tilde{a}(\omega)|^2 \exp(i\omega t)d\omega = \exp(i\omega_0 t)A(t) + \exp(-i\omega_0 t)A(-t), \tag{17}$$

where A is the inverse Fourier transform of $\tilde{A}$ and $A(-t)=A^*(t)$. If $\tilde{A}(\omega)$ is the Gaussian described above, then $A(t)$ is

$$A(t) = \frac{1}{2T\sqrt{\pi}}\exp(-t^2/4T^2). \tag{18}$$

If as before we take B in a Fourier series, we may confine our integration to positive ω if we assume $\tilde{A}(\omega-\omega_0)$ has no support at negative ω. Then

$$\int i\,\mathrm{sgn}(\omega)\exp(i\omega\tau)J_q(\omega x)|\tilde{a}(\omega)|^2 d\omega = \begin{cases} 2i\displaystyle\int_0^{\infty} J_q(\omega x)\cos(\omega\tau)\tilde{A}(\omega-\omega_0)d\omega & \text{for } q \text{ odd} \\ -2\displaystyle\int_0^{\infty} J_q(\omega x)\sin(\omega\tau)\tilde{A}(\omega-\omega_0)d\omega & \text{for } q \text{ even.} \end{cases} \tag{19}$$

So that,

$$C'_{0,x}(\tau) = \sum_{q\ \text{even}} (-1)^{q/2} B_q \int_0^{\infty} \sin(\omega\tau)J_q(\omega x)\tilde{A}(\omega-\omega_0)d\omega + \sum_{q\ \text{odd}} (-1)^{(q+1)/2} B_q \int_0^{\infty} \cos(\omega\tau)J_q(\omega x)\tilde{A}(\omega-\omega_0)d\omega. \tag{20}$$

IV. BAND-LIMITED HIGH FREQUENCY CORRELATIONS WAVEFORM NEAR NOMINAL ARRIVAL TIME

At asymptotically large ωx, Eq. (20) may be simplified by recalling

$$J_q(\omega x) \sim \sqrt{2/\pi\omega x}\cos\{\omega x - \pi q/2 - \pi/4\}.$$

With the definition $\phi_q = \pi q/2 + \pi/4$, we find

$$C'_{0,x}(\tau) = \frac{1}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ even}} (-1)^{q/2} B_q \int_0^\infty [\sin(\omega\tau + \omega x - \phi_q) + \sin(\omega\tau - \omega x + \phi_q)]\tilde{A}(\omega - \omega_0)d\omega + \frac{1}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ odd}} (-1)^{(q+1)/2} B_q \int_0^\infty [\cos(\omega\tau + \omega x - \phi_q) + \cos(\omega\tau - \omega x + \phi_q)]\tilde{A}(\omega - \omega_0)d\omega. \quad (21)$$

We again confine attention, without loss of generality, to τ and x both positive, and recognize that the first terms oscillate rapidly and contribute negligibly to the integration. Then

$$C'_{0,x}(\tau) = \frac{1}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ even}} (-1)^{q/2} B_q \operatorname{Im}\{\exp(i\omega_0(\tau - x) + i\phi_q) \times A(\tau - x)\} + \frac{1}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ odd}} (-1)^{(q+1)/2} B_q \operatorname{Re} \times \{\exp(i\omega_0(\tau - x) + i\phi_q)A(\tau - x)\}. \quad (22)$$

By choosing ω_0 such that $\tilde{A}(\varpi)$ is maximum at $\varpi = 0$, $A(t)$ can be approximated as real; its leading order imaginary part would be only of order time cubed (units of time being inverse bandwidth, T).

Thus in the vicinity of the arrival at $\tau = x$, and for $z = \tau - x$, we write

$$C'_{0,x}(\tau) = \frac{A(z)}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ even}} (-1)^{q/2} B_q \sin(\omega_0 z + \phi_q) + \frac{A(z)}{\sqrt{2\pi\omega_0 x}} \sum_{q \text{ odd}} (-1)^{(q+1)/2} B_q \cos(\omega_0 z + \phi_q). \quad (23)$$

This expression for the band-limited correlation waveform is asymptotically valid for large $\omega_0 x$. The ostensible arrival time might be determined by examining the location of the first peak, near $z = 0$. If only the B_0 term is present, then the peak is at $z = (\pi/2 - \phi_0)/\omega_0 = (\pi/4\omega_0)$. The peak is slightly delayed, by one-eighth of a cycle, beyond the true arrival time at $z = 0$. This observation could used to estimate arrival time from a band-limited Green's function. Alternatively, true arrival time could be estimated by picking the point that is one-eighth of a cycle *after* the zero that precedes the peak. The identification of a zero crossing is often used for high precision estimates of ultrasonic propagation time.[38]

If we have two or more terms B, then after inserting for the $\phi_q = \pi q/2 + \pi/4$, Eq. (23) becomes

$$C'_{0,x}(\tau) = \frac{A(z)}{\sqrt{2\pi\omega_0 x}}(B_0 + B_1 + B_2 + B_3 + \cdots) \times \sin(\omega_0 z + \phi_0). \quad (24)$$

The correlation is proportional to the full intensity on strike, $B(0) = B_0 + B_1 + B_2 + \cdots$. Asymptotically, it has the same temporal form that it has in the isotropic case. One concludes that the presence of an anisotropic diffuse field does not, at least in the asymptotic limit $\omega x \gg 1$, impair estimation of arrival time beyond the 1/8 cycle delay noted above that is present even with an isotropic field. This was Snieder's conclusion also.[2]

V. CORRECTIONS TO ZERO CROSSING TIME

It is of interest to inquire how, short of the full asymptotic limit $\omega x \to \infty$, apparent arrival time might be affected by non-isotropic $B(\theta)$. Without appealing to a Fourier decomposition of B, or to the Bessel function identities, we may write [Eq. (13)]

$$C'_{0,x}(\tau) = \frac{-1}{4\pi} \int B(\theta)d\theta i\,\operatorname{sgn}(\omega)\exp(-i\omega x\cos\theta) \times \exp(i\omega\tau)|\tilde{a}(\omega)|^2 d\omega.$$

Again we set $\tau = z + x$ and confine attention to positive ω

$$C'_{0,x}(\tau) = \frac{-i}{4\pi} \int B(\theta)d\theta \exp(i\omega x(1 - \cos\theta)) \times \exp(i\omega z)\tilde{A}(\omega - \omega_0)d\omega + \text{c.c.} \quad (25)$$

The asymptotically high ωx behavior near the arrival time is dominated by θ near 0:

$$C'_{0,x}(\tau) \sim \frac{-1}{4\pi} \int \left\{B(0) + \frac{1}{2}B''(0)\theta^2 + \cdots\right\} \times \left\{1 - \frac{1}{24}i\omega x\theta^4 + \cdots\right\}\exp(i\omega x\theta^2/2)d\theta \times \int_0^{+\infty} d\omega i\exp(i\omega z)\tilde{A}(\omega - \omega_0)d\omega + \text{c.c.}, \quad (26)$$

$$= \frac{-1}{4\pi}\int_0^{+\infty} \left\{B(0)\sqrt{\frac{2\pi}{-i\omega x}} + \frac{1}{4}B''(0)\sqrt{\frac{8\pi}{(-i\omega x)^3}} - \frac{i\omega x}{24}\left(\frac{3}{4}\right)B(0)\sqrt{\frac{32\pi}{(-i\omega x)^5}}\right\}d\omega i\exp(i\omega z) \times \tilde{A}(\omega - \omega_0) + \text{c.c.}, \quad (27)$$

or

$$= \frac{-1}{4\pi}\int_0^{+\infty} \left\{B(0)\sqrt{\frac{2\pi}{\omega x}}e^{i\pi/4} + \frac{1}{4}B''(0)\sqrt{\frac{8\pi}{(\omega x)^3}}e^{3i\pi/4} - \frac{i\omega x}{24}\left(\frac{3}{4}\right)B(0)\sqrt{\frac{32\pi}{(\omega x)^5}}e^{5i\pi/4}\right\} \times d\omega i\exp(i\omega z)\tilde{A}(\omega - \omega_0) + \text{c.c.} \quad (28)$$

There are two distinct methods by which one might in practice attempt to identify arrival time. It is not uncommon in ultrasonics to identify a zero crossing. It is more common in seismology, where signals tend to be more contaminated by noise, to cross correlate a waveform against a reference wavelet and select the time shift which maximizes the cross correlation. Here we shall investigate both methods.

At leading order, i.e., neglecting all but the first term, the correlation waveform is

$$B(0)A(z)\sqrt{\frac{2\pi}{\omega_0 x}}\sin(\omega_0 z+\pi/4), \tag{29}$$

which has a zero at $z=-\pi/4\omega_0$. This is the same waveform as in Eq. (24). We thus, as above, estimate "arrival time" ($z=0$) as one-eighth cycle later than this zero. To study corrections to the waveform near the zero, and therefore the shift of the zero, we analyze the expression (28) at $z=-\pi/4\omega_0$ and change integration variable: $\omega=\omega_0+\delta$, $d\omega=d\delta$, so $\exp(i\omega z)=\exp(-i\delta\pi/4\omega_0)\exp(-i\pi/4)=\exp(i\delta z)\times\exp(-i\pi/4)$.

Then, for $z=-\pi/4\omega_0$,

$$\begin{aligned}C'_{0,x^\sim}=&\frac{-i}{4\pi}\int_{-\infty}^{+\infty}d\delta\exp(i\delta z)\tilde{A}(\delta)\left\{B(0)\sqrt{\frac{2\pi}{\omega_0 x}}\left(1-\frac{\delta}{2\omega_0}\right)\right.\\&+\frac{1}{4}B''(0)\sqrt{\frac{8\pi}{(\omega_0 x)^3}}e^{i\pi/2}\left(1-\frac{3\delta}{2\omega_0}\right)\\&\left.-\frac{i\omega_0 x}{24}\left(\frac{3}{4}\right)B(0)\sqrt{\frac{32\pi}{(\omega_0 x)^5}}e^{i\pi}\left(1-\frac{3\delta}{2\omega_0}\right)+\cdots\right\}+\text{c.c.}\end{aligned} \tag{30}$$

The terms independent of δ integrate to $2\pi A(z)$. The terms linear in δ integrate to $2\pi i\partial_z A(z)$. Then, for $z=-\pi/4\omega_0$,

$$\begin{aligned}C'_{0,x}\sim&\frac{1}{2}\left\{B(0)\sqrt{\frac{2\pi}{\omega_0 x}}\left(-iA(z)+\frac{A'(z)}{2\omega_0}\right)\right.\\&+\frac{1}{4}B''(0)\sqrt{\frac{8\pi}{(\omega_0 x)^3}}\left(A(z)+\frac{3iA'(z)}{2\omega_0}\right)\\&\left.+\frac{\omega_0 x}{24}\left(\frac{3}{4}\right)B(0)\sqrt{\frac{32\pi}{(\omega_0 x)^5}}\left(A(z)+\frac{3iA'(z)}{2\omega_0}\right)\right\}\\&+\text{c.c.}\\=&\left\{B(0)\sqrt{\frac{2\pi}{\omega_0 x}}\left(\frac{A'(z)}{2\omega_0}\right)+\frac{1}{4}B''(0)\sqrt{\frac{8\pi}{(\omega_0 x)^3}}A(z)\right.\\&\left.+\frac{\omega_0 x}{24}\left(\frac{3}{4}\right)B(0)\sqrt{\frac{32\pi}{(\omega_0 x)^5}}A(z)\cdots\right\}\\=&\sqrt{\frac{2\pi}{\omega_0 x}}\left\{B(0)\frac{A'(z)}{2\omega_0}+\frac{1}{2\omega_0 x}B''(0)A(z)\right.\\&\left.+\frac{1}{8\omega_0 x}B(0)A(z)\cdots\right\}.\end{aligned} \tag{31}$$

Equation (31) represents the value of the correlation waveform C' at the nominal zero at $z=-\pi/4\omega_0$. The z-derivative of C' at the zero is [see Eq. (29)]

$$B(0)A(z)\sqrt{\frac{2\pi}{x\omega_0}}\omega_0.$$

Thus the waveform *near* its nominal zero at $z_0=-\pi/4\omega_0$ is

$$\begin{aligned}&B(0)A(z)\sqrt{\frac{2\pi}{x\omega_0}}\omega_0(z-z_0)+\sqrt{\frac{2\pi}{\omega_0 x}}\left\{\frac{1}{2\omega_0 x}B''(0)A(z)\right.\\&\left.+\frac{1}{8\omega_0 x}B(0)A(z)+B(0)\frac{A'(z)}{2\omega_0}\right\},\end{aligned} \tag{32}$$

which has its zero at

$$z=z_0-\left(4\frac{B''(0)}{B(0)}+1+4xA'(z)/A(z)\right)\Big/\,8\omega_0^2 x. \tag{33}$$

For the assumed $A(z)$, $4xA'/A=-2zx/T^2$ so arrival time, as evaluated by examining the time of this zero, is earlier than the true arrival time, by an amount

$$\left\{4\frac{B''(0)}{B(0)}+1+\pi(x/2\omega_0 T^2)\right\}\Big/\,8\omega_0^2 x. \tag{34}$$

Equation (33) serves as a higher order asymptotic estimate for the shift of the zero relative to its location as determined by Snieder[2] or by Eqs. (24) and (29).

The second and third terms in Eq. (34) are present even if B is constant, i.e., even if the correlation waveform is G itself. This is an indication that travel time assessment by identifying the time of the zero crossing and adding one-eighth of a cycle is only correct asymptotically; there are corrections at finite ωx and finite x/T.

The first term in Eq. (34) is the more interesting. It gives the leading order effect of nonisotropic diffuse intensity. By way of illustration, take T large, $B''(0)=-2B(0)$ [as would be the case if $B(\theta)=1+\cos(2\theta)$], and $\omega_0 x=6$ (one wavelength separation); then the zero occurs $7/48\omega_0$ later than one would have supposed, or about $1/40$ of a period. Velocity estimate would be erroneously low by about 2.5%.

VI. ARRIVAL TIME AS ESTIMATED BY CORRELATION WITH A REFERENCE WAVELET

Seismologists often evaluate arrival time by correlating the signal with a reference wavelet. Thus it is of interest also to cross correlate the waveform (13) with the same waveform obtained for the case $B=$const (i.e., with the band-limited Green's function itself). We again write Eq. (13) as

$$\begin{aligned}C'_{0,x}(\tau)=&\frac{-1}{4\pi}\int B(\theta)d\theta i\,\text{sgn}(\omega)\exp(-i\omega x\cos\theta)\\&\times\exp(i\omega\tau)|\tilde{a}(\omega)|^2 d\omega,\end{aligned}$$

whose Fourier transform is (an unimportant factor of −2 has been dropped)

$$\begin{aligned}\tilde{C}'_{0,x}(\omega)\sim&\int B(\theta)d\theta i\,\text{sgn}(\omega)\exp(-i\omega x\cos\theta)|\tilde{a}(\omega)|^2\\=&\int\left\{B(0)+\frac{1}{2}B''(0)\theta^2\right\}d\theta i\,\text{sgn}(\omega)\\&\times\exp\left(i\omega x\frac{1}{2}\theta^2\right)(1-i\omega x\theta^4/24)\exp(-i\omega x)|\tilde{a}(\omega)|^2\\=&\int\left\{B(0)+\frac{1}{2}B''(0)\theta^2-B(0)i\omega x\theta^4/24\right\}d\theta i\end{aligned}$$

$$\times \mathrm{sgn}(\omega)\exp\left(i\omega x\frac{1}{2}\theta^2\right)\exp(-i\omega x)|\tilde{a}(\omega)|^2$$
$$= \left\{ B(0)\sqrt{\frac{2\pi}{i\omega x}} + \frac{1}{2}B''(0)\sqrt{\frac{8\pi}{(i\omega x)^3}} - B(0)\frac{i\omega x}{24}\frac{3}{4}\sqrt{\frac{32\pi}{(i\omega x)^5}} \right\} i\,\mathrm{sgn}(\omega)\exp(-i\omega x)|\tilde{a}(\omega)|^2. \quad (35)$$

By expanding cos θ near $\theta=0$ we have implicitly focused on the arrival at positive lapse time. We wish to form the cross correlation between this and its version with $B=1$. At an offset of Δ, this is

$$X(\Delta) = \int \tilde{C}_{B\neq 1}(\omega)\tilde{C}^*_{B=1}(\omega)\exp(-i\Delta\omega)d\omega$$
$$= \int \left\{ \sqrt{\frac{2\pi}{i\omega x}} + \frac{B''(0)}{2B(0)}\sqrt{\frac{8\pi}{(i\omega x)^3}} - \frac{i\omega x}{24}\frac{3}{4}\sqrt{\frac{32\pi}{(i\omega x)^5}} \right\}$$
$$\times \left\{ \sqrt{\frac{2\pi}{-i\omega x}} + \frac{i\omega x}{24}\frac{3}{4}\sqrt{\frac{32\pi}{(-i\omega x)^5}} \right\}\exp(-i\omega\Delta)$$
$$\times |\tilde{a}(\omega)|^4 d\omega$$
$$= \int \frac{2\pi}{|\omega|x}\left\{ 1 + \frac{B''(0)}{2B(0)}\frac{1}{(i\omega x)} - \frac{1}{(8i\omega x)} \right\}$$
$$\times \left\{ 1 + \frac{1}{(8i\omega x)} \right\}\exp(-i\omega\Delta)|\tilde{a}(\omega)|^4 d\omega$$
$$= \int \frac{2\pi}{|\omega|x}\left\{ 1 + \frac{B''(0)}{2B(0)}\frac{1}{(i\omega x)} + \cdots \right\}$$
$$\times \exp(-i\omega\Delta)|\tilde{a}(\omega)|^4 d\omega. \quad (36)$$

We expand this for small Δ and find that it achieves its maximum, $\partial X/\partial\Delta=0$, at

$$\Delta = \frac{B''(0)}{2xB(0)}\int \frac{1}{|\omega|}|\tilde{a}(\omega)|^4 d\omega \Big/ \int |\omega||\tilde{a}(\omega)|^4 d\omega$$
$$\sim \frac{B''(0)}{2x\omega_0^2 B(0)}, \quad (37)$$

which is identical to the expression derived above (34), after its terms unrelated to anisotropy are removed.

Equations (37) and (34) provide an asymptotically valid estimate, $B''(0)/2x^2\omega_0^2 B(0)$, for the apparent fractional increase in wavespeed occasioned by having constructed a correlation waveform from smooth but not isotropic diffuse intensity. In Sec. VII, this estimate is compared to results from numerical simulations.

VII. COMPARISON WITH SYNTHETIC EXPERIMENTS

We compare the above predictions with numerical experiments, some purely synthetic and based on an assumed homogeneous medium, as pictured in Fig. 1, and others based on field data from a many component array of seismograms in Oman.

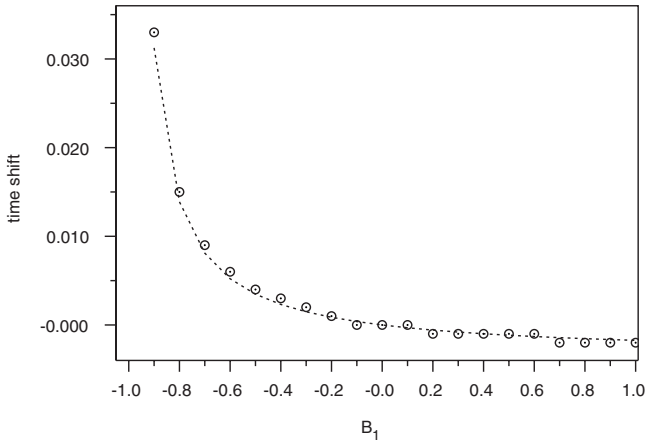


FIG. 3. Comparison of predicted time shift (dashed line) and time shift as obtained by correlation against the actual Green's function (symbols), for a set of intensity distributions given by $B(\theta)=1+B_1\cos(\theta)$ and a range of values for B_1.

For purposes of the tests with synthetic data, Eq. (11) for a distance $x=1$ was convolved with a tone burst of the form given in Eqs. (15)–(18) with $T=1/4$ and $\omega_0=12$. Thus we consider the case in which the receiver pair is separated by almost two wavelengths. This was done for a variety of choices $B(\theta)$. Each such waveform was isolated into its positive lapse time part ($\tau>0$) and cross correlated with the actual band-limited Green's function obtained by taking $B_0=1$, $B_{q\neq 1}=0$. The relative shift between these two was taken as the error in apparent arrival time as predicted in Sec. VI. As seen in Figs. 3–8, the theoretical expression (37) does a good job of predicting the time shift. This suggests that Eq. (37) may (i) be used in practice to estimate the error in apparent velocity, (ii) be used to correct for a non-isotropic distribution, or (iii) reassure a practitioner that such anisotropy does not significantly impact estimations of seismic velocity.

Figure 3 compares time shifts, for the arrival at positive lapse time, for an intensity distribution given by $B(\theta)=1+B_1\cos\theta$, for a range of B_1 values between -1 and $+1$. Except for the singular case $B_1=-1$ where the intensity on

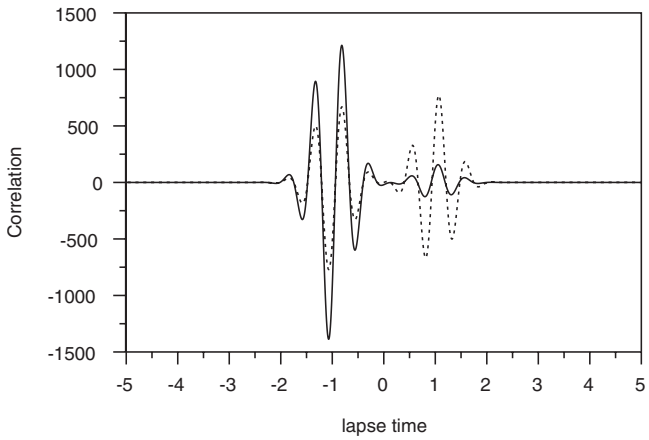


FIG. 4. Band-limited waveform (solid line) obtained from Eq. (13) using an intensity distribution $B(\theta)=1-0.8\cos\theta$. Dashed line is the time-symmetrized Green's function for the same spectrum, as obtained from Eq. (13) using $B(\theta)=1$.

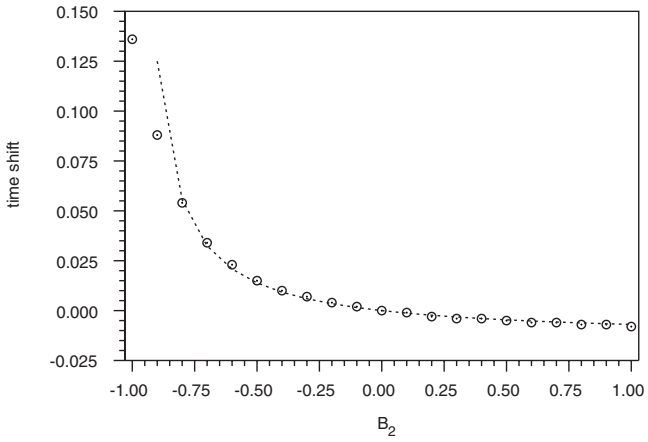


FIG. 5. Time shifts for $B(\theta)=1+B_2\cos 2\theta$, as a function of B_2.

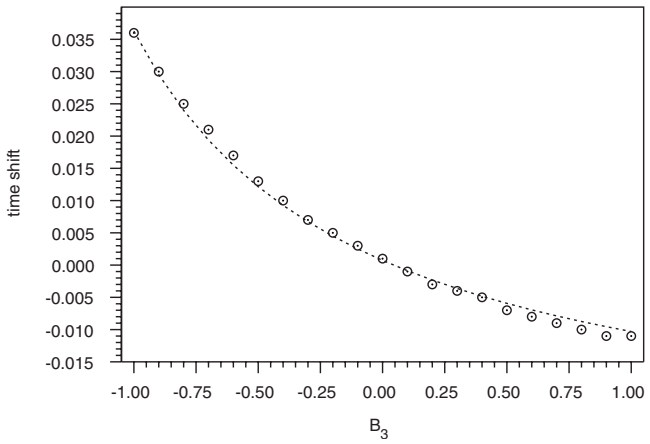


FIG. 7. Time shifts for $B(\theta)=1.7+0.4\cos\theta-0.2\cos 2\theta+B_3\cos 3\theta$.

strike $B(0)$ vanishes and Eq. (37) predicts an infinite correction, theory does an excellent job. Figure 4 shows an example of the waveforms used for this comparison, for the case $B_1=-0.8$. Figures 5–8 are similar to Figs. 3 and 4, but for more strongly varying intensity distributions.

VIII. COMPARISON USING DATA FROM FIELD MEASUREMENTS

A data set of 2 560 000 seismic responses courtesy of Petroleum Development Oman has been discussed elsewhere.[39,40] These responses were obtained from 1600 geophones in a square array of 25 m spacings, as due to 1600 active sources in a similar square array offset by 12.5 m. Here we use this data set to construct correlation waveforms from arbitrary distributions of sources. Two receivers $(\alpha=1,2)$ in the center of the array were selected. They were separated by a distance of 155 m, corresponding to a surface wave transit time of about 0.13 s at 15 Hz. Signals were studied from sources j in an annulus centered on the receivers, an annulus of inner radius 300 m, and thickness 69 m. Waveforms $\psi_{\alpha j}$ from the data set were windowed into the range 0–1.5 s so as to minimize contributions from scattered waves and emphasize ballistic waves, and correlated. The result was then summed with an angular weighting $B(\theta)$. The resulting $C_{12}(\tau)=\Sigma_j B(\theta_j)\int\psi_{1j}(t)\psi_{2j}(t+\tau)dt$ was then averaged over a set of 18 receiver pairs with different absolute orientations. This gave a correlation waveform for each of several choices $B(\theta)$.

Figure 9 compares the theoretical prediction to the time shift (divided by the nominal arrival time) as obtained by cross correlating the arrivals as constructed with $B=$const, and B as indicated in the captions. Theoretical predictions were based on power spectra having their support between 10 and 20 Hz. Arrival time, at 15 Hz, was about 0.13 s. Thus we took $\omega_0=94$ rad/s and $\omega_0 x=12$. The spectrum was only approximately Gaussian, but it shared a second moment with a Gaussian of $T=0.04$ s. The plots examine the same three cases of weighting B, examined in Figs. 3, 5, and 7. The asymptotic theory continues to describe the apparent time shift. The theory is quantitatively less accurate. The difference may be ascribed to the presence of some nonballistic waves (there is some scattering), to not being in the far field of the original sources [their emissions are not plane waves in the vicinity of the receivers $R=\infty$, as assumed for Eq. (1)], and to geometric dispersion (the soil is layered; surface waves have a frequency dependent speed.)

IX. SUMMARY

Non-isotropic distributions of ballistic specific intensity violate the assumptions behind the identification of ambient

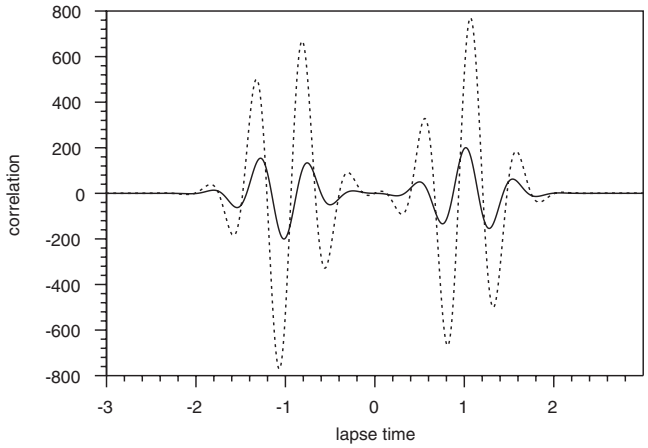


FIG. 6. Example of waveforms used in Fig. 5.

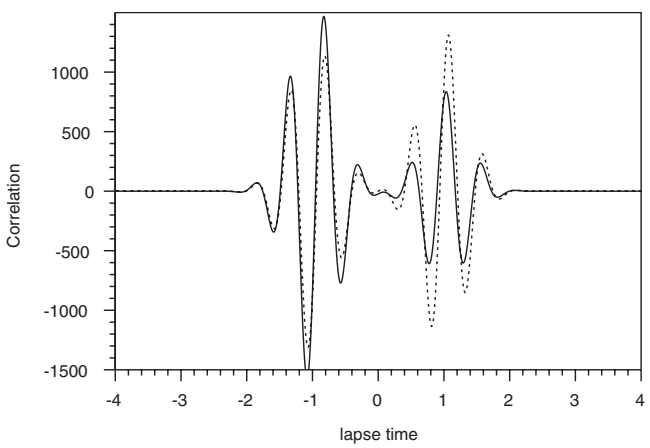


FIG. 8. Example of waveforms used in Fig. 7, the case $B_3=-1$.

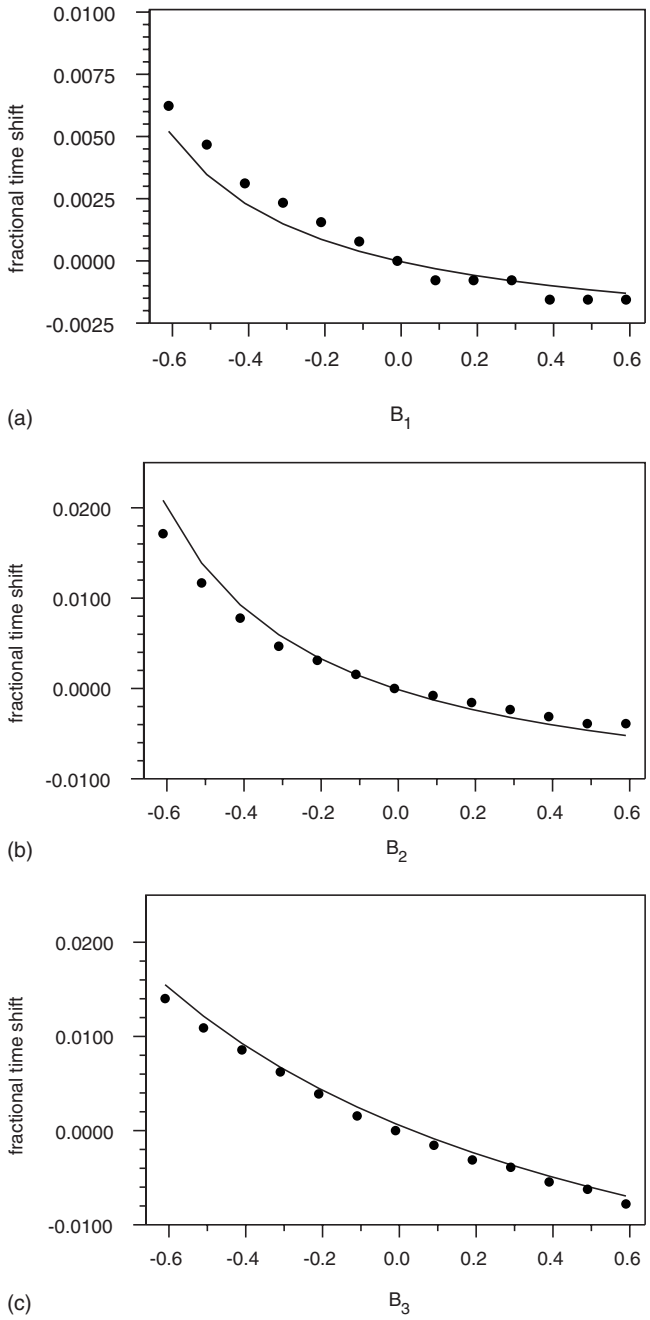


FIG. 9. (a) Comparison of predicted (solid line) and measured arrival time difference (as a fraction of the nominal arrival time 0.13 s) between cross correlations obtained by using $B=\text{const}$ and $B=1+B_1\cos\theta$ (symbols). Raw data taken from the Oman data set. (b) Comparison of predicted (solid line) and measured arrival time difference between cross correlations obtained by using $B=\text{const}$ and $B(\theta)=1+B_2\cos 2\theta$ (symbols). Raw data taken from the Oman data set. (c) Comparison of predicted (solid line) and measured arrival time difference between cross correlations obtained by using $B=\text{const}$ and $B(\theta)=1.7+0.4\cos\theta-0.2\cos 2\theta+B_3\cos 3\theta$ (symbols). Raw data taken from the Oman data set.

noise correlations and Green's functions. Here an asymptotically valid formula is derived that permits estimations of the consequent error in estimates of travel time based on such correlations. The formula is successfully compared with apparent travel times seen in synthetic waveforms. Although based on simple assumptions, the formula derived here was shown to be a good approximation when dealing with actual records of surface waves from an exploration experiment. The presence of different types of waves and of heterogeneities in that medium did not significantly degrade the accuracy of the theoretical prediction.

It is found that for sufficiently smooth distributions, and for sufficiently large receiver-receiver separations, the error in apparent arrival time is small, thus removing corresponding concern over possible inaccuracies in modern maps of seismic velocity based on arrival times seen in ambient seismic noise correlations.

ACKNOWLEDGMENTS

The authors thank the Ministry of Oil and Gas of the Sultanate of Oman, Petroleum Development of Oman, and Shell Research for permission to use the data. One of the authors (B.F.) acknowledges the support from Shell Research. R.L.W. thanks the *Laboratoire de Géophysique Interne et Tectonophysique* for support while he was in residence in Grenoble where much of this work was done.

[1]O. I. Lobkis and R. L. Weaver, "On the emergence of the Greens function in the correlations of a diffuse field," J. Acoust. Soc. Am. **110**, 3011–3017 (2001).
[2]R. Snieder, "Extracting the Greens' function from the correlation of coda waves: A derivation based on stationary phase," Phys. Rev. E **69**, 046610 (2004).
[3]R. L. Weaver and O. I. Lobkis, "Diffuse waves in open systems and the emergence of the Greens' function," J. Acoust. Soc. Am. **116**, 2731–2734 (2004).
[4]P. Roux, K. G. Sabra, W. A. Kuperman, and A. Roux, "Ambient noise cross correlation in free space: Theoretical approach," J. Acoust. Soc. Am. **117**, 79–84 (2004).
[5]R. L. Weaver and O. I. Lobkis, "Ultrasonics without a source; Thermal fluctuation correlations at MHz frequencies," Phys. Rev. Lett. **87**, 134301 (2001).
[6]A. Derode, E. Larose, M. Campillo, and M. Fink, "How to estimate the Green's function of a heterogeneous medium between two passive sensors? Application to acoustic waves," Appl. Phys. Lett. **83**, 3054–3056 (2003).
[7]R. L. Weaver and O. I. Lobkis, "The mean and variance of diffuse field correlations in finite bodies," J. Acoust. Soc. Am. **118**, 3447–3456 (2005).
[8]E. Larose, O. I. Lobkis, and R. L. Weaver, "Coherent backscattering of ultrasound without a source," Europhys. Lett. **76**, 422–428 (2006).
[9]E. Larose, O. I. Lobkis, and R. L. Weaver, "Passive correlation imaging of a buried scatterer," J. Acoust. Soc. Am. **119**, 3549–3552 (2006).
[10]A. E. Malcolm, J. A. Scales, and B. A. van Tiggelen, "Retrieving the Green function from Diffuse, equipartitioned waves," Phys. Rev. E **70**, 015601 (2004).
[11]N. M. Shapiro and M. Campillo, "Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise," Geophys. Res. Lett. **31**, L07614 (2004).
[12]S. E. Fried, W. A. Kuperman, K. G. Sabra, and P. Roux, "Extracting the local Green's function on a horizontal array from ambient ocean noise," J. Acoust. Soc. Am. **124**, EL183–EL188 (2008).
[13]P. Roux and W. A. Kuperman, "Extracting coherent wave fronts from acoustic ambient noise in the ocean," J. Acoust. Soc. Am. **116**, 1995–2003 (2004).
[14]L. Stehly, M. Campillo, B. Froment, and R. L. Weaver, "Reconstructing Green's function by correlation of the coda of the correlation (C^3) of ambient seismic noise," J. Geophys. Res. **113**, B11306 (2008).
[15]K. Wapenaar, D. Draganov, and J. Robertsson, "Seismic interferometry supplement," Geophysics **71**, SI1–SI102 (2006).
[16]P. Gouedard, L. Stehly, F. Brenguier, M. Campillo, Y. Colin de Verdiere, E. Larose, L. Margerin, P. Roux, F. J. Sanchez-Sesma, N. M. Shapiro, and R. L. Weaver, "Cross-correlation of random fields: mathematical approach and applications," Geophys. Prospect. **56**, 375–393 (2008).
[17]K. Wapenaar, "Nonreciprocal Green's function retrieval by cross correlation," J. Acoust. Soc. Am. **120**, EL7–EL13 (2006).
[18]O. A. Godin "Recovering the acoustic Green's function from ambient

noise cross correlation in an inhomogeneous moving medium," Phys. Rev. Lett. **97**, 054301 (2006).
[19]R. Snieder, "Extracting the Greens function of attenuating acoustic media from uncorrelated waves," J. Acoust. Soc. Am. **121**, 2637–2643 (2007).
[20]R. Snieder, "Retrieving the Greens function of the diffusion equation from the response to a random forcing," Phys. Rev. E **74**, 046620 (2006).
[21]R. L. Weaver "Ward identities and the retrieval of Greens functions in the correlations of a diffuse field," Wave Motion **45**, 596–604 (2008).
[22]K. G. Sabra, E. S. Winkel, D. A. Bourgoyne, B. R. Elbing, S. L. Ceccio, M. Perlin, and D. R. Dowling, "Using cross correlations of turbulent flow-induced ambient vibrations to estimate the structural impulse response. Application to structural health monitoring," J. Acoust. Soc. Am. **121**, 1987–1995 (2007).
[23]M. Rollwage, K. Ebeling, and D. Guicking, "Experimental investigation of two dimensional diffuse sound fields in a shallow water basin," Acustica **58**, 149–161 (1985).
[24]K. Aki, "Space and time spectra of stationary waves, with special reference to microtremours," Bull. Earthquake Res. Inst., Univ. Tokyo **35**, 415–456 (1957).
[25]H. Okada, "The microtremor survey method," Geophysical Monograph Series No. 12, Society of Exploration Geophysicists (2003).
[26]R. H. Lyon, "Statistical analysis of power injection and response in structures and rooms," J. Acoust. Soc. Am. **45**, 545–565 (1969).
[27]R. H. Lyon and R. G. DeJong, *Theory and Application of Statistical Energy Analysis*, 2nd ed. (Butterworth-Heinemann, Boston, 1995).
[28]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).
[29]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).
[30]M. Campillo and A. Paul, "Long-range correlations in the diffuse seismic coda," Science **299**, 547–549 (2003).
[31]R. L. Weaver and O. I. Lobkis, "Elastic wave thermal fluctuations, ultrasonic waveforms by correlation of thermal phonons," J. Acoust. Soc. Am. **113**, 2611–2621 (2003).
[32]O. Godin, "Retrieval of Green's functions of elastic waves from thermal fluctuations of fluid-solid systems," J. Acoust. Soc. Am. **125**, 1960–1970 (2009).
[33]F. Brenguier, M. Campillo, C. Hadziioannou, N. M. Shapiro, R. M. Nadeau, and E. Larose, "Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations," Science **321**, 1478–1481 (2008).
[34]F. Brenguier, N. M. Shapiro, M. Campillo, V. Ferrazzini, Z. Duputel, O. Coutant, and A. Nercessian, "Towards forecasting volcanic eruptions using seismic noise," Nature Geoscience **1**, 126–130 (2008).
[35]F. Mulargia and S. Castellaro, "Passive imaging in nondiffuse acoustic wavefields," Phys. Rev. Lett. **100**, 218501 (2008).
[36]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Sec. 4.8.
[37]*Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1972).
[38]H. F. Nicholson, G. Lowet, C. M. Langton, J. Dequeker, and G. Van der Perre, "A comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurement in trabecular bone," Phys. Med. Biol. **41**, 2421–2435 (1996).
[39]P. Gouédard, P. Roux, M. Campillo, and A. Verdel, "Convergence of the two-point correlation function toward the Green's function in the context of a seismic prospecting dataset," Geophysics **73**, V47–V53 (2008).
[40]G. C. Herman and C. Perkins, "Predictive removal of scattered noise," Geophysics **71**, V41–V49 (2006).

Estimating sound power radiated from rectangular baffled panels using a radiation factor

Dan Palumbo[a)]
NASA Langley Research Center, MS 463, Hampton, Virginia 23681

(Received 29 August 2008; revised 10 March 2009; accepted 16 July 2009)

A method is introduced which is shown to predict radiated sound power from rectangular baffled panels. The method employs a filtered wavenumber transform to extract the power in the supersonic wavenumbers on the panel and a radiation factor to scale the supersonic power to match the actual radiated sound power. Although empirically derived, the radiation factor is shown to be related to the radiation efficiency of an infinite panel. The radiation factor is simple, depending only on the ratio of the wavenumbers of the panel to the radiation medium, and the method is straightforward to use, requiring only the panel normal velocities. The computation is efficient, as much as two orders of magnitude faster than a Rayleigh integration, thus providing a means of combining sound power predictions with finite element optimizations. A formula is derived which predicts the lowest frequency for which the method is valid as a function of the bin width of the wavenumber transform. The radiation factor method is shown to produce radiated sound power estimates which favorably compare to estimates derived from intensity measurements of physical test specimens and to Rayleigh integral estimates computed using both simulated and measured velocities.
[DOI: 10.1121/1.3203930]

PACS number(s): 43.40.Rj [SFW] Pages: 1827–1837

NOMENCLATURE

c = Speed of sound in air
dk = Bin width of wavenumber transform
I = Intensity of acoustic radiation
i = $\sqrt{-1}$
k = Wavenumber
k_a = Wavenumber in air
k_p = Wavenumber in panel
L = Panel dimension
n_m = Mode order
n_p = Pad multiplier
n_r = Filtered transform accuracy factor
p = Pressure at observer
Rect = Rectangle function
r = Distance from source to observer
S = Panel surface area
sin = Sine function
sinc = Unnormalized sinc function
v = Velocity normal to surface of panel
V = Wavenumber transform of v
x_o, y_o, z_o = Observer coordinates
x_s, y_s = Source coordinates
δ = Kronecker delta
γ = Ratio of wavenumber in air to wavenumber in panel
φ = Angle of radiated plane wave to panel surface normal
Π = Radiated sound power
Π_s = Radiated sound power derived from panel supersonic wavenumbers
ρ = Density of air
σ = Radiation efficiency
σ_e = Empirical radiation factor
σ_o = Optimized radiation factor
σ_∞ = Infinite panel radiation efficiency, finite panel radiation factor
$\mathcal{F}$ = Fourier transform function

I. INTRODUCTION

The Rayleigh integral is the classical method used to compute the far-field sound power radiated from a baffled panel given the panel normal velocities. The integral returns excellent results, but can be computationally prohibitive especially in iterative optimization applications. Alternative approaches are founded in the seminal works of Maidanik[1] and Wallace[2] where the methodologies employ either select approximations to the radiation efficiency based on the problem domain (Maidanik) or analytically derived expressions for the modal radiation efficiency (Wallace). Many studies expand on these works (see Ref. 3 for a good overview) offering unique solution techniques for the specific problem addressed.

The work presented here is derived from Heckl's approach[4,5] where he explained that the finite baffled panel velocity field can be decomposed into a series of infinite fields (essentially a wavenumber transform) each able to be analyzed using the radiation efficiency for an infinite panel. This approach is shown to have some weakness, especially near the sonic wavenumber. Williams[6] provided a more formal rationale for the casting of the finite velocity field into a series of infinite fields and suggested a series expansion in the wavenumber domain for approximating the radiated sound power.[7] Perhaps the work which comes closest to that

[a)]Electronic mail: d.l.palumbo@nasa.gov

presented here is the concept of supersonic intensity developed by Williams.[8] Williams used near-field acoustic holography to extract the complex velocities and pressures on a plane. Using a wavenumber filter, the supersonic components of these quantities can be separated from the near-field components, and the intensity that radiates to the far field (the supersonic intensity) can be calculated.

The purpose of this work is to validate an alternative approach for computing the radiated sound power from panel velocities that is general enough and has the accuracy and computational efficiency necessary for use in iterative optimization. To this end, the panel normal velocities are transformed to the wavenumber domain where the supersonic components are extracted. A formula for a radiation factor is empirically derived for the case of a baffled panel radiating into free space and is used with the supersonic power in the panel to predict radiated sound power. The accuracy of the radiation factor is tested through comparisons to intensity measurements and Rayleigh integral predictions of radiated sound power and radiation efficiency. The computational overhead of the radiation factor method is compared to that of the Rayleigh integral. Sections II and III provide background theory and experimental results which include both numerical simulation and physical test.

II. THEORY

The Rayleigh integral of panel normal velocities will be used along with intensity measurements to validate the radiation factor estimate. There are several ways by which the radiated sound power can be computed using the Rayleigh integral. The most direct approach for this application would be to use the Rayleigh surface integral which yields pressures on the surface of the panel enabling direct computation of the radiated intensity from the pressure and velocity. The Rayleigh integral is evaluated using a simple rectangular numeric integral. This will be shown to introduce some error at higher modal orders and frequencies, but is the preferred approach as it is more compatible with the discrete physical data and is much faster than adaptive integration techniques. The radiation factor scales the total power in the supersonic wavenumbers on the panel to produce estimates of radiated sound power. The supersonic filtering of the wavenumber transformation requires some explanation of the Fourier transform and associated theorems as well as the process of padding the transform to increase resolution. The concept that the response of a finite panel can be considered to be the sum of the responses of a set of infinite panels is shown to be a result of the Fourier transform. The classic definitions of radiation efficiency for finite and infinite panels are given. The radiation factor is then shown to be related to the first order approximation for the radiation efficiency of an infinite panel. Sources of computation error for both the Rayleigh integral and radiation factor method are described in the Appendix.

A. Rayleigh surface integral

The Rayleigh integral is well known and written as

$$p(x_o,y_o,z_o) = \frac{-i\rho c k_a}{2\pi}\int_0^{L_y}\int_0^{L_x} v(x_s,y_s)\frac{e^{ik_a r}}{r}dx_s dy_s, \tag{1}$$

where p is the pressure at observer coordinates x_o, y_o, and z_o; ρ, c, and k_a are the density, wave speed, and wavenumber of sound in air; v is the panel normal velocity at source coordinates x_s and y_s; L_x and L_y are the panel dimensions; and r is the distance from the source to the observer.

The sound intensity radiated from the panel is computed as[9]

$$I(x_o,y_o) = \tfrac{1}{2}\,\mathrm{Re}(pv^*), \tag{2}$$

where * denotes the complex conjugate. The total radiated sound power is found by integrating Eq. (2) over a surface enclosing the panel which in this case is taken to be the surface of the panel, i.e., when $z_o=0$,

$$\Pi = \int_0^{L_y}\int_0^{L_x} I dx_s dy_s. \tag{3}$$

When computing Eq. (3) numerically, consideration must be given for the case when (x_o,y_o) approaches (x_s,y_s) and r goes to zero. When performing a numerical integration at this point with discrete square elements $(dx=dy)$ and the observer at the source, the propagator term in the integrand of Eq. (1) is converted to cylindrical coordinates and evaluated with the following approximation:

$$\int_0^{2\pi}\int_0^{dx}\frac{e^{ik_a r}}{r} r dr d\theta = 2\pi\left(\frac{e^{ik_a dx/2}-1}{ik_a}\right). \tag{4}$$

B. Wavenumber transform

A finite velocity field can be expressed as the product of an infinite velocity field and the rectangle function Rect located over the finite field.[6] The definition of Rect for a panel of unit length in one dimension is

$$\mathrm{Rect}(x) = \begin{cases} 1 & |x| < 1/2 \\ \frac{1}{2} & |x| = 1/2 \\ 0 & |x| > 1/2 \end{cases}. \tag{5}$$

The wavenumber transform in one dimension is defined as

$$\mathcal{F}(f(x)) = F(k) = \int_{-\infty}^{\infty} f(x)e^{-ikx}dx, \tag{6}$$

Equation (6) defines a continuous Fourier transform. All data analyzed in this paper were processed using a discrete Fourier transform where the results are accumulated in individual bins which have a center wavenumber k and width Δk=sample rate/N, where N is the number of data samples in each ensemble. For a time domain transform the resolution is easily increased by increasing N. This is not so easily accomplished in a wavenumber transform where the number of samples at a given sample rate is limited by the dimension of the article under test. To decrease a wavenumber transform's bin width, i.e., increase its resolution, the number of samples in the data set can be increased by adding zeros, or padding,

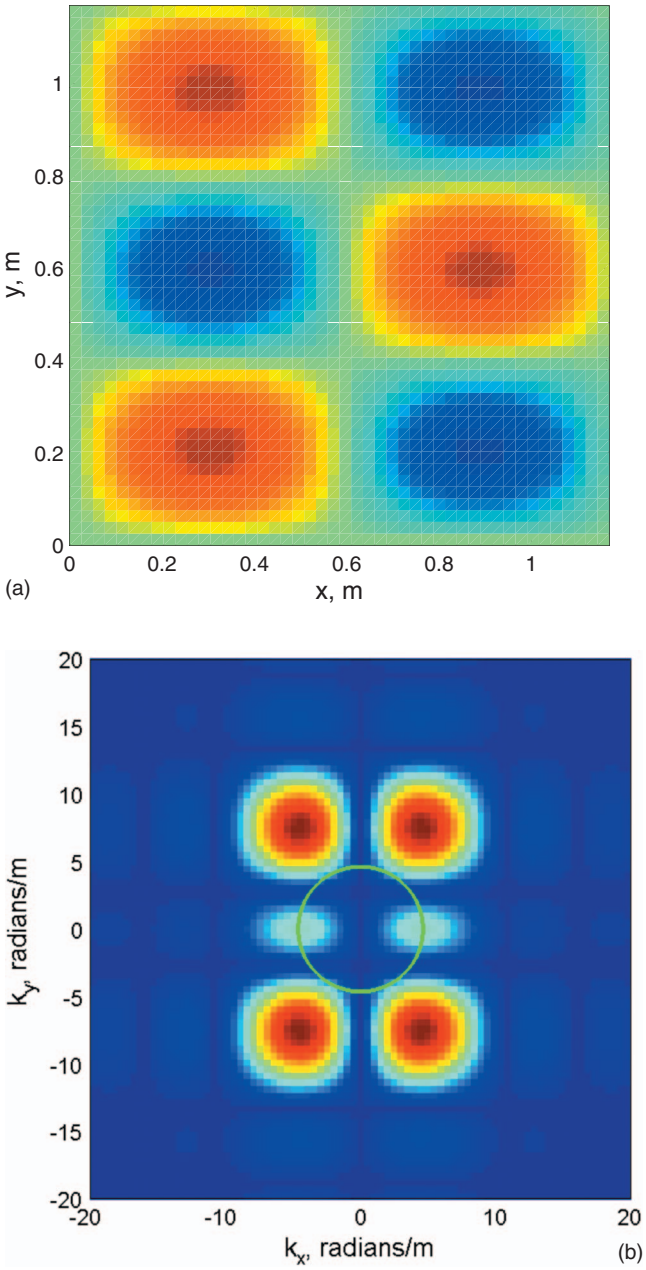


FIG. 1. Example (2,3) mode: (a) normal velocity field and (b) wavenumber transform with sonic circle in green.

to the data. Increased resolution gained by increased levels of padding comes at a cost of distortion of the result, as will be explained next, and an increase in computation requirement, as explained in Sec. II E.

The transform of the rectangle function is the sinc function,

$$\mathcal{F}(\mathrm{Rect}(x)) = \mathrm{sinc}(k/2), \tag{7}$$

where

$$\mathrm{sinc}(x) = \frac{\sin(x)}{x}. \tag{8}$$

With a finite velocity field v expressed as the product of an infinite velocity field v_∞ and the rectangle function, the

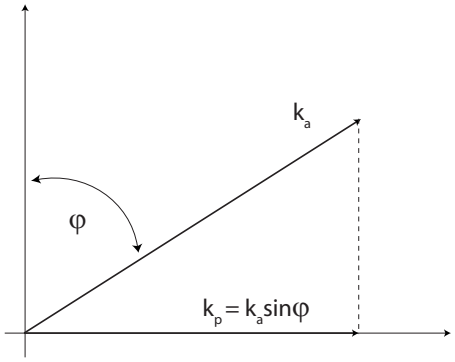


FIG. 2. Relation of plate wavenumber to that of air for a radiating, infinite panel.

Fourier transform of the finite velocity field becomes the convolution of the Fourier transform of the respective functions.

$$\mathcal{F}(v(x)) = \mathcal{F}(v_\infty(x)\mathrm{Rect}(x)) = \mathcal{F}(v_\infty(x)) \otimes \mathcal{F}(\mathrm{Rect}(x))$$
$$= \delta((k - k_x)) \otimes \mathrm{sinc}(k/2) = \mathrm{sinc}((k - k_x)/2), \tag{9}$$

where $\otimes$ is the convolution operator. Equation (9) then takes the familiar form of the sinc function centered on the velocity field's wavenumber and dispersing energy throughout the wavenumber domain. An example is shown in Fig. 1 where the transform of a (2,3) mode at 250 Hz on a simply supported panel, (a), generates the lobed pattern that is characteristic of the sinc function, (b). Each wavenumber in the transform represents an infinite wave in the spatial domain. The original finite response is the sum of the responses of the infinite waves. The supersonic, radiating energy in the mode is contained in the panel wavenumbers that are less than or equal to the sonic wavenumber, as delineated by the green circle in Fig. 1(b). As can be seen, most of the panel's energy is subsonic in this example.

C. Radiation efficiency

The radiation efficiency of a panel is defined as the panel's radiated power referenced to the power radiated by a piston with the same area and average mean square velocity,[4]

$$\sigma = \frac{\Pi}{\rho c S \langle \overline{v^2} \rangle}, \tag{10}$$

where σ is the radiation efficiency, Π the panel's total radiated power, ρ the density of air, c the speed of sound in air, S the area of the panel, and $\langle \overline{v^2} \rangle$ the area-averaged, mean square panel normal velocity.

For an infinite panel whose bending wavenumber is at or below the wavenumber in air, a plane wave will radiate from the panel at an angle φ to the panel normal,[10] Fig. 2, where k_a is the wavenumber in air and k_p the wavenumber in the panel. When the panel radiates at wavenumber k_a, the panel's wavenumber k_p is equal to the component of k_a parallel to k_p.

The radiation efficiency for an infinite panel is[1,4]

$$\sigma_\infty(k_a,k_p) = \frac{k_a}{\sqrt{k_a^2 - k_p^2}} \quad \text{for } k_p \le k_a. \tag{11}$$

Defining a wavenumber ratio $\gamma = k_a/k_p$ yields

$$\sigma_\infty(\gamma) = \frac{1}{\sqrt{1-\gamma^{-2}}} \quad \text{for } \gamma \ge 1. \tag{12}$$

Note that at any single frequency ω, a single acoustic wavenumber k_a will exist, but the plate will create many structural wavenumbers k_p and thus a range of γ.

D. The radiation factor

Given the argument presented in Sec. II B, it might seem reasonable, as suggested by Heckl,[4,5] to consider the finite panel as a series of infinite panels each with energy at a specific wavenumber. The radiated sound power can then be found by using σ_∞ with the average power in each of the "infinite" panels that have supersonic wavenumbers, i.e., those within the sonic circle. A filtered wavenumber transform can be used to extract the supersonic components of the panel velocities and compute the average supersonic power at that wavenumber,

$$\bar{V}_s^2(\gamma) = \sum_{k_p} V(\gamma)V(\gamma)^*/N \quad (\gamma \ge 1), \tag{13}$$

where N is the number of elements in the entire discrete wavenumber transform and $V(\gamma)$ is the wavenumber transform of the panel normal velocity v at a given frequency ω. Using Eq. (10), the power radiated at that wavenumber can then be written as

$$\Pi_s(\gamma) = \rho c S \sigma_\infty(\gamma)\bar{V}_s^2(\gamma). \tag{14}$$

It is important to note that σ_∞ is not the radiation efficiency of the panel but the radiation efficiency of the supersonic wavenumbers in a similar, but infinite, panel. To avoid confusing σ_∞ with the panel's actual radiation efficiency, we will refer to σ_∞ and its derivatives as the panel's radiation factor. Ignoring, for the moment, the singularity in σ_∞ at $\gamma = 1$, the total power radiated by the supersonic wavenumbers is

$$\Pi_s = \rho c S \sum_{\gamma} \sigma_\infty(\gamma)\bar{V}_s^2(\gamma) \quad (\gamma \ge 1). \tag{15}$$

The panel's radiation efficiency at a particular frequency ω can now be computed using Π_s for total radiated power Π in Eq. (10) or

$$\sigma(\omega) = \frac{\Pi_s}{\rho c S \langle \overline{v^2} \rangle}. \tag{16}$$

The radiation efficiency for the (2,3) mode shown in Fig. 1 is computed using radiated sound power estimates from both the Rayleigh integral and Eq. (16) with the results shown by the solid and dashed lines in Fig. 3. As is customary, the radiation efficiency is plotted against the wavenumber ratio at the modal frequency $\gamma(\omega_m)$.

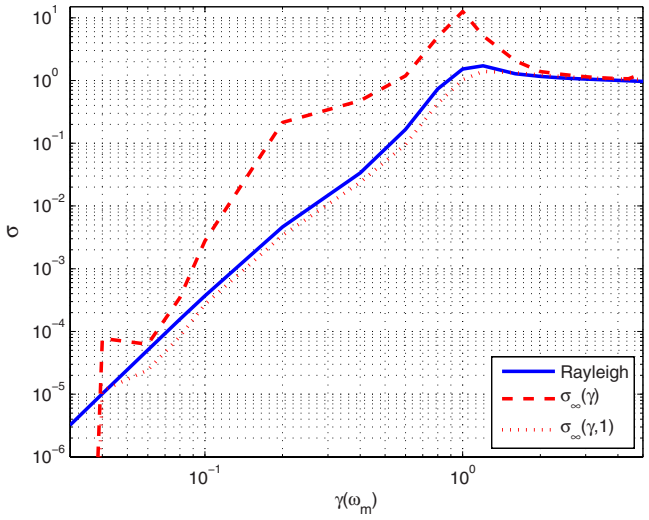


FIG. 3. (Color online) Radiation efficiency of (2,3) mode, Rayleigh, and infinite panel approximations.

$$\gamma(\omega_m) = \frac{\omega_m L}{c n_m \pi}, \tag{17}$$

where L is the panel dimension and n_m the mode order. The infinite panel assumption overestimates the radiation efficiency by a wide margin due to the singularity in the radiation factor σ_∞. The singularity comes into play at any value of $\gamma(\omega_m)$ because a wide range of wavenumbers exist on a finite panel at any one frequency so that the likelihood of a wavenumber with appreciable power near the sonic circle is high. The question remains, then, if an alternative form for the radiation factor, other than Eq. (12), can be found which can produce a better estimate.

Equation (12) can be rewritten as

$$\sigma_\infty(\gamma) = \frac{1}{\sqrt{1-\gamma^{-2}}} = (1 + \gamma^{-2} + \gamma^{-4} + \gamma^{-6} + \cdots)^{1/2}, \tag{18}$$

and can be approximated as

$$\sigma_\infty(\gamma,n) = (1 + \gamma^{-2} + \gamma^{-4} + \cdots + \gamma^{-2n})^{1/2}. \tag{19}$$

The first order approximation to σ_∞ is then

$$\sigma_\infty(\gamma,1) = (1 + \gamma^{-2})^{1/2}. \tag{20}$$

The result of using the first order approximation to σ_∞, $\sigma_\infty(\gamma,1)$, to compute the panel radiation efficiency is plotted as the dotted line in Fig. 3. It is shown to produce a better estimate of the panel radiation efficiency than σ_∞. The error can be reduced by adding more terms, but, as more terms are added to make this particular example a closer match, a different example with modes closer to the supersonic circle would overestimate the result because of the increased power in the vicinity of the supersonic circle. This sensitivity depends on the distribution of wavenumber power near the sonic circle and is problem dependent. If one case has little power adjacent to the sonic circle, it may require (and tolerate) higher order terms, yet, another case with more power at the sonic circle will overestimate the radiated power due to

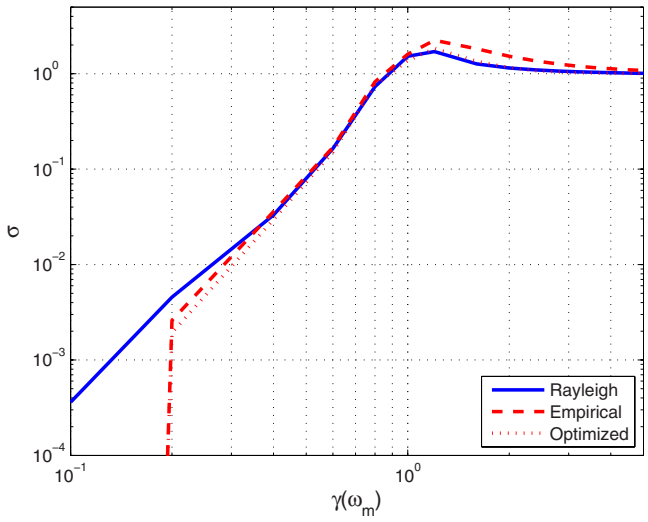


FIG. 4. (Color online) Radiation efficiency of (2,3) mode, Rayleigh, empirical, and optimized approximations.

the singularity. The same trade-offs exist if one tries to approach the sonic circle to within some limit to avoid the singularity.

In an effort to find a more stable solution, trial and error suggested the square of $\sigma_\infty(\gamma,1)$,

$$\sigma_e = (1+\gamma^{-2}). \tag{21}$$

Termed the empirical radiation factor, σ_e returns a reasonably close approximation to the panel radiation efficiency, dashed line in Fig. 4. To produce a better match to the Rayleigh integral result, a least squares fit of a parametrized version of Eq. (21) yielded a radiation factor optimized for the case of a baffled panel radiating into free space,

$$\sigma_o = (1+1.1\gamma^{-4}). \tag{22}$$

The optimized radiation factor, dotted line in Fig. 4, is shown to closely match the Rayleigh integral result in the interval $0.3 \leqslant \gamma(\omega_m) \leqslant 5$. The reason for the error for small values of $\gamma(\omega_m)$ is that as the acoustic wavenumber reduces, the radius of the sonic circle approaches the bin width of the discrete wavenumber transform. Once the sonic circle radius equals the transform bin width, the result is a constant equal to the value of a single bin, which can be very small. The bandwidth of the radiation factor method can be extended to the lower frequencies by padding the velocity data to reduce the transform bin width, a topic discussed in Sec. II E.

E. Increasing the effective bandwidth of the radiation factor

The most straightforward way to increase the bandwidth of the radiation factor is to reduce the transform bin width by padding the velocity data, i.e., to extend the dimensions of the panel by adding zeros to the velocity matrix. The nature of the error introduced by filtering the discrete wavenumber transform, the effect of padding on the accuracy of the result, and the performance penalty incurred by increasing the amount of padding are discussed in Sec. 2 of the Appendix.

The transform bin width can be expressed as a function of the amount of padding by

$$\Delta k = \frac{2\pi}{n_p L}, \tag{23}$$

where Δk is the transform bin width, L is the panel length, and n_p is an integer pad multiplier. In Sec. 2 of the Appendix, it is shown that the accuracy of the result depends on the number of transform bin widths in the acoustic wavenumber. The acoustic wavenumber can be expressed in terms of the transform bin width to specify a desired accuracy,

$$k_a = n_r \Delta k, \tag{24}$$

where n_r is the number of bin widths in the sonic circle radius, k_a. The panel wavenumber is

$$k_p = \frac{n_m \pi}{L}, \tag{25}$$

where n_m is the mode order. An expression for the minimum $\gamma(\omega_m)$ that can achieve the specified accuracy can now be written as

$$\gamma(\omega_m)_{\min} = \frac{2n_r}{n_p n_m}. \tag{26}$$

It can be seen from Eq. (26) that high accuracy, i.e., $n_r \geqslant 5$, must be compensated with increased padding to reach small values of γ. For example, using $n_r=5$, $n_p=1$, $n_m=1$, and $L=1$ would yield $\gamma(\omega)_{\min}=10$. The amount of padding necessary depends on the frequency at which the mode occurs. Continuing the example, at 85 Hz $k_a=\pi/2$ rad/m and with $m=1$, $k_p=\pi$ rad/m, so $\gamma(\omega_m)=1/2$. A pad multiplier of 20 would then be needed to bring $\gamma(\omega_m)_{\min}$ to 1/2. Similarly, an expression for the minimum frequency for a given accuracy and pad level can be written as

$$\omega_{\min} = \frac{2\pi n_r c}{n_p L}. \tag{27}$$

The penalty for increasing the padding is increased computation time due to the larger matrix sizes. A performance metric for the operation can be taken as the ratio of the time it takes to compute the Rayleigh surface integral to that of

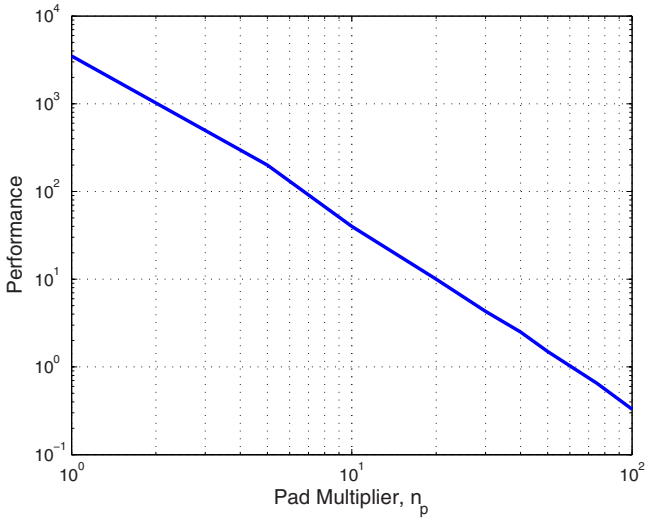


FIG. 5. (Color online) Filtered wavenumber transform performance metric.

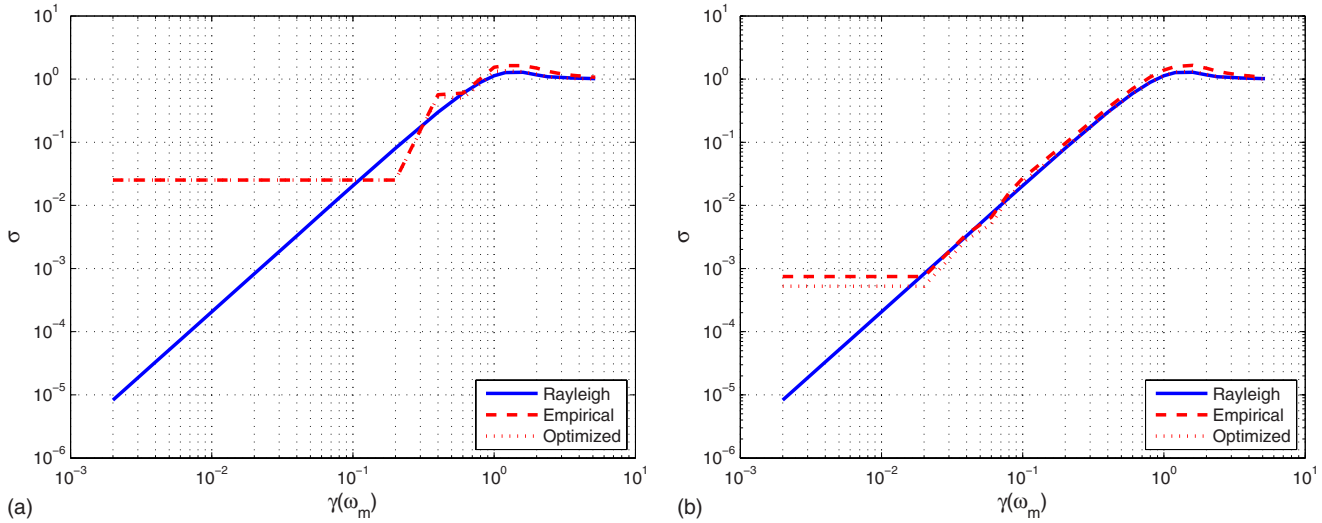


FIG. 6. (Color online) Radiation efficiency for square panel, mode 1,1: (a) n_p=5 and (b) n_p=80.

the filtered wavenumber transform, Fig. 5. At a pad multiple of 1, the filtered wavenumber transform is over three orders of magnitude faster than the Rayleigh integral. This performance advantage disappears at pad multipliers over 50. There are signal processing techniques that can be used to increase the performance by specializing the discrete Fourier transform to this application, but these techniques will not be addressed here. As discussed in Sec. III a pad multiplier in the range of 5–20, depending on the wave speeds in the panel, returns very good results.

III. EXPERIMENTAL DATA

Two sets of data are presented to validate the filtered wavenumber transform. The first set is simulated modal data for panels with varying aspect ratios. These are used to compute radiation efficiency for a specific mode over a range of gamma. The second set is physical data taken from panels of varying construction and wave speeds which are used to compute radiated sound power across a frequency band. The empirical and optimized radiation factor results taken at different pad levels are compared to the Rayleigh integral results in both sets. The numerically derived predictions are compared to measured intensity for the physical data set.

A. Simulated modal data

The simulated data were taken from simply supported, baffled panels of unit base dimension. Panels with aspect ratios of 1.0, 1.5, and 2.0 were evaluated. The modal velocities with mode numbers of 1–5 were computed on a grid with a spacing of 0.01. The lowest sample rate was then 20 samples per half-wavelength for a fifth order mode in the unit direction.

In Fig. 6, the first mode clearly shows the effect of increased padding where the usable range of gamma is increased by more than an order of magnitude by an increase in the pad level from 5 to 80. Equation (26) suggests that increased mode order will extend the valid range of gamma. This effect can be seen by comparing the 3,3 mode in Fig. 7 with the 1,1 mode in Fig. 6. At both pad levels, the lower limit of valid gammas is reduced by about a factor of 3.

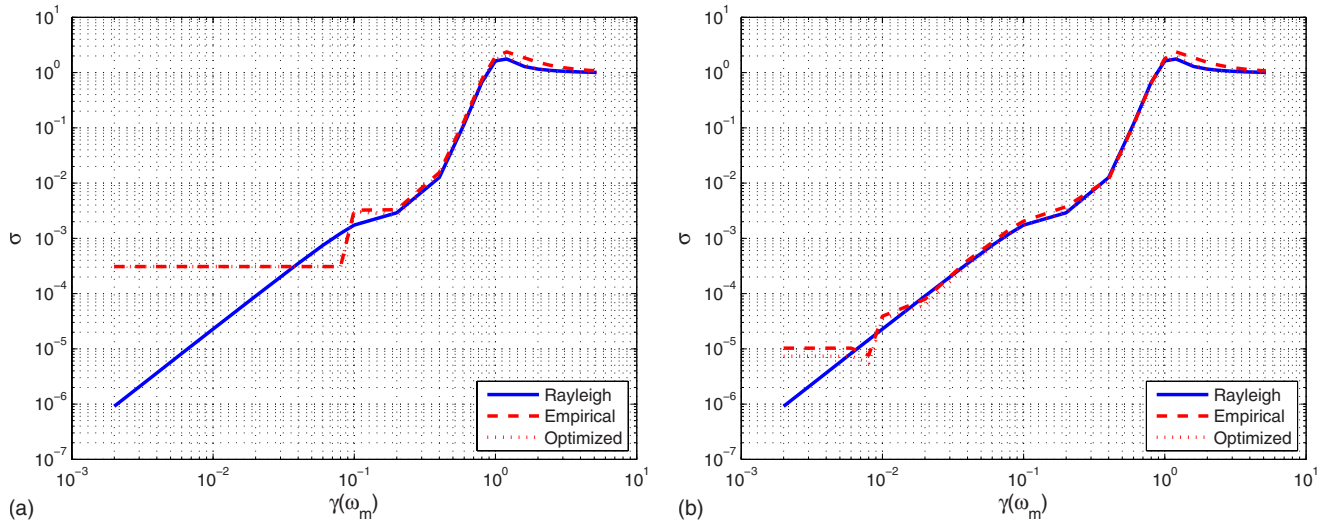


FIG. 7. (Color online) Radiation for square panel, mode 3,3: (a) n_p=5 and (b) n_p=80.

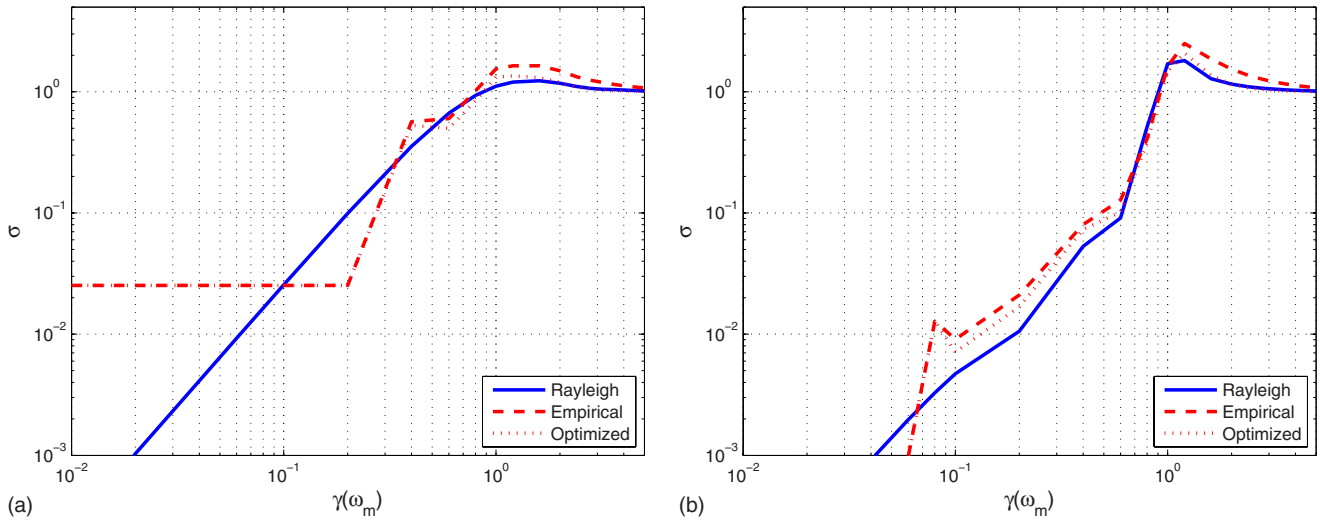


FIG. 8. (Color online) Radiation efficiency for rectangular panel with aspect ratio of 2 and n_p=5: (a) mode 1,1 and (b) mode 1,5.

When analyzing a rectangular panel, an increased mode order in the longer axis will improve the analysis. This is demonstrated in Fig. 8 where a rectangular panel with aspect ratio of 2 is analyzed with a pad level of 5. The 1,1 mode result (a) begins to oscillate at γ=0.5 and reaches its final value at γ=0.2. The 1,5 mode result (b) remains smooth but accumulates some error down to γ=0.1, then reaches its final value at γ=0.06.

Equation (26) predicts the lower limit of the analysis. For example, with n_r=1, n_m=1, and n_p=80, $\gamma_{\min}$=0.025. This is very close to the first valid value in Fig. 6(b) which occurs at γ=0.02. The analysis converges to the Rayleigh prediction at about γ=0.08, corresponding to n_r=4.

B. Vibration and acoustic data

Physical data were acquired from three panels in NASA Langley's Structural Acoustics Transmission Loss Facility (SALT). The panels were 46×46 in.^{2} (1.17×1.17 m^2) and were constructed of two different materials. Two panels were honeycomb composite panels with 0.75 in. (1.9 cm) Nomex core and 0.02 in. (0.51 mm) aluminum facesheets. The baseline honeycomb panel was supersonic for a substantial portion of the analysis bandwidth. A modified honeycomb panel had nine 10 in. (25.4 cm) square recesses cut into the core which were arranged in a 3×3 matrix, Fig. 9(a). The core in the recessed areas was reinforced with 0.016 in. (0.41 mm) aluminum panels. The third panel was a stiffened aluminum panel shown mounted in SALT in Fig. 9(b). The sub-panel dimensions are 17.25 in. (43.8 cm)×5.5 in. (14 cm).

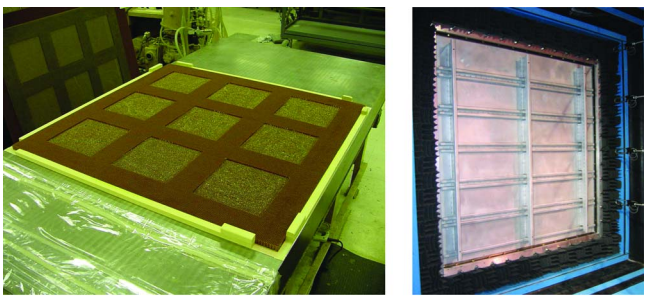


FIG. 9. (Color online) Test panels: (a) Nomex core with 10 in. (25.4 cm) recesses and (b) stiffened aluminum panel in SALT test fixture.

The panels were excited by pseudo-random noise with a bandwidth of 100–2000 Hz using a 10 N shaker through a PCB 288D01 impedance head. The frame into which the panels were clamped exhibits measurable response below 200 Hz, so some error can be expected at these low frequencies. The normal velocity structural response was measured with a Polytec PSV-300 scanning laser vibrometer. Velocity measurements were taken on a 1×1 in.^{2} (2.54×2.54 cm^2) grid. Normal acoustic intensity was measured using Bruel & Kjaer 2683 intensity probes on a 2×3 in.^{2} (5.08×7.62 cm^2) grid 5 in. (12.7 cm) from the panel's surface. A 12 mm spacer was used with the intensity probe which results in a 15 dB pressure-residual intensity index over the analysis band. The measured pressure-intensity index was 5 dB. Both the velocities and the intensities were normalized by the input force. Data were sampled at 5120 Hz providing a 2000 Hz bandwidth. Ensemble lengths were set to 800 to yield a 2.5 Hz bin width. The spatial Nyquist frequency for the velocity measurements was 124 rad/m. The data from the baseline honeycomb panel were well below the Nyquist frequency. The modified honeycomb panel had some subsonic energy exceeding the Nyquist frequency above 1500 Hz. The energy in the stiffened aluminum panel was below the Nyquist frequency. The spatial Nyquist frequency for the intensity measurements was 41 rad/m which just exceeds the acoustic wavenumber of 37 rad/m at 2000 Hz.

1. Baseline honeycomb panel

The radiated sound power based on measured intensity is compared to the Rayleigh integral estimate and the empirical and optimized estimates (n_p=20) in Fig. 10. Figure 10(a) shows the radiated sound power of all four measurements and Fig. 10(b) the difference between the intensity based result and the Rayleigh integral, empirical, and optimized results. The first observation is that all velocity based estimates are higher than the intensity based measurement ex-

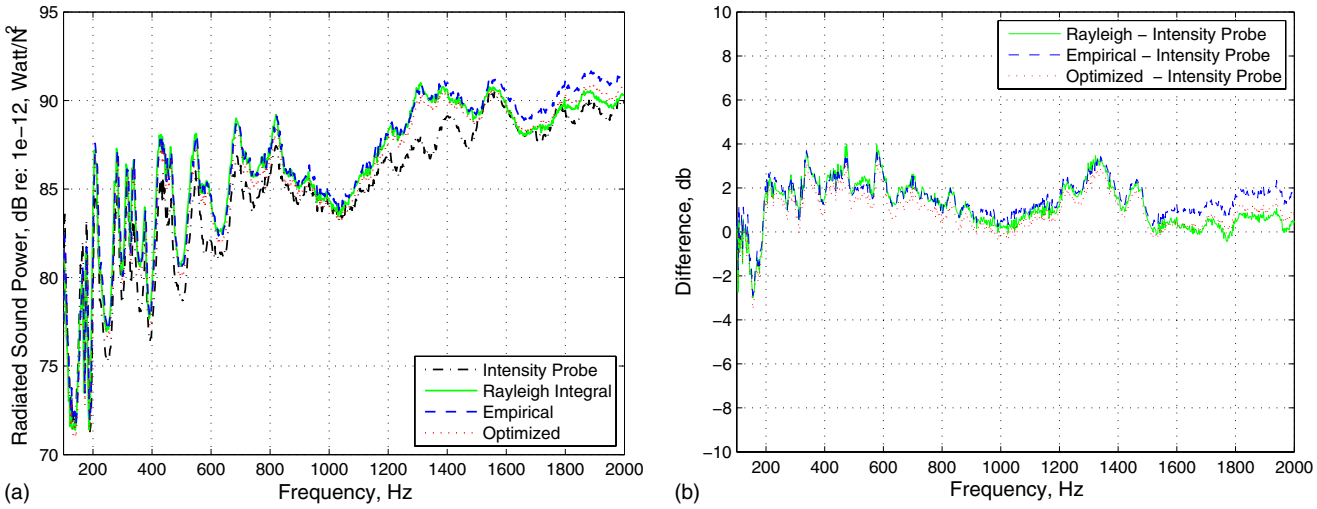


FIG. 10. (Color online) Baseline panel comparison of measurement techniques: (a) radiated sound power and (b) differences in radiated sound power.

cept for the band below 200 Hz where, as explained previously, some error is to be expected. The elevated radiated sound power returned by the velocity based estimates may be due to the nature of the intensity scan, being a planar grid instead of a hemisphere. At the critical frequency, i.e., when $k_a = k_p$, the vibration energy is just becoming supersonic and the radiation angle φ would be 90°. The acoustic radiation at this frequency would be grazing the panel's surface and may not be sensed by the instrumentation. The second observation is that the optimized estimate tracks the Rayleigh integral very well, with the empirical estimate returning a slightly higher value. Both the empirical and optimized estimates tend to drift above the Rayleigh integral at higher frequencies. The modal testing in Sec. III A indicated that the empirical estimate should over-predict at higher gamma levels. The fall-off with respect to the optimized estimate may be due to integration error in the Rayleigh integral, as discussed in Sec. 1 of the Appendix. In the remaining results only the optimized estimate will be shown as the general behavior of the empirical estimate is captured in Fig. 10.

The benefit of higher pad levels can be seen in Fig. 11(a) where the values obtained at a pad level of 5 begin to oscillate at about 600 Hz with increasing diversions from the $n_p = 20$ values as frequency decreases. The radiation efficiencies are plotted in Fig. 11(b) where the estimates derived from the velocities return values greater than those from the intensity measurements, as expected based on the radiated sound power results. The panel is supersonic above 1000 Hz. The reason for the pronounced dip of the intensity based value with respect to the velocity based values at about 1300 Hz is unknown, but may be due to the panel having substantial non-normal radiated power in this frequency band.

2. Modified honeycomb panel

Due to the recessed core, the modified honeycomb panel is less stiff than the baseline panel and should be expected to have lower radiation efficiency. This is confirmed in Fig. 12(a) where the radiation efficiency for the modified panel is 5 dB less than the baseline panel. As in the baseline case, the velocity based estimates are higher than the intensity estimate [Fig. 12(b)], but only by about 1 dB as compared to 2 dB for the baseline. The optimized estimate tracks just below the Rayleigh integral below 1000 Hz, after which the

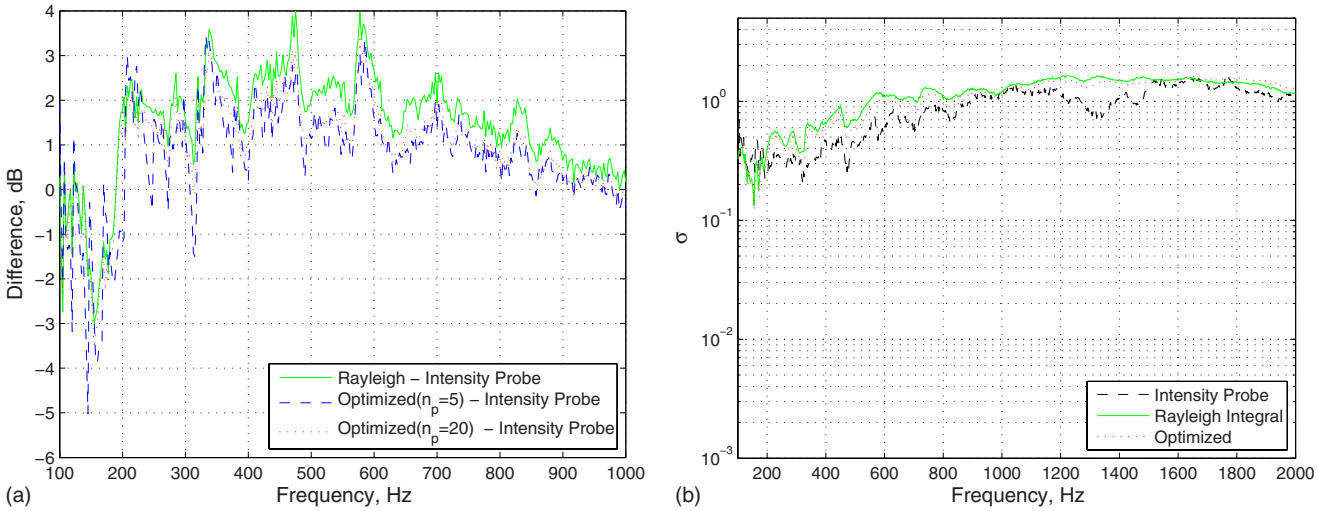


FIG. 11. (Color online) Baseline honeycomb panel: (a) effect of pad level and (b) radiation efficiency.

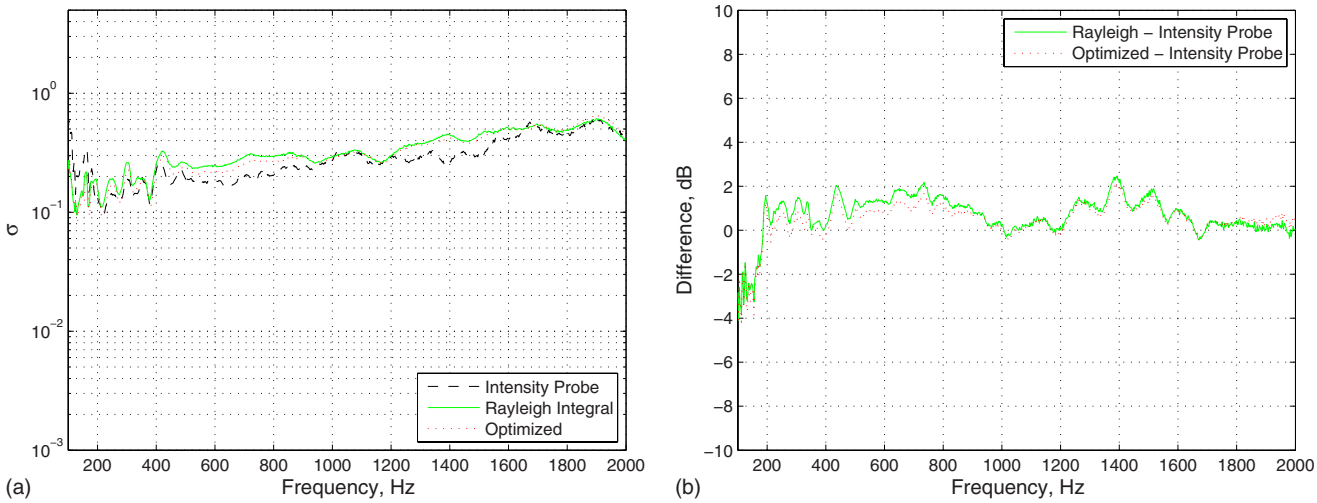


FIG. 12. (Color online) Modified honeycomb panel: (a) radiation efficiency and (b) difference in radiated sound power.

Rayleigh integral estimate begins to fall-off. This occurs at a lower frequency than in the baseline case where the fall-off begins at about 1500 Hz. If, as suggested previously, the fall-off is due to error in the numerical integral, the modified panel should exhibit a lower fall-off frequency as the shorter wavelengths created by the slower wave speeds would produce lower spatial sampling rates in the modified panel at any given frequency. The measurements were made on the side of the panel where the recesses are cut in the core.

3. Stiffened aluminum panel

The stiffened aluminum panel's material and construction are very different from the honeycomb panels. The aluminum panel is heavier, 8.8 kg vs 5.7 kg, than the baseline honeycomb panel and less stiff. The radiation efficiency of the stiffened aluminum panel, Fig. 13(a), is 5 dB below the modified honeycomb panel and 10 dB below the baseline. Even with these dissimilarities, the observed trends are similar. The velocity based estimates of radiated sound power, Fig. 13(b), are higher than the intensity based estimate, but not as high as the baseline honeycomb panel, supporting a trend of decreasing difference of the velocity based measurements to intensity measurements with lower wave speeds. The optimized estimate tracks the Rayleigh integral up to about 1000 Hz where it appears that the Rayleigh integral begins to fall off as it does with the honeycomb panels. The measurements were made on the skin side of the panel, thus avoiding the irregular structure that the stiffeners present.

C. Discussion of results

The behavior of the velocity based radiated sound power estimates is summarized in Table I as the mean error of the estimates with respect to the measured intensity. The results are given for the full analysis bandwidth as well as for the low frequency, 50–1000 Hz, and high frequency, 1000–2000 Hz, bands. The error in the Rayleigh integral falls from a high of 1.25 dB for the baseline honeycomb to a low of 0.36 dB for the stiffened aluminum panel. A large percentage of this reduction in error is due to a general reduction in error with slower wave speeds in the panel and the Rayleigh integral's tendency to under-predict at higher modal densities and frequencies. The latter behavior can be seen in the 0.4 dB drop in error between the low frequency

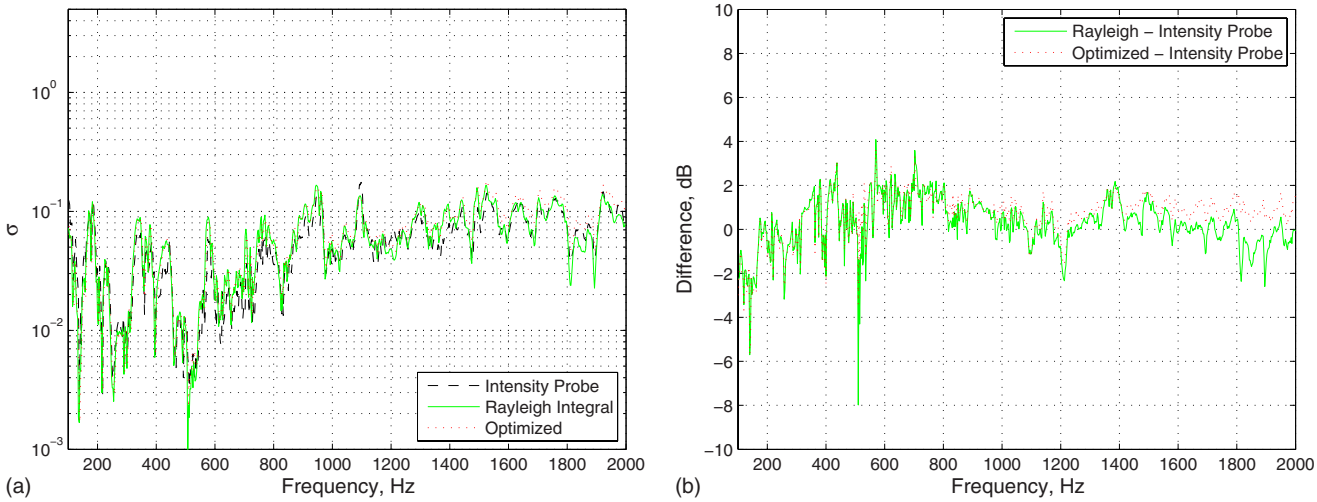


FIG. 13. (Color online) Stiffened aluminum panel: (a) radiation efficiency and (b) difference in radiated sound power.

TABLE I. Mean error of velocity based estimates with respect to intensity measurements.

Frequency range	Method	Baseline honeycomb	Modified honeycomb	Stiffened aluminum
Mean error (dB) 50–2000 Hz	Rayleigh	1.25	0.73	0.36
	Empirical	1.53	1.20	1.44
	Optimized	0.94	0.60	0.75
Mean error (dB) 50–1000 Hz	Rayleigh	1.42	0.93	0.56
	Empirical	1.50	1.14	1.31
	Optimized	0.96	0.55	0.63
Mean error (dB) 1000–2000 Hz	Rayleigh	1.07	0.50	0.15
	Empirical	1.56	1.25	1.56
	Optimized	0.93	0.64	0.86

and high frequency bands for all three panels. The radiation factor predictions are fairly consistent across frequency bands and panel types. The error in the optimized radiation factor's prediction varies by at most 0.4 dB across panels and 0.25 dB across frequency bands with no real pattern except that the optimized method returned the lowest error for the modified honeycomb panel. The error behavior of the empirical estimates mirrors that of the optimized estimate albeit 0.6 dB higher.

The performance of the methods can be evaluated as a combination of accuracy and computational efficiency. The accuracy of the Rayleigh integral can be improved by using adaptive integration methods; however, this comes at great cost in computational overhead as interpolation of the fixed grid of experimental data is required for compatibility with adaptive error reduction methods. The radiation factor method's accuracy can also be improved by expending additional computational resources, but this penalty can be greatly reduced using straightforward signal processing techniques. For example, increasing the pad level increases the virtual dimension of the panel by padding the velocity matrix with zeros. When computing the Fourier transform, the product of the velocity field and the sine/cosine need only be computed for the non-zero data and only at wavenumbers of interest, that is, from 0 to the acoustic wavenumber, k_a. This, in effect, integrates the supersonic filter with the computation of the transform. As the accuracy of the radiation factor method depends on the number of transform bins in k_a, the transform could be made adaptive, increasing the pad dynamically as necessary at the frequency of interest.

One last requirement for the method is generality. While generality cannot be proven by example, an effort has been made here to cover a wide variety of conditions to test the method's domain of applicability. In simulated modal tests, the Rayleigh integral was the limiting factor at high modal densities. This limitation could be addressed using adaptive integration methods as it lends itself to functional representation which would not require interpolation. This was not done here to keep the solution methods identical for the simulation and physical tests. The physical tests used models of very different construction, materials, and dynamics. In all cases, the radiation factor method has been shown to return results to within a fraction of a decibel of both measured intensity and predicted Rayleigh integral values.

IV. CONCLUSIONS

The radiation factor can be used with the filtered wavenumber transform to predict the sound power radiated from baffled panels with a high degree of confidence. The method as described here has both accuracy and performance advantages over the Rayleigh integral when used with experimental data and is thus the preferred method for preliminary design studies. Formulas have been derived, Eqs. (26) and (27), which predict the minimum gamma and frequency for which the radiation factor method is valid for a given level of accuracy and computational overhead. The formulas are of value in setting analysis parameters to meet a given study's requirements or in adapting the transform dynamically to achieve a desired accuracy at a particular frequency. On a cautionary note, the radiation factor method has been empirically derived and therefore cannot be shown to apply in general beyond those cases tested here. Further testing using simulated data that is amenable to adaptive integration meth-

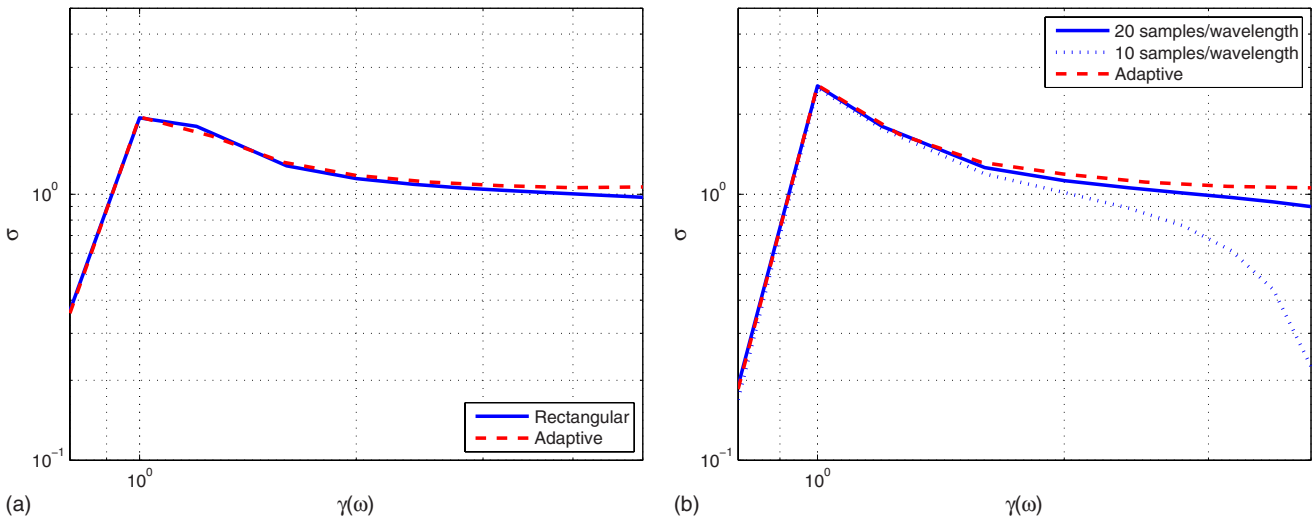


FIG. 14. (Color online) Rayleigh estimate of radiation efficiency for square panel: (a) n_p=80, mode 5,5 and (b) mode 10,10 at different sample rates.

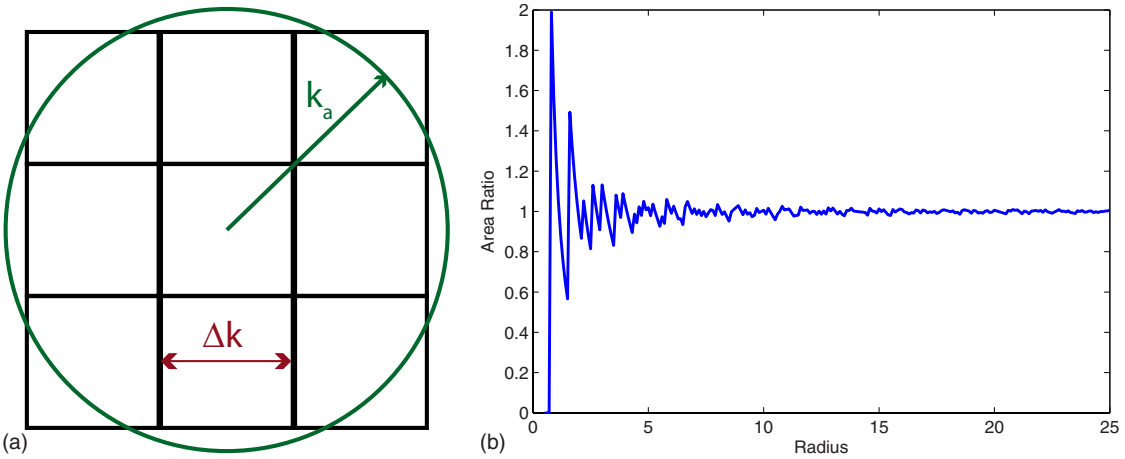


FIG. 15. (Color online) Estimating area of sonic circle: (a) sonic circle and elements of wavenumber transform and (b) ratio of area estimate to area of sonic circle.

ods will increase the accuracy of the Rayleigh integral allowing the expansion of the validated domain in frequency and modal order. This may also lead to a better realization of the optimized radiation factor. Alternatively, the radiation factor might be derived analytically, an endeavor that remains an open research question.

APPENDIX: COMPUTATIONAL ERROR

The Rayleigh integral and filtered wavenumber transform methods require numerical integration. Both methods are then subject to error which is a product of the discrete approximation inherent in the data. Although techniques exist to reduce the error, for example, by interpolating the data, the benefit comes at a cost of lost performance to the point where it can become computationally prohibitive. The two methods are thus compared using simple rectangular approximations to the integrals. This will produce error, the nature of which is described for each method in Sec. 1 and 2.

1. Rayleigh integral error

There appears to be error in the implementation of the Rayleigh surface integral at high mode numbers despite what would seem to be an adequate sample rate of 40 samples per wavelength for a 5,5 mode. In Fig. 14(a), the rectangular numerical integral is compared with an adaptive integral solution. At high frequencies the rectangular integral falls below the adaptive solution and even falls slightly below 1, indicating that it returns a low estimate for the radiated sound power. Although this error is only a fraction of a decibel, it increases with higher mode numbers, as can be seen for the 10,10 mode in Fig. 14(b). Notice also that the error increases considerably when the sample rate is halved. The tendency for the Rayleigh surface integral as implemented in this study to underestimate radiated sound power at high frequencies was observed in the physical data as well. Higher sample rates can reduce the error, but the processing time grows by over an order of magnitude for each doubling of sample rate. The adaptive integration solutions are more accurate in these regimes, but require interpolation of the data, which greatly increases processing time.

2. Padding the wavenumber transform

The fundamental operation in computing the filtered wavenumber transform estimates the area of the sonic circle in terms of the discrete wavenumber elements. This is shown graphically in Fig. 15(a), where Δk is the width of a transform element. If the transform element is considered to be located at its center, all nine transform elements would be included in the area estimate, overestimating the area. If k_a is reduced slightly, the four corner elements would not be included, underestimating the area.

This behavior is captured in Fig. 15(b) where the ratio of the estimated to the actual area of the sonic circle is plotted against the radius of the sonic circle normalized by the transform element bin width Δk. The oscillation from over to under estimation will continue until a single bin remains, in which case a constant, and potentially very small, value is returned. From the graph, it can be seen that it is necessary to have at least five transform bin widths in the sonic circle radius at the frequency of interest. There are graphical processing techniques that can be used to remove the aliasing effect of the discrete transform, but these methods are not considered here.

[1]G. Maidanik, "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826 (1962).
[2]C. E. Wallace, "Radiation resistance of a rectangular panel," J. Acoust. Soc. Am. **51**, 946–952 (1972).
[3]G. Xie, D. J. Thompson, and C. J. C. Jones, "The radiation efficiency of baffled plates and strips," J. Sound Vib. **280**, 181–209 (2005).
[4]L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, New York, 1987), pp. 526–534.
[5]M. Heckl, "Radiation from plane sound sources," Acustica **37**, 93–102 (1977).
[6]E. G. Williams, *Fourier Acoustics* (Academic, San Diego, 1999), pp. 43–46.
[7]E. G. Williams, "A series expansion of the acoustic power radiated from planar sources," J. Acoust. Soc. Am. **73**, 1520–1524 (1983).
[8]E. G. Williams,"Supersonic acoustic intensity on planar sources," J. Acoust. Soc. Am. **104**, 2845–2850 (1999).
[9]F. J. Fahy, *Sound Intensity* (Elsevier Science, New York, 1989), pp. 55–61.
[10]L. Cremer, M. Heckl, and E. E. Ungar, *Structure-Borne Sound* (Springer-Verlag, New York, 1987), pp. 502–505.

A comparative analysis of acoustic energy models for churches

Umberto Berardi,[a] Ettore Cirillo, and Francesco Martellotta
Dipartimento di Architettura e Urbanistica, Politecnico di Bari, via Orabona 4, I70125 Bari, Italy

(Received 3 February 2009; revised 17 July 2009; accepted 24 July 2009)

Different models to improve prediction of energy-based acoustic parameters in churches have been proposed by different researchers [E. Cirillo and F. Martellotta, J. Acoust. Soc. Am. **118**, 232–248 (2005); T. Zamarreño *et al.*, J. Acoust. Soc. Am. **121**, 234–250 (2006)]. They all suggested variations to the "revised" theory proposed by Barron and Lee [J. Acoust. Soc. Am. **84**, 618–628 (1988)], starting from experimental observations. The present paper compares these models and attempts to generalize their use taking advantage of the measurements carried out in 24 Italian churches differing in style, typology, and location. The whole sample of churches was divided into two groups. The first was used to fine-tune existing models, with particular reference to the "μ model," which was originally tested only on Mudejar-Gothic churches. Correlations between model parameters and major typological and architectural factors were found, leading to a classification that greatly simplifies parameter choice. Finally, the reliability of each model was verified on the rest of the sample, showing that acoustic parameters can be predicted with reasonable accuracy provided that one of the specifically modified theories is used. The results show that the model requiring more input parameters performs slightly better than the other which, conversely, is simpler to apply.
 [DOI: 10.1121/1.3205398]

PACS number(s): 43.55.Br, 43.55.Gx [NX] Pages: 1838–1849

I. INTRODUCTION

Theoretical models to predict acoustical parameters have been attracting renewed interest in recent years, with particular reference to churches. Computer simulations may provide detailed and reliable results, but they require a detailed three-dimensional model of the room under investigation, which may be a quite difficult task, especially for buildings such as churches. Conversely, theoretical models interpreting how sound propagates in such complex spaces provide simple prediction formulas to calculate reference values with little computational effort, aiding the general understanding of room acoustics in the process.

An energy model generally defines a law that describes the way in which sound energy propagates. These models can be divided into two categories: the theoretical ones, based on the interpretation of the sound field and on the definition of mathematical laws capable of describing its variations, and the empirical or semi-empirical ones, derived from correlations of data. The theoretical approach is generally more sophisticated, even though more complex, while the empirical approach gives relations that, even without a substantial improvement of the understanding of the physical aspects, may be as accurate as the others. A key issue that these models should satisfy is their reasonable ease of use. In fact, the success of the reverberation time among the acoustic descriptors of a space relies not only on its correlation with perceived subjective quality, but also on its steadiness throughout the space, and, above all, on its predictability with simple formulas,[1,2] which cover most of the cases, even churches. However, a number of other acoustic parameters have been defined in order to better match subjective perception of the sound field, even though the prediction of such acoustic parameters depends on many factors, such as the relative position of sources and receivers and any mathematical formulation becomes more complex.

According to the classical theory of diffuse sound propagation in rooms with uniformly distributed absorption, the sound pressure level is the sum of a direct component and a diffuse one. Barron and Lee[3] proposed a "revised theory" assuming that the reflected sound cannot arrive earlier than direct sound. This theory was proposed for concert halls or other proportionate spaces,[4] providing considerably better predictions of the actual behavior. A modification of this theory was proposed by Vörlander[5] for reverberant chambers.

Unfortunately, several studies[6–10] show that churches and other places of worship can hardly be included among the proportionate spaces. Measurements of both strength and clarity carried out in Mudejar-Gothic churches,[6–8] in mosques,[9] and in Italian churches[10] show that reflected sound level is below that predicted by previous theories. Possible explanations of this difference may be found in the "disproportionate" nature of these kinds of buildings, in the non-uniform distribution of sound absorbing materials, and in their architectural elements such as side aisles, chapels, vaults, and domes. These elements scatter or hinder the sound, especially affecting the early reflections where the energy value is greater. A simple empirical approach is to obtain prediction equations based on simple regression formulas.[11] However, the above-mentioned works[6–10] show that, despite some fluctuations, the acoustic energy parameters are well related to source-receiver distance. So, researchers attempted to modify the revised theory in order to take into account, in a more or less empirical way, the physical phenomena, that cause the early reflections to decrease. Those models were validated on a specified group of Spanish

[a]Author to whom correspondence should be addressed. Electronic mail: u.berardi@virgilio.it

churches[6,12] and on a wider sample of Italian churches,[13,14] leading to improved prediction accuracy. A brief overview of such models is reported in Sec. II.

Starting from the results of a preliminary study where the more recent model defined for Spanish churches[12] was used on a large sample of Italian churches,[15] the present work is aimed at investigating the possibility of generalizing this model to different churches by means of a typological classification. In addition, differences between models and, where possible, their similarities were investigated. Finally, the prediction accuracy of different models was analyzed through a comparison between measured and predicted values of sound strength, clarity, and center time.

II. OVERVIEW OF ENERGY MODELS

A. The classical theory and the "revised" theory

According to the classical theory of sound propagation in enclosed rooms,[1] if the absorption is uniformly distributed and if the sound field is diffuse, the relative sound pressure level (also known as strength, G) at a distance r from the source, assuming as a reference the level of the direct sound at a distance of 10 m from the source, and expressing total acoustic absorbing area as a function of the reverberation time (T) and of the room volume (V), through Sabine's equation is

$$G(r) = 10\ \log(100/r^2 + 31\ 200T/V) \quad [\mathrm{dB}]. \tag{1}$$

However, according to Eq. (1), when the direct sound becomes negligible, G only depends on T and V, and should be the same throughout the space. Other energy-based acoustic parameters, such as clarity (C_{80}) and center time (T_S), may also be calculated by taking into account that the instantaneous reverberant energy follows a decay function that decreases exponentially (with a time constant $T/13.8$) and, after integration, must yield the diffuse-field contribution in Eq. (1) as follows:

$$g(t) = (13.8 \times 31\ 200/V)e^{-13.8t/T} \quad [\mathrm{s}^{-1}]. \tag{2}$$

However, as observed for G, the resulting values of the parameters are substantially independent of the distance.

The classical formulation shows limited effectiveness in predicting the acoustic behavior in large rooms as measured acoustic parameters show much larger variations as a function of the distance from the source. Barron and Lee[3] found that the sound level decay was linear soon after the direct sound in the majority of the halls and the reflected sound level decreased with increasing source-receiver distance. So, they proposed a model based on the following four assumptions. (i) The direct sound is followed by linear level decay at a rate corresponding to the reverberation time. (ii) The instantaneous level of the late decaying sound is uniform throughout the space. (iii) The time $t=0$ corresponds to the time the signal is emitted from the source; therefore the direct sound reaches a point at a distance r from the source after a time $t_D=r/c$. In this way the integrated energy decreases when the source-receiver distance increases, while the early/late reflected energy ratio remains constant. (iv) The integrated value for the reflected sound level is assumed to be, at $r=0$, equal to the value predicted by the classical theory [Eq. (1)].

According to the revised theory, the integrated reflected energy from time $\tau=t-t_D$ (i.e., assuming the time origin at $t=t_D$) to infinity, at a point at a distance r from the source is given by

$$i(\tau,r) = (31\ 200T/V)e^{-0.04r/T}e^{-13.8\tau/T}. \tag{3}$$

In order to ease the calculation of the clarity index, the sound energy is divided into three components: the direct sound (d), the early reflected sound (from 0 to 80 ms, E_0^{80}), and the late reflected sound (from 80 ms to infinity, E_{80}^{∞}). From Eq. (3), the corresponding energies become

$$d(r) = 100/r^2, \tag{4}$$

$$E_0^{80}(r) = (31\ 200T/V)e^{-0.04r/T}(1 - e^{-1.11/T}), \tag{5}$$

$$E_{80}^{\infty}(r) = (31\ 200T/V)e^{-0.04r/T}e^{-1.11/T}. \tag{6}$$

So G, C_{80}, and T_S may be calculated according to Eqs. (10)–(12) given in Ref. 13.

B. The "μ model" for Mudejar-Gothic churches

The lack of accuracy of the revised theory when applied to churches was first pointed out by Sendra *et al.*,[6] who proposed an empirical correction of the theory, known as the β model. This model replaced the coefficient 0.04 (resulting from $13.8/c$) appearing in Eq. (3) with a coefficient β. The application of the model to some churches led to the empirical estimation of β values providing the best accuracy in predicting G. The resulting values varied between 0.06 and 0.12, suggesting a loss of energy that the authors attributed to the geometrical complexity of the churches.

However, as observed by Zamarreño *et al.*,[12] this model proved to be quite ineffective in predicting other monaural parameters, mostly because the β coefficient equally affects both early and late reflected energies. Taking advantage of measurements carried out in ten Mudejar-Gothic churches the same authors proposed an alternative approach capable of providing improved prediction accuracy. According to the new formulation the empirical coefficient, named μ, only affects the early part (from 0 to 80 ms) of the reflected sound. The μ coefficient is then derived from regression analysis in order to minimize the differences between measured and estimated values of a given acoustical parameter. This assumption is aimed at improving the agreement with measured values, even though it determines a discontinuity in the reflected energy function [Fig. 1(a)]. The authors justify this discrepancy by the discrete nature of the early reflections so that they consider it unnecessary to assume a continuous function.

According to the new approach only Eq. (5) needs to be rewritten as follows:

$$E_{\mu 0}^{80}(r) = (31\ 200T/V)e^{-\mu r/T}(1 - e^{-1.11/T}), \tag{7}$$

so that G and C_{80} may be calculated by simply replacing $E_{\mu 0}^{80}$ with E_0^{80}, while center time needs to be rewritten to account

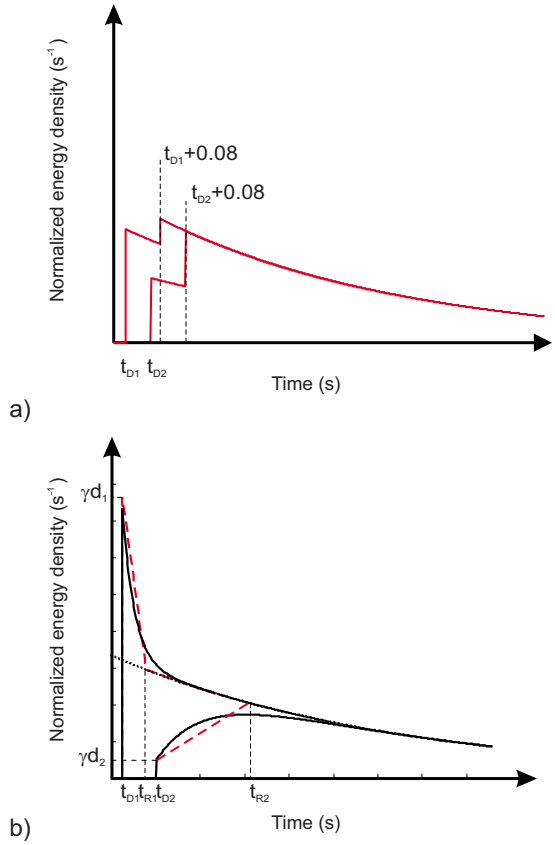


FIG. 1. (Color online) (a) Shape of the energy decay curve obtained with μ model observed at different distances from the source. The early reflected energy (from t_D to t_D+0.08 s) is given by an exponential energy decay having the same time constant of the late part but estimated at a later time (multiplied by $\mu/0.04$). (b) Shapes of the energy decay curve obtained with the linear model (---) and with the multi-rate exponential model (——) observed at different distances from the source. Dotted curve represents classical exponential decay.

for the displacement of the center of mass of the early reflected energy:

$$T_s(r) = \{[T - 1.11/(e^{1.11} - 1)]E_{\mu 0}^{80} + (1.11 + T)E_{80}^{\infty}\}/[13.8(d + E_{\mu 0}^{80} + E_{80}^{\infty})]. \quad (8)$$

Assuming that prediction error on G values predicted by the revised theory is relatively low, the authors proposed calculating μ values by minimizing the prediction error on C_{80}. In fact, this parameter shows the greatest variations inside a room, partly due to the arbitrary definition of the 80 ms limit for energy integration. The values calculated for Mudejar-Gothic churches show an average value of about 0.13, with a standard deviation of 0.02, so the authors assumed the value of 0.13 as specific for this group of churches. The use of the average μ gives reliable predictions of both G and T_S.

C. The "modified" theory

The analysis of the acoustic results measured in a sample of Italian churches[10,13] showed that the basic hypothesis of the revised theory, that is, the uniformity of the reverberant sound field throughout the space, was generally satisfied. However, the time at which the decay began to be linear was later, the farther the measurement position was from the source. Furthermore, at points near the source, the early reflections brought more energy than the ideal classical reverberant field while conversely, when the distance from the source grew, the early reflections became weaker.

TABLE I. Values of the mean scattering coefficient according to the chancel typology.

Chancel typology	s
Raised, or bounded by very close hard, flat surfaces	0.20
Slightly raised, bounded by relatively flat reflecting walls with few obstacles between source and walls	0.40
Very slightly raised, bounded by scattering (decorated) walls with some obstacles between source and walls	0.60
Not raised, bounded by distant reflecting walls and full of scattering furniture	0.80

In order to fit with these observations two modifications were introduced. The first one was to assume the reverberant sound field to be uniform, as it is in Barron and Lee's theory,[3] but that linear level decay starts with a certain delay (t_R) after the arrival of the direct sound. The measurements showed that this delay was proportional to the source-receiver distance; therefore, in general, it could be written as $t_R = kt_D$. The k coefficient depended on the room characteristics and it was demonstrated that it could be expressed as a function of the architectural features of each church, assuming integer values from 1 to 3 growing with church complexity (Table III in Ref. 13).

The second modification was to schematize the early reflected sound arriving between the direct sound and the reverberant sound field as a continuous linear function varying from an initial value (at time t_D), proportional through a factor γ to the energy of the direct sound, and a final value (at time $t_D + t_R$), equal to the energy of the reverberant field at the same time [Fig. 1(b)]. The factor γ depended on the mean absorption coefficient (α), the mean scattering coefficient of the surfaces close to the source (s), and on the mean free path. The estimation of the mean scattering coefficient was simplified by assigning values varying from 0.2 to 0.8 as a function of the mean characteristics of the area surrounding the sound source (Table I).

However, as observed by Zamarreño *et al.*,[12] the assumption of a linearly decreasing energy density made the mathematical formulation of the model a bit complex, with particular reference to the subdivision of the reflected energy into the early and late contributions. Taking advantage of a more detailed analysis of the fine structure of the early reflections, Martellotta[14] proposed expressing the reflected energy function in the form of a double-rate decay as a linear combination of two exponential decay functions. This gives a more elegant mathematical formulation and a considerable simplification of the calculations without any loss in accuracy. According to this refined model Eq. (2) may be rewritten as follows:

$$g'(t) = A_1 e^{-13.8t/T_1} + A_2 e^{-13.8t/T_2}, \quad (9)$$

where $T_1 = T$ and $A_1(r) = (13.8 \times 31\,200/V)e^{-0.04r/T}$, so that the first exponential decay coincides with Barron and Lee's,[3]

while T_2 and A_2 need to be adapted in order to fit the modified linear function [Fig. 1(b)]. As explained in Ref. 14, a convenient choice is to assume that $T_2=6.9t_R$ (so that the center of gravity of the second exponential falls in the middle of the t_R interval), and $A_2(r)=\gamma d-(13.8\times 31\ 200/V)e^{-0.04r/T}$, so that the initial value of the function g' is still γd. According to the proposed modifications the early and late energies (respectively, $E_0'^{80}$ and $E_{80}'^{\infty}$) can be rewritten as reported in Eqs. (13) and (14) in Ref. 14. In this way G and C_{80} may be calculated by simply replacing E with E' in Eqs. (7) and (8), while T_S needs to be calculated using Eq. (15) in Ref. 14 in order to account for the different positions of the centers of gravity of the two exponential functions. The new shape of the energy curve causes a slight decrease in the early energy at points close to the source, and an increase at farther points [Fig. 1(b)]. Consequently, it overestimates clarity and slightly underestimates center time in comparison with the "old" linear model.

III. THE ACOUSTIC SURVEY

A. Measurement technique

The measurements were carried out using an omnidirectional sound source made up of 12 120 mm loudspeakers (with a flat response from 100 Hz up to 16 kHz) mounted on a dodecahedron, together with an additional sub-woofer to cover the frequencies from 40 to 100 Hz. A calibrated measurement chain consisting of a GRAS 40-AR omnidirectional microphone together with a 01 dB Symphonie system was used to measure the sound pressure levels. A MLS signal was used to get the calibrated impulse responses to obtain the strength values. The other acoustic parameters were obtained by using high-quality impulse responses collected with a Soundfield Mk-V microphone, an Echo Audio Layla 24 sound card, and a constant envelope equalized sine sweep[16] to excite the room. Both the microphones had a flat response within the frequency bands considered in the present paper.

At least two source positions were used in each church. The source was placed 1.5 m above the floor. Ten receiver positions were used on average. In very large but symmetrical churches the receivers were only placed in one-half of the floor; otherwise they were spread to cover the whole floor area uniformly (Fig. 2). The microphone was placed 1.2 m from the floor surface. All the measurements and the calculations of the indices were made in unoccupied conditions, according to the ISO-3382 standard.[17] In particular, for the measurement of the sound strength (G) the sound power of the source was calibrated in a reverberation chamber, employing the same measurement chain and the same settings used during the on site survey.

B. The churches surveyed

Twenty-four churches located in Italy were considered in the present survey. The churches were chosen in order to include different typologies of buildings for age, style, dimensions, volume, and interior finishes. The whole sample of churches was divided into two sub-sets. The first one, including the same churches used in Ref. 13, was employed to find a general rule to assign values for the μ model, while the second sub-set was used to compare the performances of all the models when parameter values were assigned.

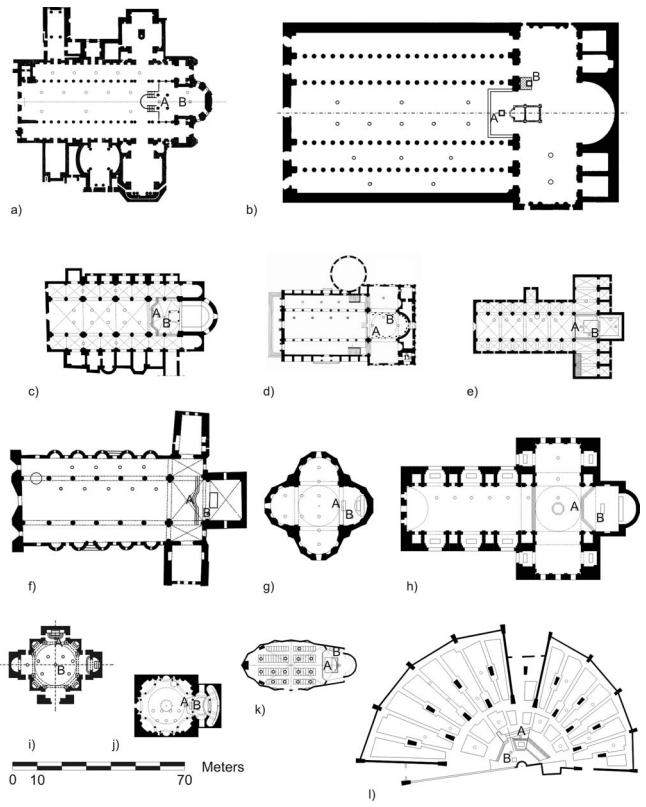


FIG. 2. Plans of the second set of 12 churches surveyed. (a) Basilica of Santa Maria Maggiore in Rome, (b) Basilica of Saint Paul outside-the-walls in Rome, (c) Sant'Ambrogio in Milan, (d) Bari Cathedral, (e) Abbey of Chiaravalle della Colomba in Alseno, (f) Orvieto Cathedral, (g) Santa Maria della Consolazione in Todi, (h) Sant'Andrea in Mantua, (i) Sant'Agnese in Agone in Rome, (j) San Lorenzo in Turin, (k) Santa Maria del Carmelo in Bari, and (l) Padre Pio Pilgrimage church in San Giovanni Rotondo. (Same scale for all the churches.)

For the sake of brevity the plans and descriptions of the first 12 churches are not included in this paper, as they can be found in Ref. 13, but Table II gives a summary of the most important architectural data.

The churches belonging to the second group were chosen adopting the same criteria used for the first one. In this way, even though the churches were clearly different, they could be classified according to the same geometrical and architectural characters (Table III). A short description of these churches is provided below; more detailed information may be found in Ref. 18.

The Basilica of Santa Maria Maggiore in Rome [Fig. 2(a)] is an Early-Christian church built at the beginning of the fifth century. It is built according to the typical basilica plan with side aisles even though several side chapels have been added over the centuries. It is covered with a coffered wooden ceiling with deep carvings, while the aisles are vaulted.

The Basilica of Saint Paul outside-the-walls [Fig. 2(b)] was one of the earliest churches built in Rome and, after a fire in the 1830s, it was faithfully rebuilt. The interior is a typical basilica plan 100 m long subdivided in a central nave

TABLE II. Basic details of the first set of churches surveyed.

Church	ID	Period	Style	Volume (m^3)	Total area (m^2)	Length (m)	$T_{1\ kHz}$ (s)
St. Sabina Basilica, Rome	SSA	432	Early-Christian	17 500	6 000	52	4.1
St. Apollinare in Classe, Ravenna	SAP	549	Byzantine	22 500	7 200	57	3.6
Modena Cathedral (Duomo)	MOD	1099	Romanesque	20 000	8 000	62	5.0
St. Nicholas Basilica, Bari	SNI	1197	Romanesque	32 000	10 500	59	4.4
Lucera Cathedral	LUC	1301	Gothic	33 100	10 500	64	5.3
St. Petronius Basilica, Bologna	SPB	1390	Gothic	160 000	42 000	130	9.8
Basilica Laurentiana, Florence	BLA	1419	Renaissance	39 000	18 000	82	7.9
The Holy Name of Jesus, Rome	JES	1568	Renaissance	39 000	13 000	68	5.1
St. Luca and Martina, Rome	SLM	1664	Baroque	8 700	3 500	35	3.1
St. Martin Basilica, Martina Franca	SMA	1763	Baroque	16 400	6 500	45	6.9
Concattedrale, Taranto	CCT	1970	Modern	9 000	6 200	50	4.2
S. Maria Assunta Church, Riola	RIO	1978	Modern	5 500	3 700	35	6.1

and four aisles by granite columns. The nave and the aisles are covered with a coffered ceiling made of wood.

The Basilica of Sant'Ambrogio in Milan [Fig. 2(c)] is an example of Lombard-Romanesque architecture. The interior is a conventional basilica plan without transept and with apses. The central nave is made of four square spans emphasized by cluster columns and ribbed cross vaults. The aisles are cross vaulted and topped with galleries.

Bari Cathedral [Fig. 2(d)] is a typical Apulian-Romanesque church, characterized by a basilica plan, with a main nave and two side aisles, a transept that is slightly larger than the total church width, and three semi-circular apses. The nave is flanked by marble columns alternated with pillars supporting rounded arcades, and covered by a roof with trusses.

The Abbey of Chiaravalle della Colomba in Alseno [Fig. 2(e)] is a typical Cistercian church following rigorous geometrical proportions. The walls, the pillars, and the ribs are made of facing bricks, while the vaults are finished in plaster. The transept is large, having the same width of the nave. The walls are free of decoration apart from some paintings.

The Cathedral of Orvieto [Fig. 2(f)] is a clear example of an Italian Gothic church. The interior is organized on a basilica plan, divided into three naves by ten columns and two pillars with richly decorated capitals supporting rounded arcades. The nave and aisles are covered by a wooden roof with trusses, while the transept and the choir are covered with ribbed cross vaults painted with frescoes.

The church of Santa Maria della Consolazione in Todi [Fig. 2(g)] reflects the ideal principles theorized by Renaissance architects. The Greek-cross plan derives from the combination of a square and four semi-circular apses topped by a dome on a high tambour. The decorations are shallow and made of stone.

The church of Sant'Andrea in Mantua [Fig. 2(h)] was built by Alberti in Renaissance style. The interior is a Latin cross, with a single nave flanked by four deep chapels on each side, alternating with smaller chapels. The nave is topped by an impressive barrel vault painted to simulate a coffered effect. The walls are finished in plaster and painted with rich decorative patterns.

The church of Sant'Agnese in Agone in Rome [Fig. 2(i)] is a typical baroque church, based on a central plan with elongated transversal braces and niches on the diagonals. Eight Corinthian columns in marble support the dome, emphasizing the octagonal shape of the central volume. The lower part of the church is finished in marble and stuccoes, while the upper is painted with frescoes and gilded decorations.

The church of San Lorenzo in Turin [Fig. 2(j)] was built

TABLE III. Basic details of the second set of churches surveyed.

Church	ID	Period	Style	Volume (m^3)	Total area (m^2)	Length (m)	$T_{1\ kHz}$ (s)
S. Maria Maggiore, Rome	SMM	410	Early-Christian	39 000	12 000	80	4.1
St. Paul outside-the-walls, Rome	SPX	383	Early-Christian	130 000	33 650	130	7.5
Sant'Ambrogio, Milan	SAB	1099	Romanesque	23 000	10 200	67	5.7
Cathedral of San Sabino, Bari	BAC	1100	Romanesque	30 150	9 500	59	4.8
Abbey of Chiaravalle, Alseno	ACC	1136	Gothic	13 500	7 500	59	5.6
Orvieto Cathedral	ORV	1308	Gothic	68 000	15 000	90	6.9
S. Maria della Consolazione, Todi	TOD	1508	Renaissance	19 000	4 400	39	7.9
Sant'Andrea, Mantua	SAD	1472	Renaissance	78 000	19 500	100	8.8
Sant'Agnese in Agone, Rome	SAA	1672	Baroque	14 500	5 300	28	5.0
San Lorenzo, Turin	SLO	1680	Baroque	12 000	4 500	34	4.0
Church of Carmelo, Bari	SMC	1960	Modern	9 700	3 000	46	4.2
Padre Pio Church, San Giovanni Rotondo	SGR	2004	Modern	51 000	15 600	56	5.5

by Guarini according to a central plan with a complex structure obtained by combining eight convex sides. On each side there are Corinthian columns supporting convex arches. The central space is characterized by the unique design of both the tambour and the dome with lantern.

The church of Santa Maria del Carmelo in Bari [Fig. 2(k)] is a contemporary church characterized by an almost elliptical space, even though the side walls are fragmented in order to prevent focusing effects. The roof is covered with perforated panels mostly absorbing low frequencies.

The Padre Pio Pilgrimage church in San Giovanni Rotondo [Fig. 2(l)] was designed by Piano. The church is organized according to a spiral movement focused on the chancel area, from which a series of stone arches of decreasing height radiates to hold the curved roof. The latter is made of wood and is supported by a sub-structure made of steel and wood. The interior part of the ceiling is finished with thin gypsum panels.

IV. A TYPOLOGICAL CLASSIFICATION FOR THE MODELS

A. Generalization of μ model

At the end of their study, Zamarreño *et al.*[12] expressed the hope to extend the model "to other types of closed rooms (…) and to establish μ values for different typologies." A preliminary investigation[15] showed that with changing church typology, the resulting μ values vary over a wider range than observed in Ref. 12, suggesting that a proper choice of the parameter might allow a generalization of the model. In order to check this possibility the μ model was applied to the first 12 churches. The analysis took into account C_{80} values, at a frequency of 1 kHz as a function of source-receiver distance. For each church the μ value, which minimizes the rms error between measured and predicted values, was found by numerical iteration. Predicted values were determined using room volume and measured reverberation time at the frequency of 1 kHz.

The values found in each of the churches surveyed are reported in Table IV. They vary in the interval from 0.13 to 0.42, respectively, corresponding to the churches of Sant'Apollinare in Classe (Ravenna) and St. Petronius Basilica (Bologna). The interval shows a much greater variability than in the study of Mudejar-Gothic churches.

The analysis of rms errors for sound strength, clarity, and center time shows that using μ values specifically derived for each church leads to reasonably accurate estimates of the acoustic parameters. According to the theory, μ can be considered as an attenuation coefficient that reduces the early reflected energy by increasing the source-receiver distance fictitiously. In the analyzed churches, μ is up to ten times greater than the coefficient originally used in the revised model, equal to 0.04. This observation suggests that in churches the early reflected energy is considerably lower than that predicted by the revised theory, consequently explaining the bad performance of the latter approach.

The observation that μ is quite stable in the homogeneous sample of Spanish churches, which was used to define the model, suggests that similar values of the parameter should be expected in churches with similar characteristics. A careful analysis of the correlations between μ values and architectural aspects hopefully might lead to a classification similar to that proposed for the modified theory, allowing generalization. A typological classification of religious buildings, although necessary for acoustical purposes, represents a difficult task. The μ model depends on a single parameter (excluding V and T), and this parameter should take into account both geometry and material distribution; therefore, it is somewhat equivalent to a combination of both the parameters considered in the modified approach. Those parameters have a well-defined physical meaning, related to architectural features, so the possible relationship between them and μ was investigated. A weak correlation appeared but the statistical significance was quite low ($R^2=0.421$, with a residual probability $p=0.022$), mostly because of the discrete nature of the k parameter.

A further attempt was made to relate μ with the stylistic characteristics of the buildings. The sample was divided according to the architectural styles and the average value of μ was calculated for each one. Taking into account the strong typological connotation that characterizes some architectural styles, and the specific use of materials or ornamentations, it is possible to interpret the influence of both these aspects. In particular, lower μ values were observed in compact churches, growing when the geometrical articulation increases.[15] Early-Christian churches often have a basilica plan with few decorations and many smooth wall surfaces, their spaces are substantially open because of the limited depth of the side aisles and the slenderness of the columns. Consequently, their mean μ was 0.15. Romanesque churches are finished using nearly the same materials used in early-Christian period, but the plan is often more complex, generally with deeper aisles, so the average μ value was 0.25. Gothic churches have much larger volumes, especially because of their height and their greater spatial complexity (mostly due to lateral chapels and thick pillars). These elements determine a considerable lack of early reflections with

TABLE IV. Values of the μ parameter in the first set of churches surveyed, and corresponding rms errors in the prediction of strength, clarity, and center time.

		rms error		
Church	μ	G (dB)	C_{80} (dB)	T_S (ms)
SSA	0.17	0.65	1.14	21.0
SAP	0.13	0.51	0.83	14.9
MOD	0.28	0.81	1.32	37.8
SNI	0.23	0.58	1.12	23.9
LUC	0.29	0.79	1.50	53.3
SPB	0.42	0.96	1.49	65.9
BLA	0.22	0.59	1.14	62.4
JES	0.34	1.32	1.52	50.5
SLM	0.23	0.71	1.12	19.5
SMA	0.33	2.22	1.39	113.7
CCT	0.17	1.94	0.95	48.2
RIO	0.16	0.47	0.99	25.8
Mean		0.96	1.20	44.7

TABLE V. Classification of μ values according to typological style, with corresponding examples, and mean values.

Typological style	Examples	Measured range	Interval	Mean μ
Auditorium-like church or basilica with narrow aisles	SSA, SAP, CCT, and RIO	0.13–0.17	0.12–0.20	0.16
Typical basilica plan with deep aisles or thicker columns	MOD, SNI, and BLA	0.22–0.28	0.20–0.28	0.24
Basilica plan with deep aisles with large transept or chapels between source and receiver	LUC, JES, and SMA	0.29–0.34	0.28–0.36	0.32
Basilica plan with deep aisles with transept and chapels and very high ceiling/vaults	SPB	0.42	0.36–0.44	0.40

high μ values (average 0.35) compared to the other styles. In Renaissance churches, the basilica plan and an average amount of decoration led to an average μ value of 0.28. Baroque buildings are characterized by a greater spatial complexity, with side chapels, domes, and often a Greek-cross plan. These elements determine a lack of uniformity in the sound field, so their average μ value was again 0.28. Modern churches have large reflecting surfaces, due to the use of rigid materials such as concrete, which, combined with their simpler plans, determined an average μ value of 0.16.

In conclusion, studying how μ varies as a function of stylistic aspects confirms that different plan typologies, together with materials and architectural characteristics, can significantly contribute to determining the most suitable μ value for a given church. Table V shows a classification of the analyzed churches according to their architectural typologies, with the corresponding intervals of variation for calculated μ values. According to the proposed classification it appeared reasonable to subdivide the full range of variation into equally spaced intervals, and to assign the corresponding mean values to each category. Among the analyzed churches, only St. Luca and Martina in Rome did not fit the classification in Table V, probably because having a central plan and a high dome may be hardly compared with the other churches. The typological difference between the latter church and the others suggests that some adjustments in the classification are required in order to obtain a full generalization. Nonetheless, the observed results demonstrate that the classification reported in Table V gives a predictive character to the μ model, so that when the spatial articulation is known, the corresponding μ may be chosen and the energetic parameters may be calculated at each point.

It is interesting to observe that the subdivision into equally spaced intervals led to a linear growth of the mean μ value assigned to each interval, suggesting that, starting from the reference value of 0.16, any addition of given architectural elements might correspond to an increase of 0.08 in the μ value. Taking into account that an increase in μ corresponds to a subtraction of early energy, and that the addition of architectural elements between source and receiver leads to a weakening of the early reflections (both because of scattering or masking), this incremental approach appears also physically consistent. For example, the presence of either the transept or chapels in a basilica plan increases μ by 0.08, while the simultaneous presence of both increases μ by 0.16. This "incremental" approach is particularly interesting in order to deal with peculiar combinations of architectural elements, not included in Table V.

The proposed typological classification was first verified on the same group of churches in order to determine whether the errors resulting from using group values for μ instead of church-specific values were acceptable. With clarity being the parameter on which μ estimation was based, its values were calculated using the new typological values given in Table V. The resulting prediction error was slightly increased in comparison with the corresponding errors obtained using church-specific coefficients. However, the variations were generally negligible with a maximum of 0.2 dB observed in just one case and an average variation of 0.03 dB, confirming that the proposed classification can be conveniently used to generalize the μ model.

B. Toward a unified typological classification for different models

The typological classification of μ values proved to be quite effective in assigning a suitable value to a church starting from the analysis of its architectural features. Furthermore, the incremental approach appeared to be very promising in its ability to ease the difficult task of deciding which typology is best suited for a given church, in particular, when dealing with churches having complex or unusual shapes. Given this premise, a similar approach can also be proposed for the modified model in order to limit the number of arbitrary decisions and provide a unified typological classification for both models.

As already mentioned the modified model takes into account two parameters: s, which is related to the characteristics of the "sending end" of the church, and k, which is strictly related to the church typology. Consequently, a unified treatment based on typological classification should take into account μ and k as they both refer to typological features.

The classification reported in Ref. 13 defined three classes for k, which could assume integer values varying from 1 to 3. However, experimental values of this parameter varied between 1.24 (in the church of Santa Sabina in Rome), and 3.54 (in Lucera Cathedral), with typological values often differing significantly from experimental values, suggesting that despite the greater simplicity resulting from using only integer values, this approximation could lead to less accurate classification.

Taking into account the three churches belonging to the first class of the μ classification, the resulting average of experimental k values is 1.39. A straightforward extension of the μ classification to k values could then be to assign four classes of values from 1.4 to 3.5 with increments of 0.7. However, in order to take advantage of the incremental ap-

TABLE VI. Incremental values of k and μ according to the architectural characteristics.

Typical features	$k=t_R/t_D$	μ
Reference value to assign for any church	0.7	0.08
Additional scattering elements (decorations, thin columns, and roof)	+0.7	+0.08
Additional volumes between source and receiver as transept	+0.7	+0.08
Additional scattering/coupling elements such as big pillars or chapels along walls	+0.7	+0.08
Very high vaults or domes on tambour	+0.7	+0.08

proach and provide a simpler instrument to define typological values for both μ and k, the classification given in Table VI is finally proposed.

It can be observed that a new "reference" category was included to account for very simple spaces, which nonetheless cannot be considered sufficiently "proportionate" to fit Barron and Lee's model.[3] Groups of additional architectural features are listed, each corresponding to an incremental value of 0.7 for k and 0.08 for μ. In this way each church may assume k values from 0.7, for auditorium-like shapes, to 3.5, which should be considered only for very complex and large rooms with many additional scattering elements. This classification is particularly interesting, because it allows the rapid definition of the value of the typological parameter (be it k or μ). Simple volumes get a very low value, while complex spaces get larger values with increasing complexity. It is interesting to observe that the original typological classification given in Table V can be easily obtained by direct application of the incremental approach. In addition, the usefulness of the incremental approach can be observed, for example, in assigning the correct value to the church of St. Luca and Martina, which was difficult to assess according to Table V. In fact, starting from the reference value (0.08) and taking into account the presence of scattering elements due to decorations (+0.08), the transept braces between source and receivers (+0.08), and the dome (+0.08), the resulting μ value is 0.32, in good agreement with the empirical value.

Table VII shows that according to the new classification, the agreement with experimental values improved for some churches while it worsened for others, with substantially negligible differences (the largest rms variations being 0.08 dB for G, 0.09 dB for C_{80}, and 3.8 ms for T_S), so that the average absolute difference was substantially unchanged. This proves that the original classification and the new incremental classification are equivalent in terms of accuracy of results. So, the incremental approach is to be preferred if only for ease of use and for providing a unified selection criterion shared with the μ model.

V. COMPARISON BETWEEN ENERGETIC MODELS

The models investigated in this study together with the proposed typological classification were finally validated using the second sub-set of 12 churches. The performance of the revised model,[3] the modified double-rate model, and the μ model were compared by means of rms errors between measured and predicted values of G, C_{80}, and T_S.

Table VIII reports the parameters μ and k, resulting from application of the incremental typological classification, together with the parameter s required by the modified model and depending on the scattering characteristics of the chancel where the sound sources are located. In order to clarify the reasoning behind the assignment of the typological values of μ and k and, at the same time, provide a useful example of the way in which the incremental classification should be applied, Table VIII also reports the "sum" of the architectural features that individually contributed to the final rating.

As stated above, churches are complex buildings and assigning typological values may be a difficult task in which subjective factors (architectural and acoustic background, detailed knowledge of the building, and so on) may somewhat influence the final result. In order to understand how subjective factors may influence the choice of the input parameters μ, k, and s, five students with good architectural background (but limited acoustic knowledge) were asked to assign parameter values starting from the plans reported in Fig. 2 and

TABLE VII. Summary of experimental k values for each church surveyed, together with those assigned according to the old and the new classifications and absolute differences from experimental values. The variations in rms error resulting from application of the new classification, with reference to prediction of G, C_{80}, and T_S, are also reported for each church.

Church	k exp	Old k	New k	Δrms G	Δrms C_{80}	Δrms T_S
SSA	1.45	1.0	1.4	0.08	−0.09	3.80
SAP	1.24	1.0	1.4	0.07	−0.06	−1.71
MOD	1.90	2.0	2.1	0.01	−0.02	0.29
SNI	2.00	2.0	2.1	0.01	−0.03	−1.00
LUC	3.54	3.0	2.8	−0.04	0.04	0.64
SPB	2.84	3.0	3.5	−0.04	0.05	−1.25
BLA	3.25	2.0	2.8	−0.03	−0.00	0.26
JES	2.97	3.0	2.8	−0.00	−0.02	−2.36
SLM	3.30	3.0	2.8	0.03	0.02	−0.42
SMA	2.99	3.0	2.8	−0.03	0.03	0.16
CCT	3.00	3.0	2.8	0.07	−0.01	2.59
RIO	1.48	1.0	1.4	−0.02	0.01	−0.07
Mean absolute difference				0.04	0.03	1.21

TABLE VIII. Values of the coefficient derived for an application of the μ and modified models to the second set of churches. C=columns (scattering), Ch=chapels, T=transept, S=other scattering elements, and D =dome/vaults.

Church	Typological features	μ typology	μ calculation	k typology	k calculation	s typology
SMM	C	0.16	0.20	1.4	1.3	0.2
SPX	C	0.16	0.16	1.4	1.1	0.2
SAB	C+Ch	0.24	0.20	2.1	2.0	0.4
BAC	C+T	0.24	0.29	2.1	1.5	0.4
ACC	C+Ch+T	0.32	0.29	2.8	2.7	0.4
ORV	C+Ch+T	0.32	0.38	2.8	2.3	0.4
TOD	T+D	0.24	0.31	2.1	1.7	0.2
SAD	Ch+T+D	0.32	0.28	2.8	3.1	0.6
SAA	S+T+D	0.32	0.19	2.8	2.1	0.2
SLO	C+S+D	0.32	0.19	2.8	2.3	0.6
SMC	S	0.16	0.13	1.4	1.8	0.4
SGR	S	0.16	0.06	1.4	1.3	0.2

the classification reported in Tables I and VI. Median values corresponded to the values reported in Table VIII even though different evaluations were observed in a number of cases. Anyway, for all the subjects, the "error" in the typological choice was circumscribed to one incremental class above or below the median value. The inaccuracies resulting from errors in the choice of input parameters are discussed in Sec. V D.

The comparison between measured and predicted values of the three acoustic parameters (Fig. 3) was limited to receivers located in unobstructed areas in order to prevent the possibility that both direct sound and early reflections might be hindered, thus determining abnormal variations in the measured parameters that could arduously be accounted for by the predictive models.

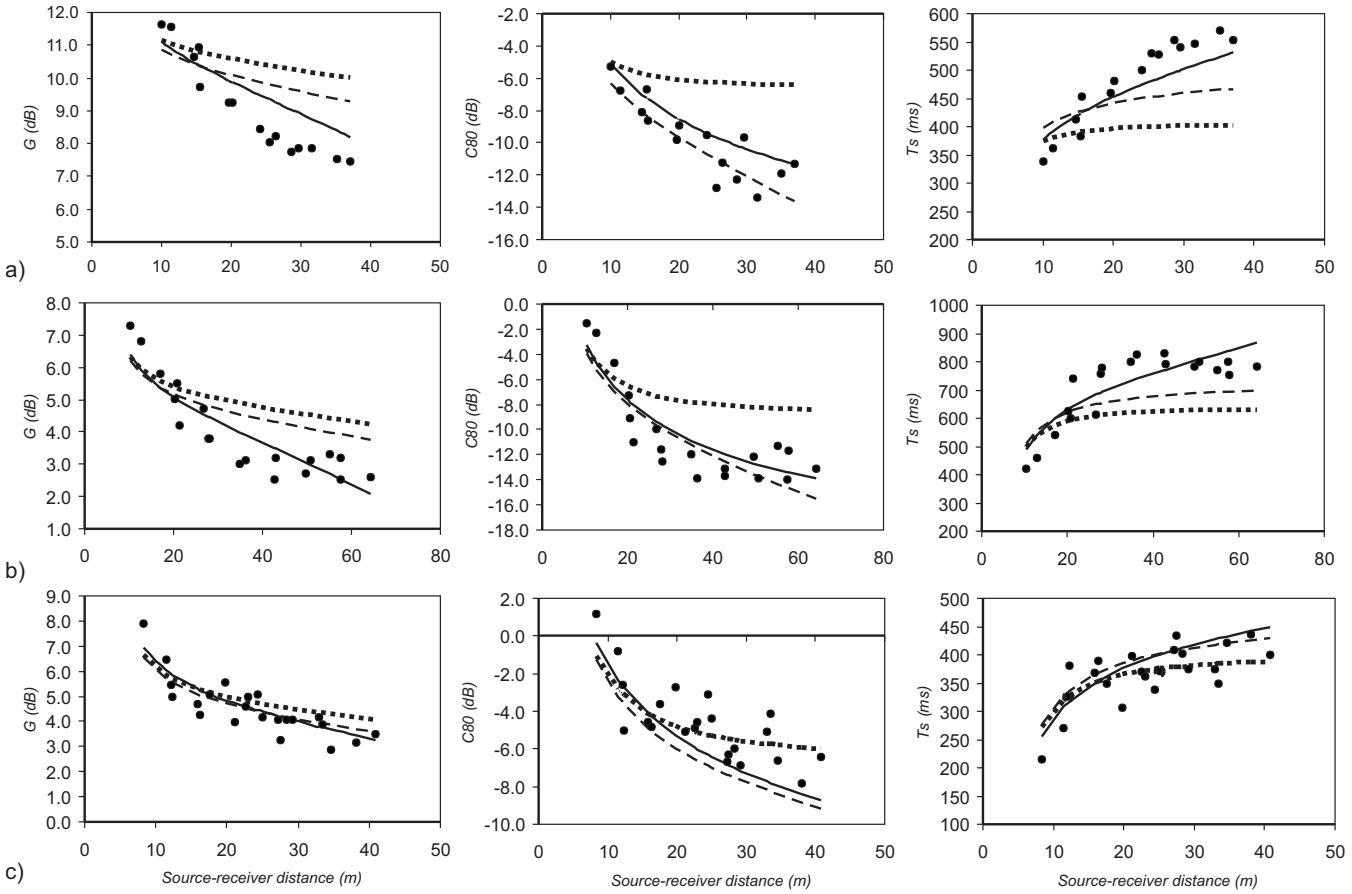


FIG. 3. Plot of measured (●) and predicted values of the strength (left), clarity (center), and center time (right) index at 1 kHz vs source-receiver distance according to the three models considered, revised theory (…), μ model (- - -), and modified theory (—). (a) ACC, (b) SAD, (c) SGR.

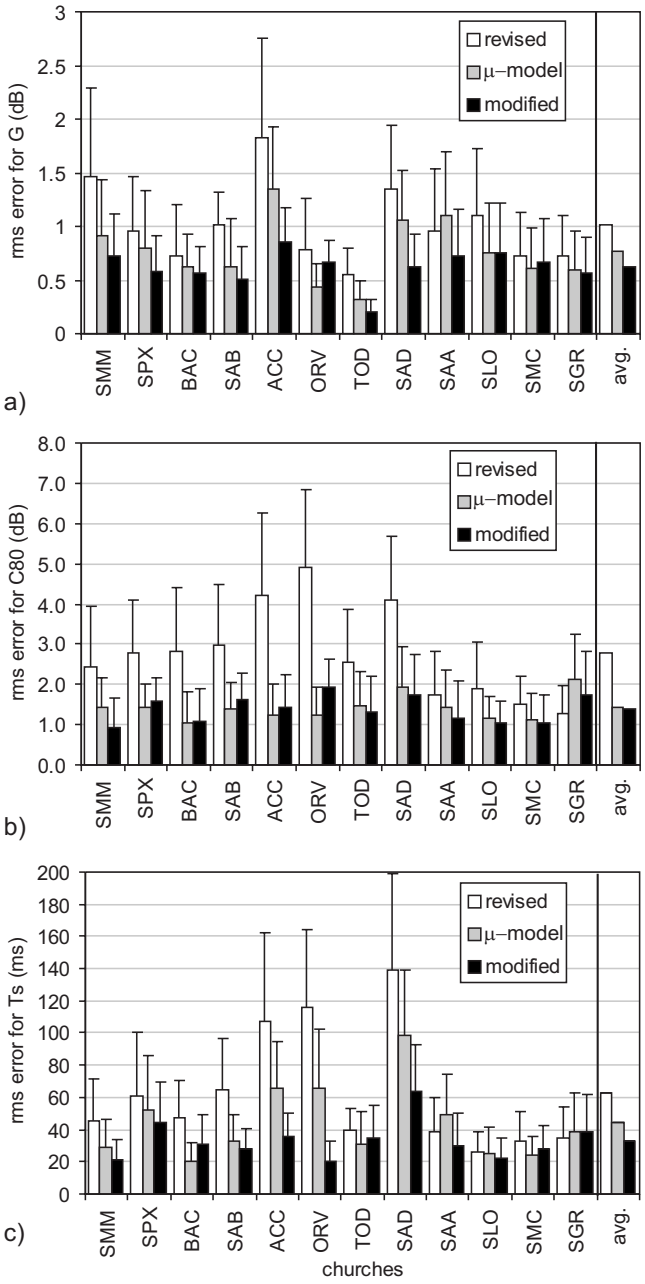


FIG. 4. rms error between measured and theoretical values of sound strength (a), clarity (b), and center time (c), at 1 kHz frequency band: theoretical values calculated with revised model, μ model, and modified model with the new classification. Error bars represent the standard deviation of point-to-point differences between measured and predicted values.

A. Sound strength

Among the acoustic parameters taken into account, sound strength shows the smallest variations in churches, possibly because it is largely influenced by long reverberation rather than by early reflections. As a consequence, this parameter might be predicted with reasonable accuracy even using the revised theory. However, as shown in Fig. 4(a), using either the modified or the μ model the prediction accuracy may be further improved. In fact, the average rms error varied from 1.01 dB for the revised model, to 0.77 dB for the μ model, to the smallest error of 0.62 dB obtained with the modified model.

The largest error was observed in Chiaravalle Abbey [Fig. 3(a)], where measured G showed a steep decrease as a function of distance (probably because of a lack of early reflections suppressed by the large transept), which was predicted with reliable accuracy only by the modified model. Similarly, the steep decrease observed in Sant'Andrea in Mantua [Fig. 3(b)] was better predicted by the modified model. The smallest errors were observed in Santa Maria della Consolazione where, probably because of the relatively small dimensions, the variation as a function of the distance was less evident.

B. Clarity

Comparisons between measured and predicted values of clarity were particularly interesting because the μ model was initially defined by calculating the μ value that provided the best agreement between predicted and measured data for clarity. Conversely, thanks to the proposed typological classification, μ values were then assigned, possibly leading to a loss in accuracy. Figure 4(b) shows that the μ model and modified model behaved the same, with an average rms error of 1.4 dB, nearly halving the error resulting from application of the revised theory (about 2.8 dB). Taking into account the performance for each church, it can be observed that in some cases the μ model provided better accuracy, while in other cases the modified model performed better, so that on average there was a substantial balance. It is interesting to observe that in Padre Pio Pilgrimage church in San Giovanni Rotondo [Fig. 3(c)] both models underpredicted clarity, while the revised theory provided the lowest error. This might depend on the relatively low ceiling that is likely to reflect early energy within the first 80 ms in a manner that is more similar to concert halls and auditoriums.

In some cases, such as Orvieto, Chiaravalle, and Mantua, the improvement resulting from using the μ or modified models was impressive, leading to errors that were one-third of those resulting from the revised theory.

C. Center time

Center time describes the temporal distribution of the sound energy without being affected by clarity drawbacks due to the arbitrary choice of a time limit to discriminate between useful and detrimental reflections. As a consequence T_S shows a distribution as a function of source-receiver distance, which is generally less scattered compared to C_{80} and, consequently, is more interesting to compare with values predicted by means of formulas. In addition, as the μ model is calibrated on clarity values (and its prediction accuracy is expected to be good for clarity), it is particularly interesting to validate its performance on a different parameter describing the same subjective attribute.

Figure 4(c) shows that on average, the modified model made it possible to nearly halve the rms error resulting from application of the revised theory (from 66 to 33 ms), while the μ model was in between (44 ms). Analysis of individual results shows that, as observed for clarity and, reasonably, for the same reasons, the Padre Pio Pilgrimage church in San Giovanni Rotondo [Fig. 3(c)] was the only one where revised theory is slightly more accurate (by 5 ms) than the

TABLE IX. rms error between measured and predicted values of sound strength, clarity, and center time, at frequency bands from 125 Hz to 4 kHz. Predicted values calculated with revised model, modified model, and μ model.

	rms error G (dB)			rms error C_{80} (dB)			rms error T_S (ms)		
Frequency	Revised	μ model	Modified	Revised	μ model	Modified	Revised	μ model	Modified
125	1.35	1.42	1.36	2.64	2.84	2.48	73	65	65
250	0.98	0.96	0.91	2.28	2.28	2.15	63	53	56
500	0.93	0.77	0.69	3.06	1.64	1.76	72	50	41
1000	1.01	0.77	0.62	2.77	1.42	1.39	63	45	33
2000	1.27	1.07	0.89	2.33	2.29	1.60	55	46	46
4000	1.87	1.33	1.05	1.89	3.47	1.79	36	43	43

other models. Among the other churches it is interesting to observe that the three churches that had already shown a significant improvement in prediction accuracy for C_{80} when the modified and μ models were used also confirmed the improvement for T_S. However, in this case the modified model provides better predictions than the μ model, probably because the latter affects only the first 80 ms, while in large churches weak early reflections are likely to arrive well beyond that time limit. Nonetheless, the modified model behaved quite badly in Mantua, with a rms error of about 66 ms (well above the average), probably as a consequence of the unusual decrease in values measured at the farthest points, possibly as a consequence of strong reflections from the back wall.

D. Frequency dependency and error analysis

In order to analyze the prediction accuracy over the whole spectrum, rms errors were calculated using the same parameter values defined at 1 kHz and assessing the difference between measured and predicted values over octave bands from 125 to 4000 Hz (Table IX). It can be observed that the models confirm the trends shown at 1 kHz even though larger errors appear at the extreme bands, where revised theory provides comparable levels of accuracy.

In particular, at 125 Hz all the models predict G values with errors above 1 dB, with revised theory performing slightly better than the others. From 250 to 4000 Hz the modified model provides the best results with errors below 1 dB except at 4 kHz.

When predicting C_{80} and T_S all the models perform badly at frequencies below 500 Hz, with larger errors than those observed at 1 kHz. From 500 to 4000 Hz modified model provides errors slightly above those observed at 1 kHz. The μ model behaves similarly when predicting T_S, but it becomes quite inaccurate when predicting C_{80} above 1 kHz; in fact, the rms error is 3.5 dB at 4 kHz. These larger errors are intrinsically due to the reduction in reverberation time typically observed at higher frequencies. Consequently, as the early reflected energy corresponds to the energy arriving $\mu/0.04$ times later, it may be significantly reduced when reverberation time gets shorter.

A further validation of the model performance was carried out by calculating the variation in prediction accuracy following an incorrect choice of the input parameters. Assuming that input parameters μ and k are badly selected by, at most, one "typological" class (i.e., respectively, by ± 0.08 and ± 0.7), the corresponding variations in the rms error are reported in Table X. Results show a general increase in the rms error, even though variations are small enough to allow both models to perform better than revised theory. The largest differences are observed for the μ model with reference to C_{80}. Figure 5 compares both models including the combined effect of the error in selecting k and s values. In general both models show larger differences as the source-receiver distance grows, with the highest variation observed when predicting C_{80} using the μ model. The error resulting from bad selection of s values is smaller than the error due to k and is negligible when k is underestimated, and a bit larger when k is overestimated.

VI. CONCLUSIONS

The results of a widespread acoustic survey taking into account a large number of Italian churches have been used to investigate and compare the prediction accuracy of energetic models in this kind of room. The whole sample of 24 churches was split into two halves in order to use the first group to define models and the other for their validation. In particular, the μ model, which was originally defined with reference to Mudejar-Gothic churches, was first generalized by showing that church-specific μ values could be grouped

TABLE X. rms average of the difference between errors corresponding to μ and k values assigned in Table VIII and errors obtained when typological class is badly selected. Upper class corresponds to $k+0.08$ and $\mu+0.7$, whereas lower class corresponds to $k-0.08$ and $\mu-0.7$. All values refer to at 1 kHz octave band.

	rms error G (dB)		rms error C_{80} (dB)		rms error T_S (ms)	
Church	Modified	μ model	Modified	μ model	Modified	μ model
Upper class	0.17	0.09	0.14	0.59	9.7	5.3
Lower class	0.22	0.16	0.25	0.51	7.9	8.6

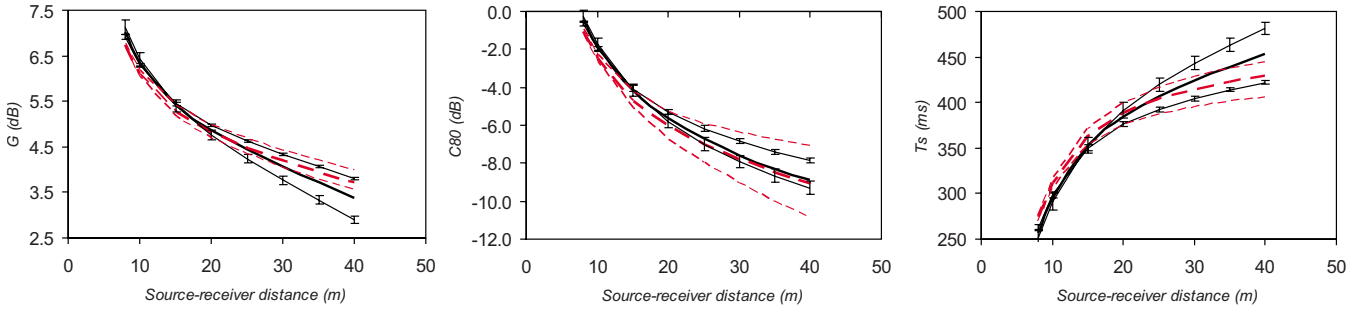


FIG. 5. (Color online) Plot of variations due to bad parameter estimation calculated for strength (left), clarity (center), and center time (right) index at 1 kHz as a function of source-receiver distance according to μ model (- - -) and modified theory (—). Thicker curves correspond to reference values calculated assuming $V=50\,000$ m^3, $T=5.5$ s, $S=15\,600$ m^2, $k=1.4$, $s=0.4$, and $\mu=0.16$. Thinner curves correspond to $k\pm0.7$ and $\mu\pm0.08$. Error bars represent variations corresponding to $s\pm0.2$.

according to typological characteristics. A typological classification was proposed using an incremental approach, so that any given architectural feature determines a corresponding increment of μ. This approach was then extended to the assignment of the parameter k for the modified theory, without significantly affecting its prediction accuracy.

The second half of the sample was first used to give an example of how the typological classification (valid for both the modified and μ models) should be applied as a function of architectural features. Then, the revised, modified, and μ models were compared in terms of their accuracy in predicting G, C_{80}, and T_S at 1 kHz. The results showed that strength could be predicted with reasonable accuracy (average rms error of 1.01 dB) even using the revised theory. However, the best results (average rms error of 0.62 dB) were obtained when the modified theory was used. Clarity showed larger variations as a function of source-receiver distance; consequently the revised theory gave on average larger rms errors of about 2.8 dB, while both the modified and μ model gave approximately the same rms error of 1.4 dB. Similarly, when dealing with center time, the modified model halved the error resulting from application of the revised theory (from 66 to 33 ms), while the μ model gave intermediate results (average rms error of 44 ms). Extension to other frequency bands showed a worse performance below 500 Hz, while at high frequencies the modified model provided consistent predictions (with slightly higher errors) for all the parameters. The μ model performed similarly for both G and T_S, while for C_{80} it showed much larger errors.

In conclusion, it can be stated that both models based on typological classification allowed an improvement in prediction accuracy compared to the revised theory. Among them, the μ model gave slightly less accurate predictions but required fewer input data (namely V, T, and μ), while the modified model gave the highest accuracy for all the acoustic parameters taken into account but required more parameters to be known (V, S, and T) and assigned (s and k). This led us to conclude that the choice of the model to use may be made according to the desired level of accuracy or, conversely, on the availability of the required input data, preferring the μ model when either less information is available or lower accuracy is desired, and the modified model when more information is available and the highest accuracy is desired. Such conclusions apply to midfrequencies, while at high frequencies the μ model should be used very carefully due to large errors in predicting C_{80}.

Further studies should be carried out in order to assess the possibility of extending model usage to different kinds of disproportionate rooms for which the revised theory cannot be reliably applied.

[1]H. Kuttruff, *Room Acoustics*, 3rd ed. (E & FN Spon, London, 1991).
[2]L. Cremer and H. A. Muller, *Principles and Applications of Room Acoustics* (Applied Science, London, 1982), Vol. 1.
[3]M. Barron and L. J. Lee, "Energy relations in concert auditoriums. I," J. Acoust. Soc. Am. **84**, 618–628 (1988).
[4]S. Chiles and M. Barron, "Sound level distribution and scatter in proportionate spaces," J. Acoust. Soc. Am. **116**, 1585–1595 (2004).
[5]M. Vörlander, "Revised relation between the sound power and the average sound pressure level in rooms and consequences for acoustic measurements," Acustica **81**, 332–343 (1995).
[6]J. J. Sendra, T. Zamarreño, and J. Navarro, "Acoustics in churches," in *Computational Acoustics in Architecture*, edited by J. J. Sendra (Computational Mechanics, Southampton, 1999), pp. 133–177.
[7]M. Galindo, T. Zamarreño, and S. Giron, "Clarity and definition in Mudejar-Gothic churches," Build Acoust. **6**, 1–16 (1999).
[8]M. Galindo, T. Zamarreño, and S. Giron, "Acoustic analysis in Mudejar-Gothic churches: Experimental results," J. Acoust. Soc. Am. **117**, 2873–2888 (2005).
[9]N. Prodi, M. Marsilio, and R. Pompoli, "On the prediction of reverberation time and strength in mosques," in Proceedings of the 17th ICA, Rome (2001).
[10]E. Cirillo and F. Martellotta, "An improved model to predict energy-based acoustic parameters in Apulian-Romanesque churches," Appl. Acoust. **64**, 1–23 (2003).
[11]A. P. O. Carvalho, "Relationship between objective acoustical measures and architectural features in churches," in Proceedings of the W.C. Sabine Centennial Symposium, 127th Acoustical Society of America Meeting, Cambridge, MA (1994), pp. 311–314.
[12]T. Zamarreño, S. Giron, and M. Galindo, "Acoustic energy relations in Mudejar-Gothic churches," J. Acoust. Soc. Am. **121**, 234–250 (2007).
[13]E. Cirillo and F. Martellotta, "Sound propagation and energy relations in churches," J. Acoust. Soc. Am. **118**, 232–248 (2005).
[14]F. Martellotta, "A multi-rate decay model to predict energy-based acoustic parameters in churches," J. Acoust. Soc. Am. **125**, 1281–1284 (2009).
[15]U. Berardi, E. Cirillo, and F. Martellotta, "Predicting energy-based acoustic parameters in churches: An attempt to generalize the μ -model," in Proceedings of the 39th Acústica, Coimbra (2008).
[16]S. Müller and P. Massarani, "Transfer-function measurement with sweeps," J. Audio Eng. Soc. **49**, 443–471 (2001).
[17]ISO-3382, "Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters," ISO, Geneva, Switzerland, 1997.
[18]E. Cirillo and F. Martellotta, *Worship, Acoustics, and Architecture* (Multiscience, Brentwood, UK, 2006).

Effect of boundary slip on the acoustical properties of microfibrous materials[a)]

Olga Umnova[b)]
Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom

David Tsiklauri
Institute for Material Research, University of Salford, Salford M5 4WT, United Kingdom

Rodolfo Venegas
Acoustics Research Centre, University of Salford, Salford M5 4WT, United Kingdom

(Received 21 November 2008; revised 21 July 2009; accepted 22 July 2009)

A variety of new porous materials with unusually small pores have been manufactured in the past decades. To predict their acoustical properties, the conventional models need to be modified. When pore size becomes comparable to the molecular mean free path of a saturating fluid, the no-slip conditions on the pore surface are no longer accurate and hence the slip effects have to be taken into account. In this paper, sound propagation in microfibrous materials is modeled analytically, approximating the geometry by a regular array of rigid parallel cylinders. It has been shown that velocity and thermal slip on a cylinder surface significantly changes the model predictions leading to lower attenuation coefficient and higher sound speed values. The influence of material porosity, fiber orientation, and size on these effects is investigated. Finite element method is used to numerically solve the oscillatory flow and heat transfer problems in a square array of cylindrical fibres. Numerical results are compared with predictions of the analytical model and the range of its validity is identified. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3204087]

PACS number(s): 43.55.Ev, 43.20.Hq, 43.20.Bi [LMW] Pages: 1850–1861

I. INTRODUCTION

In many models of porous sound absorbents[1–5] the local flow and temperature distribution are studied first. The macroscopic quantities are estimated after that by volume averaging. The theoretical basis for these models is provided by homogenization theory,[6–8] which assumes the existence of a representative volume and separation of scales. In "internal flow" models, the flow and temperature characteristics are calculated inside a pore of a given shape.[9–12] In "external flow" models, they are calculated around solid inclusions.[13–15] External flow models are more suitable for description of granular and fibrous materials where the inclusion shape is easier to define than that of the pore.

Until recently, the validity of zero velocity and fixed temperature boundary conditions on the pore or inclusion surface was assumed in most models. The dynamic tortuosity, which accounts for viscous and inertial effects, has been calculated from the solution of an oscillatory forced Stokes problem considering no-slip boundary condition at the solid-fluid interface by using the dynamic extension of Darcy's law.[8,16] The dynamic compressibility, describing thermal losses, has been obtained from the solution of an oscillatory heat transfer problem with fixed temperature boundary condition at the solid-fluid interface.[17] The domain of validity of fixed temperature boundary condition is discussed in Ref. 17.

The zero velocity and fixed temperature boundary conditions are violated if a characteristic pore or inclusion size a becomes comparable to a saturating gas molecular mean free path l_{mean}. Their ratio is described by a parameter called Knudsen number,[18] $K=l_{\mathrm{mean}}/a$. The molecular mean free path, $l_{\mathrm{mean}}=2\eta/\rho_0\sqrt{8R_gT_0/\pi}$, depends on viscosity η, equilibrium density ρ_0, temperature T_0, and the specific constant R_g of the gas.[12,18] For air at normal conditions $l_{\mathrm{mean}} \approx 60$ nm.

In the majority of porous materials currently used in acoustical applications, pores or inclusions are in excess of several micrometers. However, a substantial number of new materials are being created with much smaller pores or inclusions. For example, in aerogels and in activated carbons, pores can be in the order of 20 nm in diameter or even smaller.[19,20] The propagation of the slow compressional wave in a low density aerogel has been observed in previous studies (see Ref. 21, for example). Therefore, to model acoustical properties of microporous materials, the conventional models have to be modified to allow for velocity and thermal slip.[12] The term "thermal slip" is used here to describe the temperature jump on the wall, which is proportional to the temperature normal derivative.

It is worth mentioning that the continuum description of a gas gives physically justified results only in the so-called Knudsen regime.[22,23] This implies that $K \leq 0.1$ and imposes a lower limit on the pore or inclusion size, e.g., $a \geq 10 l_{\mathrm{mean}}$. For materials saturated by air at normal conditions, this means $a \geq 0.6\ \mu$m. However, comparisons with both measurements[12] and molecular dynamics simulations[24] show that the continuum approach still provides a good approximation when $K > 0.1$. Satisfactory agreement between data

[a)]Portions of this paper have been presented at the Acoustics-08 Conference, Paris, France, June 29–July 4, 2008.

[b)]Author to whom correspondence should be addressed. Electronic mail: o.umnova@salford.ac.uk

 0001-4966/2009/126(4)/1850/12/$25.00

and theory for rarefied gas flow through straight cylindrical tubes has been demonstrated even for Knudsen number values higher than one.[12]

The effect of boundary slip on sound propagation in porous materials has been investigated in several recent publications. The homogenization theory has been applied to account for slip effects in both static and dynamic problems.[25,26] Rigid porous materials have been studied assuming the Maxwell boundary condition on the pore walls which relates the tangential surface velocity to the stress component.[12,24] The frame elasticity has been accounted for by assuming either phenomenological frequency dependence of the velocity on the pore wall[27] or Maxwell boundary conditions.[28] In Ref. 12 the influence of thermal slip on complex compressibility has been investigated. In these publications the pore geometry was restricted to straight cylindrical channels.

However, an assumption of straight cylindrical pores does not capture the complexity of the inner structure of fibrous materials. For instance, it cannot be used to describe acoustical properties dependence on fiber orientation. In this paper, fibrous materials are modeled as regular arrays of straight circular cylinders.

This paper is organized as follows. In Sec. II, the oscillatory forced problem of fluid flow around cylindrical fibres is solved analytically. The method is based on the cell model approach described in Refs. 29, 2, and 4. The problem of the oscillatory heat transfer between the fiber and the surrounding gas has been solved accounting for the thermal slip. Analytical expressions for complex tortuosity and complex compressibility are derived for fibers parallel and perpendicular to sound propagation direction. In Sec. III, analytical expressions for high frequency tortuosity, flow resistivity, characteristic viscous and thermal lengths and thermal permeability in the presence of the boundary slip are derived. The effects of boundary slip on sound speed and attenuation coefficient are also estimated. Section IV describes numerical solutions performed using finite element method (FEM). Comparisons with numerical results help to investigate the restrictions of the analytical model in the absence of experimental data. Main findings are summarized in Sec. V.

II. COMPLEX TORTUOSITY AND COMPLEX COMPRESSIBILITY FUNCTIONS

A. Complex tortuosity

1. Fibers perpendicular to sound propagation direction

A detailed analysis of the model assuming no-slip boundary conditions can be found elsewhere.[2,4] In this subsection a brief outline of the equations and their solutions along with the necessary boundary conditions modifications is presented.

According to homogenization theory,[6–8,25,26] the fluid can be considered as incompressible at the pore/inclusion scale. The motion of the fluid around an infinitely long cylindrical fiber of radius a under oscillatory constant small-amplitude pressure gradient $(\partial P/\partial x)e^{-i\omega t}\vec{e}_x$ is governed by the following set of equations:

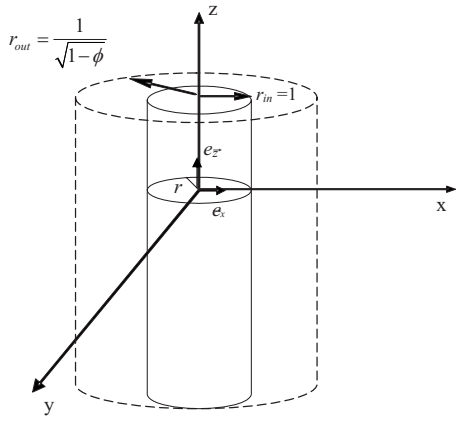


FIG. 1. Cell geometry for the analytical solution.

$$-i\omega\rho_0\vec{u} = -\frac{1}{a}\overrightarrow{\text{grad}}\, p + \frac{\eta}{a^2}\Delta\vec{u} - \frac{\partial P}{\partial x}\vec{e}_x, \tag{1a}$$

$$\text{div}\,\vec{u} = 0, \tag{1b}$$

where $\vec{u}$ and p are velocity and pressure of the fluid, ω is angular frequency, and η and ρ_0 are gas viscosity and equilibrium density, respectively. An arrow denotes a vector quantity. Throughout this paper all spatial coordinates are normalized by the fiber radius a. The geometry of the problem is illustrated in Fig. 1. Because of fluid incompressibility, one can assume that $\vec{u}=\vec{v}+\text{curl}\,\vec{A}$, where $\vec{v}=v\vec{e}_x$, $v=(\partial P/\partial x)(1/i\omega\rho_0)$ and $\vec{A}$ is a vector potential. Due to symmetry, in cylindrical coordinates (r,θ,z) velocity does not have a z-component and does not vary along the z-axis. This implies that the vector potential has only an A_z component with both A_r and A_θ being zero. Velocity components are $u_r=v\cos\theta+(1/r)(\partial A_z/\partial\theta)$ and $u_\theta=-v\sin\theta-\partial A_z/\partial r$ and assuming $A_z=A(r)\sin\theta$, they can be simplified as

$$u_r = \left(v + \frac{A(r)}{r}\right)\cos\theta, \tag{2a}$$

$$u_\theta = -(v + A'(r))\sin\theta, \tag{2b}$$

where the prime symbol is used to denote the derivative.

Vorticity of velocity $\vec{u}$ can now be calculated as

$$\text{curl}\,\vec{u} = \left(\frac{u_\theta}{r} + \frac{\partial u_\theta}{\partial r} - \frac{1}{r}\frac{\partial u_r}{\partial\theta}\right)\vec{e}_z = f(r)\sin\theta\vec{e}_z,$$

where

$$f(r) = -\left(A'(r) + \frac{A(r)}{r}\right)'. \tag{3}$$

For the incompressible fluid the following identity is valid:

$$\Delta\vec{u} = -\,\text{curlcurl}\,\vec{u} = -\,\text{curl}(f(r)\sin\theta\vec{e}_z)$$

$$= -\frac{f(r)}{r}\cos\theta\vec{e}_r + f'(r)\sin\theta\vec{e}_\theta. \tag{4}$$

Now Eq. (1a) has to be rewritten separately for two velocity components:

$$-i\omega\rho_0\frac{A(r)}{r}\cos\theta = -\frac{\eta}{a^2}\frac{f(r)\cos\theta}{r} - \frac{\partial p}{\partial r}\frac{1}{a}, \tag{5a}$$

$$i\omega\rho_0 A'(r)\sin\theta = \frac{\eta}{a^2} f'(r)\sin\theta - \frac{\partial p}{\partial\theta}\frac{1}{ra}, \tag{5b}$$

and can be solved for p to give

$$p = \frac{\eta}{a^2}(\kappa^2 A'(r) - f'(r))ra\cos\theta, \tag{6}$$

where

$$\kappa = a\sqrt{\frac{i\omega\rho_0}{\eta}}. \tag{7}$$

The modulus of κ is equal to the ratio between the fiber radius and the viscous boundary layer thickness:

$$\delta_{\text{visc}} = \sqrt{\frac{2\eta}{\omega\rho_0}} \quad \text{i.e.,} \quad |\kappa| = \frac{a}{\delta_{\text{visc}}}. \tag{8}$$

Equations (1a) and (1b) can be reduced to a single equation, $\kappa^2 curl\,\vec{u} = curlcurl(curl\,\vec{u})$, which is equivalent to

$$f''(r) - \frac{f'(r)}{r} - \frac{f}{r^2} + \kappa^2 f = 0. \tag{9}$$

Assuming $y=\kappa r$, this can be transformed to a standard Bessel equation. Its solution can be expressed as[30]

$$f(r) = -\alpha\kappa H_1^{(1)}(\kappa r) - \beta\kappa J_1(\kappa r), \tag{10}$$

where α and β are arbitrary constants.

The solution for $A(r)$ can now be found from Eq. (3):

$$A(r) = -\gamma\frac{r}{2} + \frac{\delta}{r} - \alpha\frac{H_1^{(1)}(\kappa r)}{\kappa} - \beta\frac{J_1(\kappa r)}{\kappa}, \tag{11}$$

where γ and δ is another pair of arbitrary constants. Using Eqs. (2a), (2b), (6), (10), and (11), velocity components and pressure can be rewritten as

$$u_r = \left(v - \frac{\gamma}{2} + \frac{\delta}{r^2} - \alpha\frac{H_1^{(1)}(\kappa r)}{\kappa r} - \beta\frac{J_1(\kappa r)}{\kappa r}\right)\cos\theta, \tag{12a}$$

$$u_\theta = -\left(v - \frac{\gamma}{2} - \frac{\delta}{r^2} - \alpha\left(H_0^{(1)}(\kappa r) - \frac{H_1^{(1)}(\kappa r)}{\kappa r}\right) - \beta\left(J_0(\kappa r) - \frac{J_1(\kappa r)}{\kappa r}\right)\right)\sin\theta, \tag{12b}$$

$$p = -\frac{\eta}{a^2}\kappa^2\left(\frac{\gamma}{2} + \frac{\delta}{r^2}\right)ra\cos\theta. \tag{13}$$

Constants α, β, γ, and δ must be found from the boundary conditions.

According to the cell model approach, the influence of the neighboring cylinders on the flow can be taken into account assuming that each cylinder is placed at the center of an imaginary cylindrical cell (Fig. 1). The outer radius of the cell is chosen so that the volume fraction of air in a single cell matches the material porosity ϕ, i.e., $r_{\text{out}} = 1/\sqrt{1-\phi}$. Conditions on the cell boundary are set to allow for the interactions between the cylinders. The assumption of zero vorticity on the cell boundary r_{out} was first used by Kuwabara in Ref. 29 for a steady flow and later generalized for the case of an oscillatory flow.[2,5] The validity of this condition has recently been rigorously investigated[31] in an application to granular materials. It has been found that although zero vorticity assumption does not follow from any first principles, it gives results very close to those obtained using energy consistent boundary conditions. Further investigation of the zero vorticity assumption will be presented in Sec. IV. Using Eq. (10) vorticity can be expressed as $\text{curl}\,\vec{u} = (-\alpha\kappa H_1^{(1)}(\kappa r) - \beta\kappa J_1(\kappa r))\sin\theta\vec{e}_z$. The following equation for constants α and β can be derived if vorticity is set equal to zero on the cell boundary:

$$\alpha H_1^{(1)}\left(\frac{\kappa}{\sqrt{1-\phi}}\right) + \beta J_1\left(\frac{\kappa}{\sqrt{1-\phi}}\right) = 0. \tag{14}$$

It is also assumed that the pressure drop over the cell is equal to the macroscopic pressure gradient.[2] This condition is equivalent to $p(r=1/\sqrt{1-\phi},\theta)=0$. Using Eq. (13) for pressure, this can be reformulated as

$$\frac{\gamma}{2} + \delta(1-\phi) = 0. \tag{15}$$

The normal component of the fluid velocity is zero on the fiber surface, which in combination with Eq. (12a) gives

$$\frac{\gamma}{2} - \delta + \alpha\frac{H_1^{(1)}(\kappa)}{\kappa} + \beta\frac{J_1(\kappa)}{\kappa} = v. \tag{16}$$

The condition of zero tangential velocity on the fiber surface is replaced by the Maxwell boundary condition:

$$u_\theta(r=1) = \frac{l_{\text{mean}}}{\eta}\tau_{r\theta}(r=1), \tag{17}$$

where $\tau_{r\theta} = \eta/a((1/r)(\partial u_r/\partial\theta) + (\partial u_\theta/\partial r) - (u_\theta/r))$ is the component of the stress tensor. For simplicity, the momentum accommodation coefficient is assumed to be equal to 1, which means that fiber surface reflects incident molecules diffusively.[22] Using Eqs. (12a) and (12b) Maxwell boundary condition (17) can be reduced to

$$\frac{\gamma}{2} + \delta(1+4K) + \alpha\left(H_0^{(1)}(\kappa)(1+2K) - \frac{H_1^{(1)}(\kappa)}{\kappa}(1+4K - K\kappa^2)\right) + \beta\left(J_0(\kappa)(1+2K) - \frac{J_1(\kappa)}{\kappa}(1+4K-K\kappa^2)\right) = v, \tag{18}$$

where $K = l_{\text{mean}}/a$ is Knudsen number.

In derivation of Eq. (18), the following identities for Bessel and Hankel functions have been used:

$$J_1'(x) = J_0(x) - J_1(x)/x \quad \text{and} \quad J_0'(x) = -J_1(x). \tag{19}$$

Equations (14)–(16) and (18) constitute a set of linear algebraic equations for the unknown constants, which can be solved to get

$$\alpha = \frac{2v(1-\phi)}{Z_0(\kappa)(2-\phi) - Z_1(\kappa)\left(2(1-\phi) - (2-\phi)\dfrac{K\kappa^2}{(1+2K)}\right)}, \tag{20a}$$

$$\beta = -\frac{\alpha H_1^{(1)}\left(\dfrac{k}{\sqrt{1-\phi}}\right)}{J_1\left(\dfrac{k}{\sqrt{1-\phi}}\right)}, \tag{20b}$$

$$\gamma = v \times \frac{2(1-\phi)\left(Z_0(\kappa) - Z_1(\kappa)\left(2 - \dfrac{K\kappa^2}{(1+2K)}\right)\right)}{Z_0(\kappa)(2-\phi) - Z_1(\kappa)\left(2(1-\phi) - (2-\phi)\dfrac{K\kappa^2}{(1+2K)}\right)}, \tag{20c}$$

$$\delta = -\frac{\gamma}{2(1-\phi)}, \tag{20d}$$

where

$$Z_i(\kappa) = \frac{1}{\kappa^i}\left(H_i^{(1)}(\kappa) - \frac{H_1^{(1)}\left(\dfrac{\kappa}{\sqrt{1-\phi}}\right)}{J_1\left(\dfrac{\kappa}{\sqrt{1-\phi}}\right)} J_i(\kappa)\right), \quad i = 0,1. \tag{21}$$

Complex tortuosity $\alpha_\perp(\omega)$ is defined as the ratio between the macroscopic pressure gradient $-\partial P/\partial x$ and the resulting averaged force per unit volume of the fluid $\langle F_x\rangle = -i\omega\rho_0\langle u_x\rangle$

$$\alpha_\perp(\omega) = \frac{\partial P}{\partial x}\frac{1}{i\omega\rho_0\langle u_x\rangle} = \frac{v}{\langle u_x\rangle}, \tag{22}$$

where subscript $\perp$ indicates that fibers are perpendicular to sound propagation direction. Velocity averaging can be performed over a single cell surface, as has been shown in Ref. 2:

$$\langle u_x\rangle = \frac{1}{S_{\text{cell}}}\int_0^{2\pi}\int_1^{1/\sqrt{1-\phi}} u_x(r,\theta) r dr d\theta, \tag{23}$$

where $S_{\text{cell}} = \pi[\phi/(1-\phi)]$ is the normalized cross sectional area of the cell. Using the fact that $u_x = u_r\cos\theta - u_\theta\sin\theta$ together with Eqs. (2a) and (2b), the x-component of velocity can be expressed as $u_x = v + A(r)/r\cos^2\theta - A'(r)\sin^2\theta$. Averaging over the cell surface gives

$$\langle u_x\rangle = = v + \frac{1-\phi}{\phi}\left(A\left(\frac{1}{\sqrt{1-\phi}}\right)\frac{1}{\sqrt{1-\phi}} - A(1)\right). \tag{24}$$

Using Eq. (11), $A(1)$ and $A(1/\sqrt{1-\phi})$ are expressed in terms of constants determined by Eqs. (20a)–(20d). This results in the following expressions for the averaged velocity:

$$\langle u_x\rangle = \frac{v}{\phi}\frac{\dfrac{Z_0(\kappa)}{Z_1(\kappa)}\phi + \left(2(1-\phi) + \dfrac{K\kappa^2}{(1+2K)}\phi\right)}{\dfrac{Z_0(\kappa)}{Z_1(\kappa)}(2-\phi) - \left(2(1-\phi) - \dfrac{K\kappa^2}{(1+2K)}(2-\phi)\right)} \tag{25}$$

and, consequently, complex tortuosity

$$\alpha_\perp(\omega,K) = (2-\phi) - \frac{4(1-\phi)}{\dfrac{Z_0(\kappa)}{Z_1(\kappa)}\phi + 2(1-\phi) + \dfrac{K\kappa^2}{(1+2K)}\phi}. \tag{26}$$

Equation (26) presents complex tortuosity as a function of angular frequency and Knudsen number. This is convenient for both comparisons with no-slip results and the scaling, which will be applied later to calculate complex compressibility. In the no-slip case ($K=0$) Eq. (26) coincides with that presented in Ref. 4.

2. Fibers parallel to sound propagation direction

In this case the oscillatory pressure gradient $(\partial P/\partial z)e^{-i\omega t}\vec{e}_z$ is applied along the fiber axes and $u_z(r)$ is the only non-zero velocity component in an incompressible fluid. Due to these equations of motion can be replaced by one simpler equation:

$$-i\omega\rho_0 u_z = \frac{\eta}{a^2}\left(u_z'' + \frac{1}{r}u_z'\right) - \frac{\partial P}{\partial z}. \tag{27}$$

Its general solution is[31]

$$u_z(r) = \frac{1}{i\omega\rho_0}\frac{\partial P}{\partial z} + AJ_0(\kappa r) + BH_0^{(1)}(\kappa r). \tag{28}$$

Constants A and B can be found from Maxwell boundary condition on the fiber surface, which in this case is expressed as

$$u_z(r=1) = Ku_z'(r=1), \tag{29}$$

and boundary condition of zero vorticity on the cell boundary:

$$u_z\left(r = \frac{1}{\sqrt{1-\phi}}\right) = 0. \tag{30}$$

Here cells are defined as in Part I. This gives the following pair of equations for constants A and B:

$$A(J_0(\kappa) + \kappa K J_1(\kappa)) + B(H_0^{(1)}(\kappa) + \kappa K H_1^{(1)}(\kappa)) = -\frac{1}{i\omega\rho_0}\frac{\partial P}{\partial z}, \tag{31a}$$

$$AJ_1\left(\frac{\kappa}{\sqrt{1-\phi}}\right) + BH_1^{(1)}\left(\frac{\kappa}{\sqrt{1-\phi}}\right) = 0. \tag{31b}$$

Thus, the velocity averaged over the cell surface can be calculated as

$$\langle u_z\rangle = \frac{1}{i\omega\rho_0}\frac{\partial P}{\partial z} + \left(\frac{1}{\kappa^2}\right)\frac{2(1-\phi)}{\phi}\left[A\left(J_1\left(\frac{\kappa}{\sqrt{1-\phi}}\right)\frac{\kappa}{\sqrt{1-\phi}} - J_1(\kappa)\kappa\right) + B\left(H_1{}^{(1)}\left(\frac{\kappa}{\sqrt{1-\phi}}\right)\frac{\kappa}{\sqrt{1-\phi}} - H_1{}^{(1)}(\kappa)\kappa\right)\right]. \tag{32}$$

Solving equations for A and B and substituting the results into this expression, the averaged velocity is given by

$$\langle u_z\rangle = \frac{1}{i\omega\rho_0}\frac{\partial P}{\partial z}\left(1 + \frac{2(1-\phi)Z_1(\kappa)}{\phi(Z_0(\kappa) + K\kappa^2 Z_1(\kappa))}\right). \tag{33}$$

Now an expression for complex tortuosity $\alpha_\parallel(\omega) = -(\partial P/\partial z) \times (1/-i\omega\rho_0\langle u_z\rangle)$ can be derived as

$$\alpha_\parallel(\omega,K) = 1 - \frac{2(1-\phi)}{\frac{Z_0(\phi)}{Z_1(\kappa)}\phi + 2(1-\phi) + \phi K\kappa^2}, \tag{34}$$

where subscript $\parallel$ indicates sound propagation along the fiber axes.

B. Complex compressibility

Consider the heat transfer from a fiber into the surrounding fluid subject to an oscillating pressure $Pe^{-i\omega t}$. As shown in Ref. 12 distribution of temperature $T(r)$ is not dependent on the sound propagation direction and can be found from the following equation:

$$N_{\mathrm{Pr}}\kappa^2 T + \left(T'' + \frac{1}{r}T'\right) = \frac{N_{\mathrm{Pr}}\kappa^2}{\rho_0 c_p}P, \tag{35}$$

where $N_{\mathrm{Pr}} = \eta c_P/q$ is Prandtl number, q is thermal conduction coefficient of the fluid, and c_p is heat capacity under constant pressure.

Thermal slip boundary condition on the fiber surface is[12]

$$T(r=1) = \frac{2\gamma}{(\gamma+1)N_{\mathrm{Pr}}}KT'(r=1), \tag{36}$$

where γ is the adiabatic constant and the accommodation coefficient is assumed to be equal to 1. Zero temperature flux is assumed on the outer cell boundary[3]

$$T'\left(r = \frac{1}{\sqrt{1-\phi}}\right) = 0. \tag{37}$$

Comparison of Eqs. (35)–(37) with Eqs. (27), (29), and (30) suggests that

$$\frac{T(\omega,K)}{\frac{N_{\mathrm{Pr}}\kappa^2}{\rho_0 c_p}P} = \frac{u_z\left(\omega N_{\mathrm{Pr}}, K\frac{2\gamma}{N_{\mathrm{Pr}}(\gamma+1)}\right)}{\frac{a^2}{\eta}\frac{\partial P}{\partial z}}. \tag{38}$$

It should be noted that the scaling here is applied to both frequency and Knudsen number. The normalized complex compressibility is defined in Ref. 17 as

$$C(\omega) = \gamma - (\gamma-1)\frac{\rho_0 c_p\langle T\rangle}{P}. \tag{39}$$

Here averaging $\langle T\rangle$ can be performed over a single cell surface.

Now the relationship between complex compressibility and complex tortuosity $\alpha_\parallel$, can be derived as

$$C(\omega,K) = \gamma - \frac{\gamma-1}{\alpha_\parallel\left(\omega N_{\mathrm{Pr}}, K\frac{2\gamma}{N_{\mathrm{Pr}}(\gamma+1)}\right)}, \tag{40}$$

where $\alpha_\parallel(\omega N_{\mathrm{Pr}}, K2\gamma/N_{\mathrm{Pr}}(\gamma+1))$ is described by Eq. (34). This result is equivalent to that derived in Ref. 12 for straight cylindrical pores.

C. Characteristic impedance and propagation constant

Expressions for complex tortuosity and complex compressibility can be further used for the description of macroscopic material properties. The momentum conservation and continuity equations for plane acoustic waves propagating in a porous material are

$$-i\omega\rho_0\alpha(\omega,K)V = -ik(\omega,K)P, \tag{41a}$$

$$-i\omega C(\omega,K)P = ik(\omega,K)\rho_0 cV, \tag{41b}$$

where c is sound speed in air, P and V are acoustic pressure and fluid particle velocity amplitudes, respectively, and $k(\omega,K)$ is the propagation constant expressed here as a function of angular frequency and Knudsen number. These equations can be combined with Eqs. (26), (34), and (40) to give expressions for characteristic impedances

$$z_{\parallel,\perp}(\omega,K) = \frac{\rho_0 c}{\phi}\sqrt{\frac{\alpha_{\parallel,\perp}(\omega,K)}{C(\omega,K)}}$$

and propagation constants $k_{\parallel,\perp}(\omega,K) = \omega/c\sqrt{\alpha_{\parallel,\perp}(\omega,K)C(\omega,K)}$ for two different fiber orientations.

III. LOW AND HIGH FREQUENCY ASYMPTOTICS

A. Asymptotics of complex tortuosity

Viscous friction between the fluid and the fibers has a strong influence on the acoustic properties of the material at low frequencies. In this case boundary layer thickness exceeds the fiber radius, which requires $|\kappa| \ll 1$. This leads to the following limitation on frequency f:

$$f \ll \frac{\eta}{\pi\rho_0 a^2}. \tag{42}$$

Low frequency asymptotics for complex tortuosity can be found by retaining the leading terms in the expansions of Bessel and Hankel functions for small arguments given in Ref. 30. For complex tortuosity $\alpha_\parallel(\omega,K)$, this becomes $\alpha_\parallel(\omega\to 0,K) = -\phi/\kappa^2 8(1-\phi)/(-2\ln(1-\phi) - 2\phi - \phi^2 + 4K\phi^2)$. It is shown in Ref. 16 that the low frequency limit of complex tortuosity is related to steady state flow resistivity σ as $\alpha(\omega\to 0) = -(\phi/\kappa^2)(a^2\sigma/\eta)$. Consequently, steady state flow resistivity $\sigma_\parallel$ in direction parallel to fiber axes is described as

$$\sigma_{\parallel}(K) = \frac{\sigma_{\parallel}(K=0)}{1+4KF(\phi)}, \tag{43}$$

where it is shown as a function of the Knudsen number. Here

$$\sigma_{\parallel}(K=0) = \frac{8(1-\phi)\eta}{(-2\ln(1-\phi)-2\phi-\phi^2)a^2}, \tag{44}$$

$$F(\phi) = \frac{\phi^2}{-2\ln(1-\phi)-2\phi-\phi^2}. \tag{45}$$

Equation (42) confirms the well-known fact that flow resistivity is a decreasing function of Knudsen number.[32] The expression for $\sigma_{\parallel}(K=0)$ coincides with that of Ref. 4 which was derived assuming no slip boundary conditions.

Similar expansions applied to Eq. (26) lead to the expression for flow resistivity $\sigma_{\perp}$

$$\sigma_{\perp}(K) = \frac{\sigma_{\perp}(K=0)}{1+\dfrac{4K}{1+2K}F(\phi)}, \tag{46}$$

where $\sigma_{\perp}(K=0)=2\sigma_{\parallel}(K=0)$ in agreement with Ref. 4.

Low frequency expansion of complex tortuosity derived in Ref. 12 yields the following dependence of steady state flow resistivity σ_{cyl} of cylindrical pore network on Knudsen number:

$$\sigma_{\mathrm{cyl}}(K) = \frac{\sigma_{\mathrm{cyl}}(K=0)}{1+4K}, \tag{47}$$

where $\sigma_{\mathrm{cyl}}(K=0)=8\eta/\phi a^2$ is in agreement with results presented in Refs. 33 and 34. A degree of similarity between Eqs. (42) and (46) can be immediately recognized. However, for the array of cylinders, boundary slip effect strongly depends on porosity. This is an expected result as the effective size of the sound propagation "channels" is determined by the distance between the cylinders and consequently by material porosity. When porosity is high, these channels get bigger which leads to weaker slip effect.

In the case of perfect slip ($K\to\infty$), both $\sigma_{\parallel}$ and σ_{cyl} are equal to zero. However, as can be noted from Eq. (45), $\sigma_{\perp}$ does not vanish. Although the shear stress is always zero on a perfectly slippery surface, additional pressure variation in the fluid is induced by the flow perpendicular to fibers due to the curvature of the flow paths. This gives a contribution to the viscous drag.

In the limit of high frequencies, the effect of fluid inertia becomes dominant. In this range $|\kappa|\gg 1$, which leads to the condition on frequency,

$$f \gg \frac{\eta}{\pi\rho_0 a^2} = \frac{\eta}{\pi\rho_0 l_{\mathrm{mean}}^2}K^2 = K^2\times 1.34\times 10^9 \text{ Hz}. \tag{48}$$

It should be noted that all previous calculations have been performed assuming the absence of scattering effects which requires

$$f \ll \frac{c}{2\pi a} = \frac{c}{2pl_{\mathrm{mean}}}K = K\times 0.91\times 10^9 \text{ Hz}. \tag{49}$$

Equations (48) and (49) can only be satisfied simultaneously for $K\ll 0.1$. In practice, this means that for air at normal conditions inertial regime can be attained only if fibers are very large compared to the molecular mean free path and hence slip effects are not important. For smaller fibers, both viscous and inertial effects influence sound propagation for most of the ultrasonic frequencies, except exotically high ones. For the latter, however, frequency limitation (49) will be violated and scattering has to be accounted for.

Nonetheless, for the sake of completeness, the high frequency drag parameters will be estimated here. Asymptotic expansions for Bessel and Hankel functions of large arguments (9.2.5) and (9.2.7) from Ref. 30 lead to the following approximation:

$$\frac{Z_0(\kappa\to\infty)}{Z_1(\kappa\to\infty)} = i\kappa + \frac{1}{2} + \frac{1}{i\kappa}\frac{3}{32} + O\left(\frac{1}{|\kappa|^2}\right). \tag{50}$$

Its substitution into Eqs. (34) and (26) gives

$$\alpha_{\parallel}(\omega\to\infty,K) = 1 + \frac{i}{\kappa}\frac{2(1-\phi)}{\phi} \times \frac{1}{1-\dfrac{i}{\kappa}\dfrac{(4-3\phi)}{2\phi} - iK\kappa + \dfrac{3}{32}\left(\dfrac{i}{\kappa}\right)^2 + O\left(\dfrac{1}{|\kappa|^3}\right)}, \tag{51a}$$

$$\alpha_{\perp}(\omega\to\infty,K) = (2-\phi) + \frac{i}{\kappa}\frac{4(1-\phi)}{\phi} \times \frac{1}{1-\dfrac{i}{\kappa}\dfrac{(4-3\phi)}{2\phi} - \dfrac{iK\kappa}{1+2K} + \dfrac{3}{32}\left(\dfrac{i}{\kappa}\right)^2 + O\left(\dfrac{1}{|\kappa|^3}\right)}. \tag{51b}$$

Further expansion in powers of $1/\kappa$ depends on the value of $K|\kappa|$, which is equal to the ratio between the molecular mean free path and the boundary layer thickness. A small value of K does not automatically mean that this parameter is small, as at high frequencies $|\kappa|$ is large. However, as pointed out in Ref. 12, for air saturation at normal conditions large values of parameter $K|\kappa|$ can be achieved only for frequencies well above 100 MHz. Moreover, the continuum approach is only applicable when all characteristic lengths of the problem are large compared to the molecular mean free path. Viscous boundary layer can be considered as one of these lengths, which means that the current analysis cannot be extended to frequencies, where $K|\kappa|\geq 1$. Condition $K|\kappa|\ll 1/|\kappa|$ is satisfied when $f\ll K\times 1.34\times 10^9$ Hz and in the absence of scattering. Comparison with Eq. (49) confirms that this condition is valid in a scattering-free range of frequencies. Assuming that $K|\kappa|=O(1/|\kappa|^2)$ further expansion of Eqs. (51a) and (51b) gives

$$\alpha_{\parallel}(\omega\to\infty,K)=1-\frac{2(1-\phi)}{\phi}K+\frac{i}{\kappa}\frac{2(1-\phi)}{\phi}+\left(\frac{i}{\kappa}\right)^2\frac{(1-\phi)(4-3\phi)}{\phi^2}+\left(\frac{i}{\kappa}\right)^3\frac{2(1-\phi)}{\phi}\left(\left(\frac{4-3\phi}{2\phi}\right)^2-\frac{3}{32}\right)+O\left(\frac{1}{|\kappa|^4}\right), \tag{52a}$$

$$\alpha_{\perp}(\omega\to\infty,K)=2-\phi-\frac{4(1-\phi)}{\phi}\frac{K}{1+2K}+\frac{i}{\kappa}\frac{4(1-\phi)}{\phi}+\left(\frac{i}{\kappa}\right)^2\frac{2(1-\phi)(4-3\phi)}{\phi^2}+\left(\frac{i}{\kappa}\right)^3\frac{4(1-\phi)}{\phi}\left(\left(\frac{4-3\phi}{2\phi}\right)^2-\frac{3}{32}\right)+O\left(\frac{1}{|\kappa|^4}\right). \tag{52b}$$

According to Ref. 16, high frequency limit of dynamic tortuosity is related to viscous characteristic length Λ and tortuosity α_∞ as $\alpha(\omega\to\infty)=\alpha_\infty+(i/\kappa)(2\alpha_\infty/\Lambda)$. If α_∞ and Λ are to be derived from the expansions (52a) and (52b), only linear terms with respect to $1/\kappa$ should be taken into account. Although $[2(1-\phi)/\phi]K$ and $[4(1-\phi)/\phi][K/(1+2K)]$ are frequency independent, they are of the same order of magnitude as cubic terms in the expansions and for this reason should be neglected in the derivations of Λ and α_∞. Simple expressions can now be derived for high frequency tortuosities as follows:

$$\alpha_{\infty\parallel}=1, \tag{53}$$

$$\alpha_{\infty\perp}=2-\phi, \tag{54}$$

and the normalized characteristic viscous lengths,

$$\Lambda_{\parallel}=\frac{\phi}{1-\phi}, \tag{55}$$

$$\Lambda_{\perp}=\frac{(2-\phi)\phi}{2(1-\phi)}, \tag{56}$$

which for both fiber orientations are independent of Knudsen number.

It should be highlighted that expressions (53)–(56) are only valid for very low values of Knudsen number. For higher values, inertial regime can no longer be considered separately from scattering effects.

B. Asymptotics of complex compressibility

The easiest way to calculate both low and high frequency approximations of the complex compressibility is to use Eq. (40) together with asymptotic expansions for $\alpha_{\parallel}(\omega,K)$. In the low frequency range, this leads to the following expression:

$$C(\omega\to 0,K)=\gamma+i\omega\frac{(\gamma-1)N_{\mathrm{Pr}}\rho_0}{\phi\sigma_{\parallel}\left(K\frac{2\gamma}{N_{\mathrm{Pr}}(\gamma+1)}\right)}. \tag{57}$$

Comparison with the general low frequency asymptotic behavior given in Ref. 17 leads to the relationship between thermal permeability $k_0'(K)$ and $\sigma_{\parallel}(K)$:

$$k_0'(K)=\frac{\eta}{\sigma_{\parallel}\left(K\frac{2\gamma}{N_{\mathrm{Pr}}(\gamma+1)}\right)}. \tag{58}$$

In the high frequency limit, relationship (40) can be transformed to $C(\omega\to\infty,K)=1+(\gamma-1/N_{\mathrm{Pr}}\kappa)(2/\Lambda_{\parallel}(K))$. Further comparisons with general form for high frequency limit of complex compressibility,[17] $C(\omega\to\infty)=1+[(\gamma-1)/N_{\mathrm{Pr}}\kappa]\times(2/\Lambda')$, gives an expression for normalized thermal characteristic length.[35]

$$\Lambda'=\Lambda_{\parallel}, \tag{59}$$

where $\Lambda_{\parallel}$ is defined by Eq. (55).

As follows from the analysis, all high frequency parameters remain independent of the slip influence when the restrictions on Knudsen number values and frequency range discussed in Sec. III A are satisfied.

C. Sound speed and attenuation coefficient

Sound speeds and attenuation coefficients for two different fiber orientations can be found as

$$c_{\parallel,\perp}(\omega,K)=\frac{\omega}{\mathrm{Re}(k_{\parallel,\perp}(\omega,K))}, \tag{60a}$$

$$\delta_{\parallel,\perp}(\omega,K)=\mathrm{Im}(k_{\parallel,\perp}(\omega,K)). \tag{60b}$$

Low frequency asymptotics of complex tortuosity combined with Eq. (60a) lead to the expression for sound speeds in the viscous regime:

$$c_{\parallel}(\omega\to 0,K)=c\sqrt{\frac{2\rho_0\omega}{\sigma_{\parallel}(K=0)}(1+4KF(\phi))}, \tag{61a}$$

$$c_{\perp}(\omega\to 0,K)=c\sqrt{\frac{2\rho_0\omega}{\sigma_{\parallel}(K=0)}\left(1+\frac{4K}{1+2K}F(\phi)\right)}, \tag{61b}$$

where $F(\phi)$ is defined by Eq. (45). Attenuation coefficients in the viscous regime can be estimated as

$$\delta_{\parallel}(\omega\to 0,K)=\sqrt{\frac{\phi\sigma_{\parallel}(K=0)\omega}{2\rho_0c^2(1+4KF(\phi))}}, \tag{62a}$$

$$\delta_{\perp}(\omega\to 0,K)=\sqrt{\frac{\phi\sigma_{\perp}(K=0)\omega}{2\rho_0c^2\left(1+4\frac{4K}{1+2K}F(\phi)\right)}}. \tag{62b}$$

It is convenient to introduce relative difference in sound speed due to boundary slip,

$$\Delta_c = \frac{c(\omega \to 0, K) - c(\omega \to 0, K=0)}{c(\omega \to 0, K=0)} \tag{63a}$$

and relative difference in attenuation coefficient,

$$\Delta_\delta = \frac{\delta(\omega \to 0, K) - \delta(\omega \to 0, K=0)}{\delta(\omega \to 0, K=0)}. \tag{63b}$$

These quantities have a simple dependence on Knudsen number and porosity:

$$\Delta_{c_{\parallel,\perp}} = \sqrt{1 + s_{\parallel,\perp} F(\phi)} - 1, \tag{64a}$$

$$\Delta_{\delta_{\parallel,\perp}} = \frac{1}{\sqrt{1 + s_{\parallel,\perp} F(\phi)}} - 1, \tag{64b}$$

where $s_\parallel = 4K$, $s_\perp = 4K/(1+2K)$.

Function $F(\phi)$ approaches to zero as porosity tends to unity. This suggests weaker slip influence in high porosity materials. Moreover, Eqs. (64a) and (64b) predict stronger slip effects when sound propagates parallel to fibers.

IV. COMPARISON WITH NUMERICAL RESULTS

The FEM was employed to solve both flow and heat transfer problems in a regular square array of cylinders. The solutions have been calculated using the FEM software COMSOL MULTIPHYSICS.[36] Second-order Lagrangian elements have been used to model the velocity components and temperature distribution, whereas the linear elements approximated the pressure field, as suggested in Ref. 37. The elements were chosen so that a good enough resolution of the geometry is achieved. Periodic boundary conditions were prescribed on the boundary of the square cell with a cylinder placed at its center. An arbitrary reference pressure was set in one of the corners of the square cell.[38] Maxwell boundary condition was applied on the cylinder surface. For no-slip computations, this was replaced by usual zero velocity boundary condition. The convergence of the numerical method has been tested by comparing its results with the analytical formulation as well as by means of a mesh refining analysis for low porosity configurations. More details about the numerical procedure can be found in Refs. 13 and 14.

By setting the frequency equal to zero in Eqs. (1a) and (1b) one can get the steady state fluid flow velocity field and then steady state flow resistivity can be estimated. Figures 2(a) and 2(b) show comparisons between flow resistivity calculated numerically and using Eqs. (43) and (46) for flow parallel and perpendicular to fiber axes. Porosity range [0.34, 0.99] was considered in order to cover the values encountered in real fibrous materials. Conventional absorbers[39] can have porosities greater than 0.95, while fiber reinforced composites[40] can present a porosity value as small as 0.4. Three cylinder radii, 120, 200, and 600 nm, were considered. At normal conditions, these radii correspond to Knudsen number values of 0.5, 0.3, and 0.1 respectively. In the case of perpendicular flow, the agreement is very good for porosity values higher than 0.6. For the parallel flow, good agreement is achieved for even lower porosity values. The discrepancies at lower porosities can be explained by considering the cell shape employed in the analytical procedure. In dense arrays,

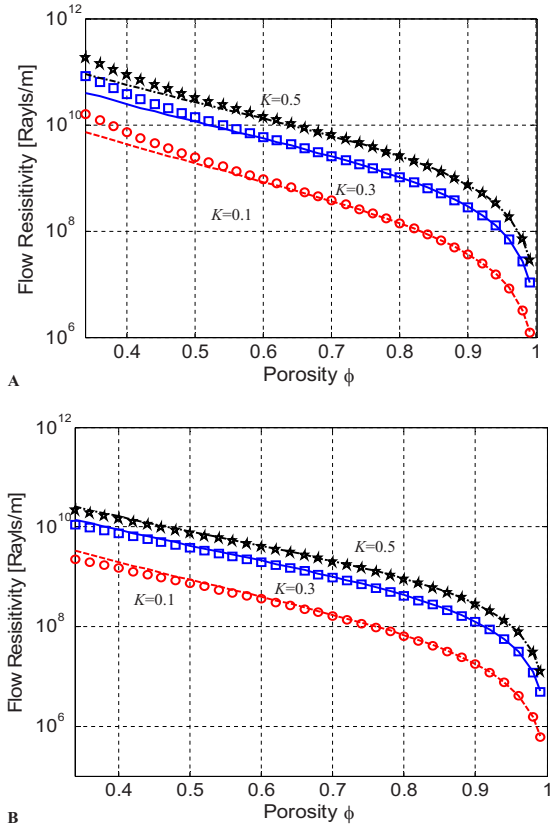


FIG. 2. (Color online) Flow resistivity for normal (a) and parallel (b) flows as a function of porosity for three different values of Knudsen number. Lines—analytical results (43)–(46); markers—numerical results.

the areas of intersections between the circular cells cannot be considered negligible. For this reason the circular cells cannot resolve the fluid flow appropriately in low porosity materials. Viscous friction is stronger when flow direction is perpendicular to cylinders which is confirmed by a higher flow resistivity value. It seems natural to expect that the error associated with the circular cell assumption is larger when viscous interactions are more pronounced. This may explain why the agreement with analytical results is worse for the case of perpendicular flow.

Analytical expressions for tortuosity and characteristic viscous length have been derived assuming very small Knudsen number values. Therefore, they have been compared with the numerical values obtained with no-slip boundary conditions on the fiber surface. The governing equation for this problem corresponds to the Laplace equation in the fluid domain (see Refs. 13 and 16 for more details). As is shown in Fig. 3(a), high frequency tortuosity for parallel flow is equal to 1 for the entire range of porosities. The linear dependence of tortuosity on porosity for perpendicular flow described by Eq. (54) is accurate for $\phi > 0.6$. Viscous characteristic length [Fig. 3(b)] shows the same trend, i.e., good agreement is obtained between numerical and analytical results when flow is parallel to cylinders. When flow direction changes, good agreement is achieved only for dilute arrays. Even in the high frequency range, when viscous boundary layer is much smaller than fiber-to-fiber distance and interactions between the cylinders are minimal, the results for the perpendicular flow in dense arrays still remain sensitive to the elementary cell shape.

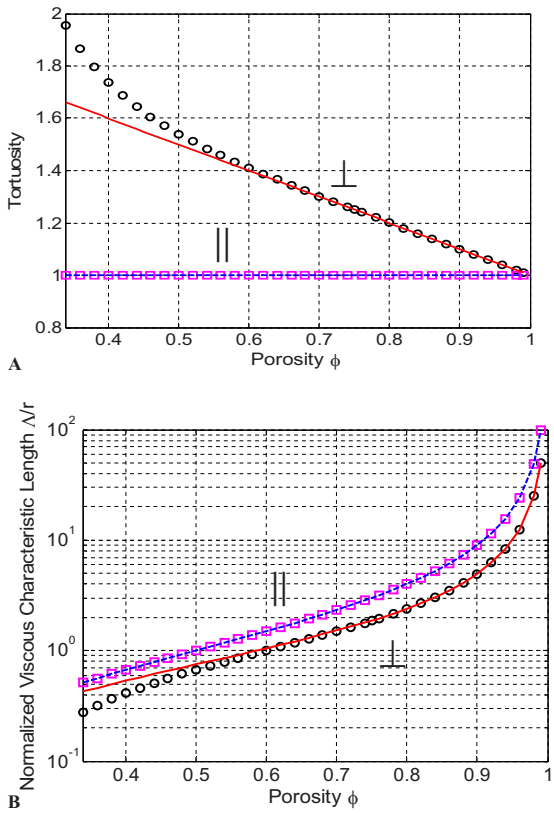


FIG. 3. (Color online) Tortuosity (a) and characteristic viscous length (b) versus porosity. ∥—fibers parallel to sound propagation direction; ⊥—fibers normal to sound propagation direction. Lines—analytical results (53)–(56); markers—numerical solutions.

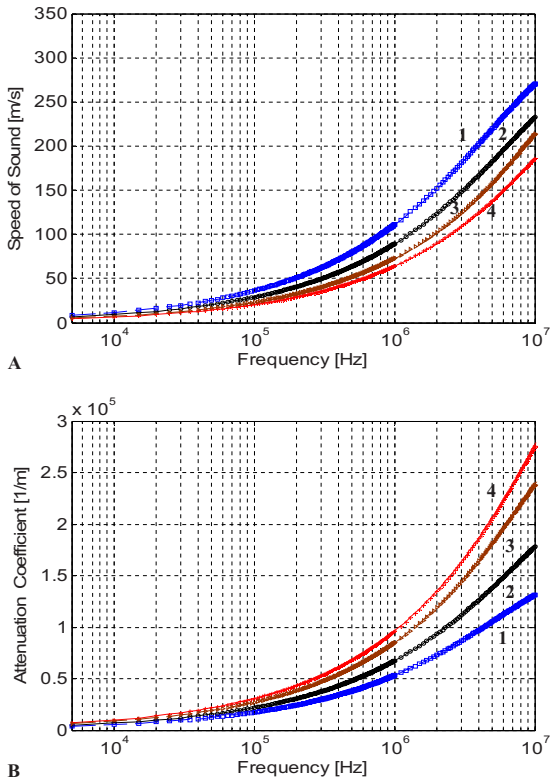


FIG. 4. (Color online) Speed of sound (a) and attenuation coefficient (b) versus frequency. Here fiber radius is 120 nm, and porosity is 0.95. Lines—analytical results (60a) and (60b); markers—numerical solutions. (1) (square)—fibers parallel to sound propagation direction, $K=0.5$; (2) (circle)—fibers parallel to sound propagation direction, no slip; (3) (triangle)—fibers normal to sound propagation direction, $K=0.5$; (4) (cross)—fibers normal to sound propagation direction, no slip.

Next, an oscillatory forced Stokes flow problem, defined by Eqs. (1a) and (1b), has been numerically solved. The macroscopic velocity field has been obtained from the solution by averaging over the cell. Complex viscous permeability $\tilde{K}_{\parallel,\perp}(\omega,K)$ was then calculated from dynamic Darcy's law.[8,26] This quantity was further related to complex tortuosity as follows:

$$\alpha_{\parallel,\perp}(\omega,K)=\frac{\eta\phi}{-i\omega\rho_0\tilde{K}_{\parallel,\perp}(\omega,K)}. \tag{65}$$

The oscillatory heat conduction problem defined by Eq. (35) has been solved in a square cell assuming thermal slip on the fluid-solid interface and periodic boundary conditions on the boundaries of the cell. The temperature was then averaged over the cell, and analogous dynamic Darcy's law for the thermal problem given by Eq. (5) in Ref. 17 was used to calculate dynamic thermal permeability $\tilde{K}^{\mathrm{th}}(\omega,K)$. The latter is related to the complex compressibility function

$$C(\omega,K)=\gamma+i\omega\frac{(\gamma-1)\rho_0 c_p\tilde{K}^{\mathrm{th}}(\omega,K)}{\phi q}. \tag{66}$$

Using complex tortuosity and complex compressibility, sound speed and attenuation coefficient have been calculated using Eqs. (60a) and (60b). Figures 4 and 5 show sound speed (a) and attenuation coefficient (b) for materials with porosity $\phi=0.95$ and fiber radius $a=120$ nm and $a=600$ nm ($K=0.5$ and $K=0.1$, respectively). For comparison calculations for the no-slip case are also shown. Simulations for frequencies between 1 kHz and 10 MHz were carried out. In general, the numerical simulations confirm a decrease in attenuation coefficient and an increase in sound speed values compared to no-slip approximation. Numerical and analytical results remain close to each other in all cases for all frequencies. Sound speed is higher when sound propagation is parallel to fiber axes and attenuation coefficient is lower in this case. For lower value of Knudsen number, $K=0.1$, the transition from the viscous to the inertial regime can be observed in Fig. 5 as sound speed approaches its limiting value. For $K=0.5$ (Fig. 4), the frequencies considered are still too low for the inertial regime to exhibit itself and sound speed continues to grow with the frequency. Naturally, the difference between slip and no-slip results is more pronounced for higher Knudsen number values.

In Figs. 6 and 7 relative differences between sound speeds ($\Delta_{c\parallel,\perp}$) and attenuation coefficients ($\Delta_{\delta\parallel,\perp}$) are shown as functions of frequency. Numerical values are compared with analytical results given by Eqs. (64a) and (64b). The numbers in the boxes represent numerically obtained values in the limit of low frequencies. The slip influence is relatively constant over the viscous regime for both sound speed and attenuation coefficient. At higher frequencies, this influence tends to decrease for speed of sound, whereas it increases for attenuation coefficient. It is also greater for higher Knudsen number at higher frequencies and does not show strong dependence on the sound propagation direction. However, the slip influence is more pronounced for the parallel

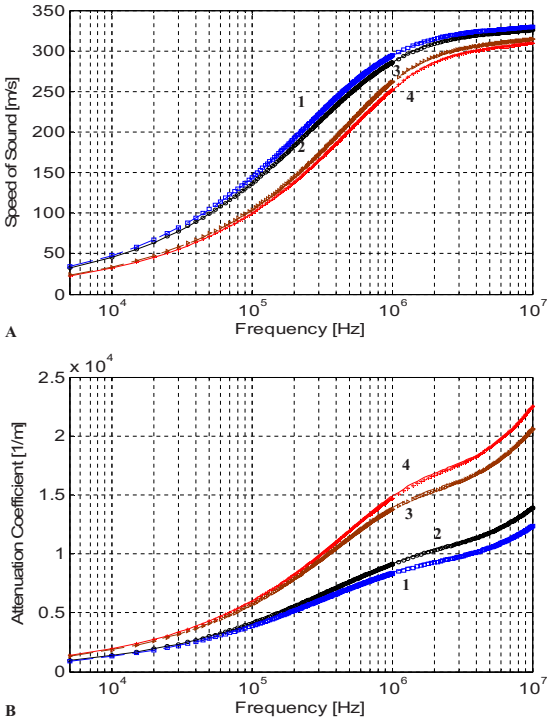


FIG. 5. (Color online) Speed of sound (a) and attenuation coefficient (b) versus frequency. Here fiber radius is $a=600$ nm, and porosity is 0.95. Lines—analytical results (60a) and (60b); markers—numerical solutions. (1) (square)—fibers parallel to sound propagation direction, $K=0.1$; (2) (circle)—fibers parallel to sound propagation direction, no slip; (3) (triangle)—fibers normal to sound propagation direction, $K=0.1$; (4) (cross)—fibers normal to sound propagation direction, no slip.

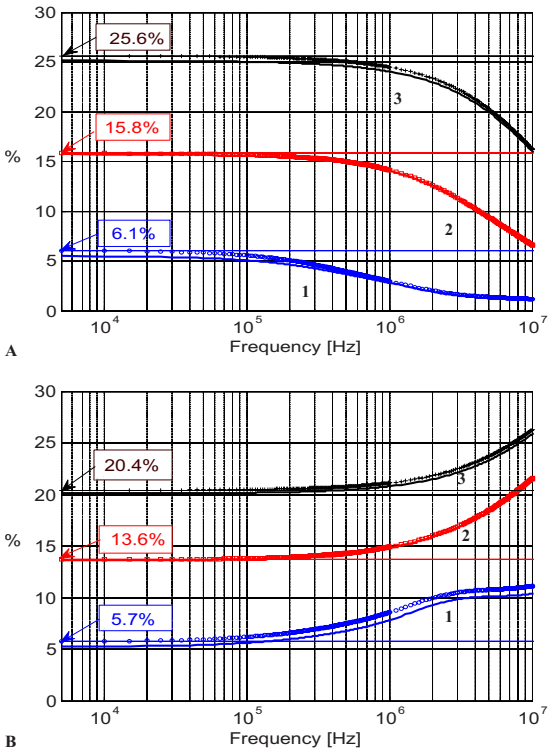


FIG. 6. (Color online) Relative difference in slip and no-slip results for sound speed (a) and attenuation coefficient (b) at different frequencies. Here fibers are parallel to sound propagation direction, and porosity is 0.95. Numbers in the boxes represent numerical low frequency values. Lines—analytical results (64a) and (64b); markers—numerical results. (1) (circle)—radius $a=600$ nm ($K=0.1$), (2) (square)—radius $a=200$ nm ($K=0.3$), (3) (plus)—radius $a=120$ nm ($K=0.5$).

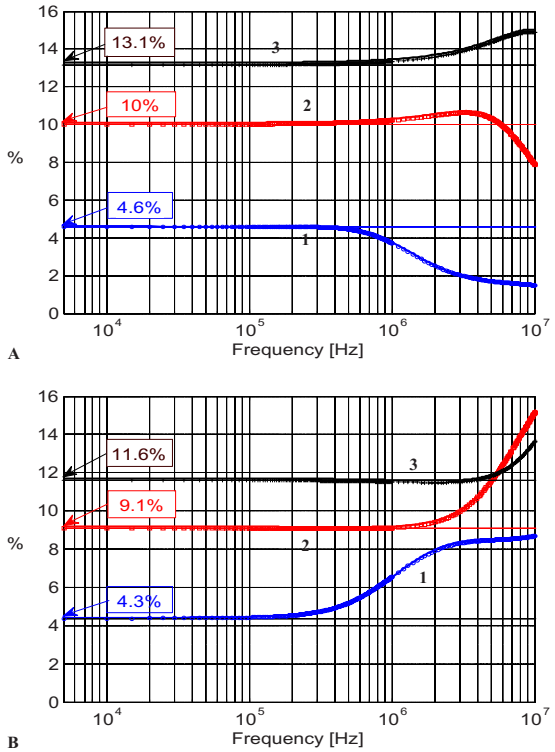


FIG. 7. (Color online) Relative difference in slip and no-slip results for sound speed (a) and attenuation coefficient (b) at different frequencies. Fibers are normal to sound propagation direction, and porosity is 0.95. Numbers in the boxes represent numerical low frequency values. Lines—analytical results (64a) and (64b); markers—numerical results. (1) (circle)—radius $a=600$ nm ($K=0.1$); (2) (square)—radius $a=200$ nm ($K=0.3$); (3) (plus)—radius $a=120$ nm ($K=0.5$).

flow case in the low frequency range. As can be seen in Fig. 6, Δ_c reaches a value of 25.6% for $K=0.5$, whereas it is lower than 13.1% for fibers perpendicular to sound propagation direction (Fig. 7). Relative errors between numerical and the analytical results for sound speed and attenuation coefficient are shown in Figs. 8 and 9. Three different porosity values, 0.35, 0.65, and 0.95, are considered. The error does not show strong frequency dependence. The analytical model is quite reliable for dilute configurations, showing an error smaller than 5% in sound speed and attenuation coefficient for both directions of sound propagation. However, as one can expect from the discrepancies found in viscous drag parameters, it becomes inaccurate at lower porosities. For $\phi=0.35$, its predictions differ from numerical results by more than 10% for sound propagating parallel to fibers and by 20% when it propagates in the perpendicular direction.

To gain a deeper understanding of the discrepancies between numerical and analytical results, one should discuss two issues. The first is related to the shape of the elementary cell, the second to the applicability of zero vorticity condition on the cell boundary. Two neighboring circular cells inevitably overlap. The shortest distance from any cylinder center to a neighboring cell boundary is $d'=\sqrt{\pi}-1/\sqrt{1-\phi}$. This distance is shown in the inset plot in Fig. 10. The neighboring cell boundary intersects the cylinder surface when this distance is equal to the cylinder radius, i.e., $d'=1$. As this situation is obviously unacceptable, it could serve as a basis for a purely geometrical criterion defining the range of

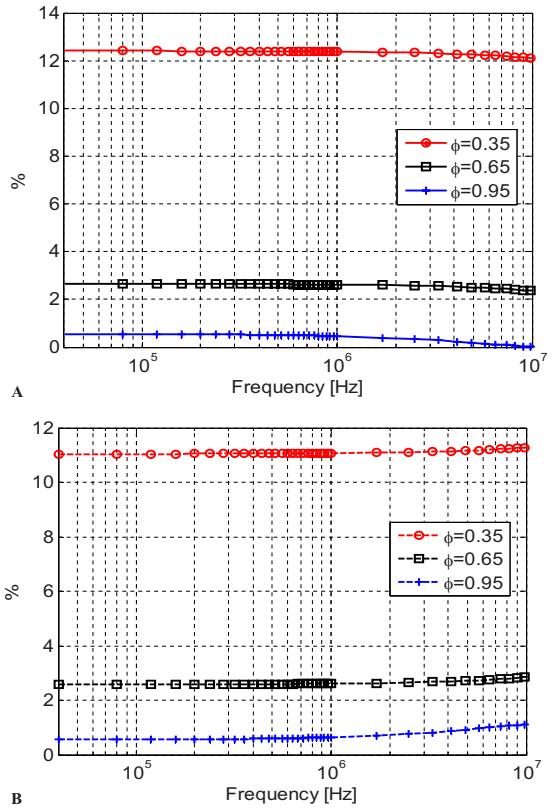


FIG. 8. (Color online) Relative error between analytical and numerical results for different array porosities: sound speed (a) and attenuation coefficient (b). Fibers parallel to sound propagation direction, and fiber radius a =200 nm (K=0.3).

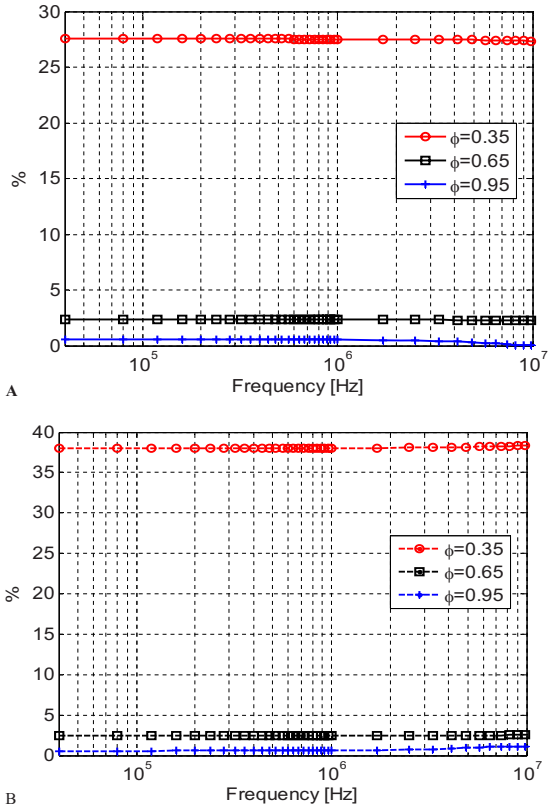


FIG. 9. (Color online) Relative error between analytical and numerical results for different porosities: sound speed (a) and attenuation coefficient (b). Fibers perpendicular to sound propagation direction, and fiber radius a =200 nm (K=0.3).

porosities for which the analytical model could be applied. This is given by $\phi \gg \pi - 2\sqrt{\pi} = 0.403$. Cylinders arranged in a dense array may also present an extra source of error associated with the zero vorticity assumption on the cell boundary. Modulus of vorticity normalized to its value on the fiber surface is shown in Fig. 10 as a function of dimensionless distance $X=(x-1)/(d_s-1)$. Here x denotes the distance from the cylinder center along the horizontal axis and $d_s = \pi/4\sqrt{1-\phi}$ is the distance to the boundary of the square cell. Surface of the cylinder corresponds to $X=0$ and the square cell boundary to $X=1$. Vorticity is calculated for two porosity values, 0.95 and 0.45, assuming stationary flow and $K=0.3$. The vorticity value is numerically zero at the square cell boundary for both directions of flow and both porosities. This means that zero vorticity assumption is accurate when the outer boundary of the circular cell d' is close to the boundary of the square cell d_s, i.e., $X'=(d'-1)/(d_s-1)\approx 1$. For porosity equal to 0.95, $X'=0.82$, and vorticity is nearly zero on the circular cell boundary. However, when porosity is equal to 0.45, $X'=0.21$ and vorticity can no longer be assumed small.

V. CONCLUSIONS

In this work, an analytical model has been developed to describe sound propagation through an array of identical parallel cylindrical fibers accounting for boundary slip effect. The results confirm that boundary slip effect can have a significant influence on the acoustical properties of microfibrous materials leading to a decrease in attenuation coefficient and an increase in sound speed. It is demonstrated that this influence depends on sound propagation direction relative to fiber axes, fiber radius, and material porosity. The effect becomes noticeable when fiber radius and/or inter-fiber distances are comparable with the molecular mean free path in air. The slip influence is stronger when sound propagates parallel to fiber axes. In this case, even for the array of relatively big fibers with radius of 0.2 μm, sound speed and attenuation coefficient can deviate by more than 13% from values obtained assuming no slip. The analytical results have been

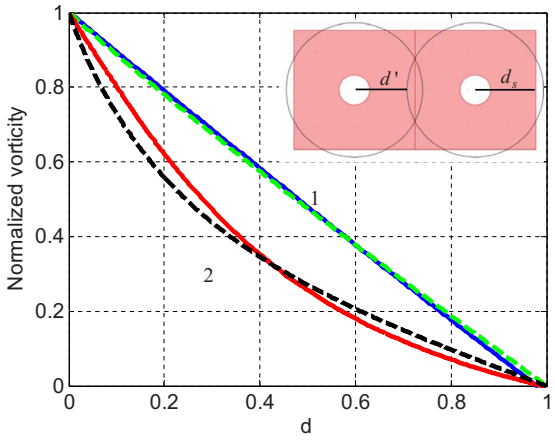


FIG. 10. (Color online) Vorticity as function of the normalized distance from the fiber surface to the edge of the square cell. (1) Normalized vorticity for perpendicular (dashed line) and parallel flow (continuous line), porosity 0.45. (2) Normalized vorticity for perpendicular (dashed line) and parallel flow (continuous line), porosity 0.95.

compared with FEM simulations. They are shown to be in a very good agreement for materials with porosities higher than 0.6. This suggests that the model can be useful for predicting acoustical properties of high-porosity microfibrous materials.

All calculations have been performed assuming air saturation at normal conditions where molecular mean free path is approximately equal to 60 nm. This dictates the size of fibers needed for experimental validation of the model. However, at different atmospheric conditions or with different gas saturations, strong influence of the boundary slip can be observed even in conventional fibrous absorbents. For instance, light glass wool described in Ref. 39 with an average fiber radius of 3.4 μm would have a Knudsen number close to 0.2 when saturated with air at 0.1 atm. High values of Knudsen number can be achieved also by saturating the material with gases such as helium, benzene, or neon where molecular mean free path is bigger than in air.[41] These situations could present an opportunity for the experimental validation of the model using conventional fibrous absorbents.

ACKNOWLEDGMENTS

R.V. gratefully acknowledges ORSAS Award and the University of Salford Research Studentship. We are grateful to anonymous referees for their valuable suggestions.

[1]K. Attenborough, "Acoustical characteristics of rigid fibrous absorbents and granular materials," J. Acoust. Soc. Am. **73**, 785–799 (1983).
[2]V. Tarnow, "Calculation of the dynamic air flow resistivity of fiber materials," J. Acoust. Soc. Am. **102**, 1680–1688 (1997).
[3]V. Tarnow, "Compressibility of air in fibrous materials," J. Acoust. Soc. Am. **99**, 3010–3017 (1996).
[4]V. Tarnow, "Airflow resistivity of models of fibrous acoustic materials," J. Acoust. Soc. Am. **100**, 3706–3713 (1996).
[5]O. Umnova, K. Attenborough, and K. M. Li, "Cell model calculations of dynamic drag parameters in packings of spheres," J. Acoust. Soc. Am. **107**, 3113–3119 (2000).
[6]A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic Analysis of Periodic Structures* (North-Holland, Amsterdam, 1978).
[7]E. Sanchez-Palencia, *Non Homogeneous Media and Vibration Theory* (Springer-Verlag, Berlin, 1980).
[8]J. L. Auriault, L. Borne, and R. Chambon, "Dynamics of porous saturated media, checking of the generalized law of Darcy," J. Acoust. Soc. Am. **77**, 1641–1650 (1985).
[9]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[10]A. Craggs and J. G. Hildebrandt, "Effective densities and resistivities for acoustic propagation in narrow tubes," J. Sound Vib. **92**, 321–331 (1984).
[11]M. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**, 550–558 (1991).
[12]V. F. Kozlov, A. V. Fedorov, and N. D. Malmuth, "Acoustical properties of rarefied gases inside pores of simple geometries," J. Acoust. Soc. Am. **117**, 3402–3411 (2005).
[13]R. Venegas and O. Umnova, "On the influence of the micro-geometry on sound propagation through periodic array of cylinders," in Proceedings of Acoustics'08 Paris, France (2008), pp. 807–812; [J. Acoust. Soc. Am. **123**, 3142 (2008)].
[14]C. Perrot, F. Chevillotte, and R. Panneton, "Dynamic viscous permeability of an open-cell aluminum foam: Computations versus experiments," J. Appl. Phys. **103**, 024909 (2008).
[15]A. Cortis, D. Smeulders, D. Lafarge, M. Firdaouss, and J. L. Guermond, "Geometry effects on sound in porous media," in IUTAM Symposium on Theoretical and Numerical Methods in Continuum Mechanics of Porous Materials, University of Stuttgart, Germany (1999).
[16]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[17]D. Lafarge, P. Lemarinier, J. F. Allard, and V. Tarnow, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[18]M. Gad-el-Hak, *The MEMS Handbook* (CRC, Boca Raton, FL, 2002).
[19]A. Soleimani Dorcheh and M. H. Abbasi, "Silica aerogel: Synthesis, properties and characterization," J. Mater. Process. Technol. **199**, 10–26 (2008).
[20]F. Bechwati, O. Umnova, and T. Cox, "New semi-empirical model for sound propagation in adsorbing microporous solids (activated carbon)," in Proceedings of the 19th ICA, Madrid, Spain (2007).
[21]J. Gross, G. Reichenauer, and J. Fricke, "Mechanical properties of SiO_2 aerogels," J. Phys. D: Appl. Phys. **21**, 1447–1451 (1988).
[22]D. A. Lockerby, J. M. Reese, D. R. Emerson, and B. W. Barker, "Velocity boundary conditions at solid walls in rarefied gas calculations," Phys. Rev. E **70**, 017303 (2004).
[23]C. Shen, *Rarefied Gas Dynamics, Fundamentals, Simulations and Micro Flows* (Springer-Verlag, Berlin, 2005).
[24]N. G. Hadjiconstantinou and O. Simek, "Sound propagation at small scales under continuum and non-continuum transport," J. Fluid Mech. **488**, 399–408 (2003).
[25]E. Skjetne and J. L. Auriault, "Homogenization of wall-slip gas flow through porous media," Transp. Porous Media **36**, 293–306 (1999).
[26]J. Chastanet, P. Royer, and J. L. Auriault, "Acoustics with wall-slip flow of gas saturated porous media," Mech. Res. Commun. **31**, 277–286 (2004).
[27]D. Tsiklauri, "Phenomenological model of propagation of the elastic waves in a fluid saturated porous solid with nonzero boundary slip velocity," J. Acoust. Soc. Am. **112**, 843–849 (2002).
[28]M. G. Markov, "Effect of the interfacial slip on the kinematic and dynamic parameters of elastic waves in a fluid-saturated porous medium," Acoust. Phys. **53**, 213–216 (2007).
[29]S. Kuwabara, "The forces experienced by two circular cylinders in a uniform flow at small Reynolds numbers," J. Phys. Soc. Jpn. **12**, 291–299 (1957).
[30]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1972), pp. 360 and 364.
[31]C. Boutin and C. Geindreau, "Estimates and bounds of dynamic permeability of granular media," J. Acoust. Soc. Am. **124**, 3576–3593 (2008).
[32]C. Ho and S. W. Webb, *Gas Transport in Porous Media* (Springer, Amsterdam, 2006).
[33]J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[34]P. Wong, *Methods in the Physics of Porous Media* (Academic, San Diego, 1999).
[35]Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[36]COMSOL MULTIPHYSICS 3.3a Documentation, www.comsol.com (Last viewed 21/11/08).
[37]T. J. Chung, *Computational Fluid Dynamics*, (Cambridge University Press, Cambridge, 2002), Chaps. 10 and 12.
[38]W. Zimmerman, *Process Modelling and Simulation With Finite Element Methods*, Series on Stability, Vibration and Control of Systems, Series a (World Scientific, Singapore, 2004), p. 182.
[39]V. Tarnow, "Measured anisotropic air flow resistivity and sound attenuation of glass wool," J. Acoust. Soc. Am. **111**, 2735–2739 (2002).
[40]P. B. Nedanov and S. G. Advani, "A method to determine 3D permeability of fibrous reinforcements," J. Compos. Mater. **36**, 241–254 (2002).
[41]J. O. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids* (Wiley, New York, 1954).

Acoustic absorption calculation in irreducible porous media: A unified computational approach

Chang-Yong Lee and Michael J. Leamy[a)]
School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Jason H. Nadler
Georgia Tech Research Institute (GTRI), Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 28 December 2008; revised 8 July 2009; accepted 24 July 2009)

A critical task in predicting and tailoring the acoustic absorption properties of porous media is the calculation of the frequency-dependent effective density and compressibility tensors, which are explicitly related to the micro-scale permeability properties. Although these two quantities exhibit strong sensitivity to physics occurring at complex micro-scale geometries, most of the existing literature focuses on employing very limited in-house and oftentimes multiple numerical analysis tools. In order to predict these parameters and acoustic absorption efficiently and conveniently, this article synthesizes multiple disparate approaches into a single unified formulation suitable for incorporation into a commercial analysis package. Numerical results computed herein for four close-packed porous media are compared to similar results available in the literature. These include simple cubic, body-centered cubic, and face-centered cubic structures, and also hexagonal close-packed, which has not appeared in the literature. Together with critical comparisons of a hybrid versus direct numerical approaches, the close agreement demonstrates the capabilities of the unified formulation to analyze and control the acoustic absorption properties at the microscopic level. [DOI: 10.1121/1.3205399]

PACS number(s): 43.55.Ev, 43.20.Jr [NX] Pages: 1862–1870

I. INTRODUCTION

Noise reduction remains a major environmental and safety issue in the automotive, aeronautical, and construction industries. One promising way to reduce noise is through the design and optimization of porous structures, where viscosity and large surface areas result in favorable acoustic absorption properties. For acoustic propagation and absorption of porous materials, Kirchhoff[1] provided first the phenomenological model of viscous and thermal effects created in cylindrical tubes with circular cross-section. For more general geometries, his analytic description presents complications making it unsuitable for use in real engineering applications. Instead, other techniques have appeared in the recent literature, which identify and quantify two important measurement parameters characterizing and predicting acoustic absorption performance, namely, the effective density and compressibility tensors. For this reason, research attention devoted to the theoretical formulation of these two parameters has received considerable attention in the past decades (see, e.g., Refs. 2–6 for a review). Here we briefly mention the most successful theoretical models in acoustics of porous materials. One of the widely accepted and used ones is the Johnson–Allard model, which was proposed through the equivalent fluid description by Johnson *et al.*[7] for the effective density and by Champoux and Allard[8] for the effective compressibility. Pride *et al.*[9] and Lafarge[10] improved this model by resolving underestimation of the imaginary part of the relaxation functions of the porous medium. Although most of the formulations in the cited works are given in an explicit analytical form, they can only be evaluated in cases of very simple pore geometries assuming isotropic material properties, such as flow in a uniform cylinder. More recently, several numerical methods have been suggested in the literature, such as the collocation method,[11] the boundary element method,[12] the lattice Boltzmann method,[13] and the finite element method.[14–18] Those offer the possibility of extending the evaluation to general geometries and materials using desktop-based workstations. However, most of these analysis techniques require the use of (multiple) in-house codes and complex numerical procedures. Furthermore, many of the cited numerical treatments are not suitable for arbitrary pore geometry and materials. Therefore, there is a need for an integrated and general computational approach for predicting the acoustic properties of periodic porous materials, which, ideally, can be applied in a form suitable for analysis using commercial analysis tools.

In this paper, we present a multi-scale numerical approach for determining acoustic absorption properties based on the effective density and compressibility parameters for periodic porous media with arbitrary three-dimensional (3D) unit cells. To the authors' knowledge, the approach presented is the first posed in a consistent and general manner such that the resulting micro-scale equations can be implemented in a single commercial off-the-shelf (COTS) simulation code. We implement the resulting micro-scale equation set in COMSOL MULTIPHYSICS®, which is a simulation and modeling tool general enough to solve many frequently encountered partial differential equations (PDEs). Importantly, this same com-

[a)]Author to whom correspondence should be addressed. Electronic mail: michael.leamy@me.gatech.edu

mercial package is equipped with tools for complex geometry modeling, mesh generation, and handling of general boundary conditions to include periodic boundary conditions, as needed herein. The resulting analysis tool is therefore very general and capable of predicting acoustic absorption in complex 3D fluid media. We also present an otherwise missing 3D comparison of acoustic absorption results generated using the direct and hybrid numerical approaches.

In Sec. II, we begin our presentation by reviewing a mathematical procedure suitable for analyzing acoustic porous medium: the multi-scale asymptotic method (MAM).[19–23] MAM enables one to determine the macroscopic material description from knowledge of the physics and geometry at the microscopic level. Section II A employs this method to derive a set of frequency-domain micro-scale PDEs, and boundary conditions, for decoupled dynamic and thermal response. In Sec. II B, the direct MAM approach and an alternative hybrid numerical procedure are detailed for calculating the effective density and compressibility parameters using only the domain of the unit fluid cell (UFC). The hybrid approach is an efficient semi-analytic means to calculate these same two tensors.[15–18] To determine the acoustic absorption properties for rigid porous media, the acoustic model[5,24,25] is briefly described in Sec. II C. Finally, in Sec. III, for each porous medium composed of four periodic rigid sphere packings—simple cubic (sc), body-centered cubic (bcc), face-centered cubic (fcc), and hexagonal close-packed (hcp)—comparisons of results generated in this study with those found in the literature[11,14–16] demonstrate very good agreement.

II. MAM

In order to describe the linear acoustic phenomena created in a porous medium, one has to solve the following set of frequency-domain viscous-thermal governing equations and associated boundary conditions for harmonic waves characterized by frequency ω.

In the fluid domain,

$$\frac{p}{\rho_0 T_0} = C_P - C_V \tag{1a}$$

or

$$\frac{p}{P_0} = \frac{\rho}{\rho_0} + \frac{\tau}{T_0} \quad \text{(state equation of a perfect gas)}, \tag{1b}$$

$$\rho_0 i\omega \mathbf{u} = -\nabla p + (\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \mu \Delta \mathbf{u}. \tag{2a}$$

$$i\omega \frac{\rho}{\rho_0} = -\nabla \cdot \mathbf{u} \quad \text{(momentum and mass balance)}, \tag{2b}$$

$$\rho_0 i\omega C_p \tau = i\omega p + K\Delta\tau \quad \text{(energy balance)}. \tag{3}$$

On the fluid-solid interface,

$$\mathbf{u} = 0 \tag{4a}$$

and

$$\tau = 0 \quad \text{(no-slip and isothermal conditions)}. \tag{4b}$$

Here P_0, ρ_0, and T_0 denote the pressure, density, and temperature of the air at the rest, while $\mathbf{u}$, p, ρ, and τ denote the fluid velocity, pressure variation, density variation, and temperature variation. In addition, the shear and bulk viscosities, specific heats at constant pressure and constant volume, and heat conductivity are denoted, respectively, by μ, λ, C_P, C_V, and K. As has been noted by several authors, any problem of linear acoustics involving porous media can be dealt with using this formalism. However, because the equations presented above are based on interdependent macroscopic variables, they cannot take into account explicitly the micro-scale physics and geometry of the porous media at the microscopic level.

MAM is a multi-scale approach based on an asymptotic analysis of the governing equations (1)–(4). It is used to further derive a set of well-posed micro-scale equations necessary for computing effective macro-scale variables, via averaging over the unit cell. The asymptotic approach begins by introducing two space variables: x for the macro-variations and $y = \varepsilon^{-1}x$ for the micro-variations. Here the small parameter $\varepsilon = l/L \ll 1$ is a scale ratio of a characteristic unit cell length l and frequency-dependent wavelength L. As such, the small parameter denotes a ratio of a micro-scale characteristic length to a macro-scale characteristic length. The solution variables sought and the differential operators are next split into their macroscopic and microscopic components via power series involving ε,

$$\mathbf{u} = \mathbf{u}^0(x,y) + \varepsilon \mathbf{u}^1(x,y) + \varepsilon^2 \mathbf{u}^2(x,y) + \cdots,$$

$$p = p^0(x,y) + \varepsilon p^1(x,y) + \varepsilon^2 p^2(x,y) + \cdots,$$

$$\tau = \tau^0(x,y) + \varepsilon \tau^1(x,y) + \varepsilon^2 \tau^2(x,y) + \cdots. \tag{5}$$

The gradient (∇) and Laplacian (Δ) operators take the forms

$$\nabla = \nabla_x + \frac{1}{\varepsilon}\nabla_y \quad \text{and} \quad \Delta = \Delta_x + \frac{2}{\varepsilon}\Delta_{xy} + \frac{1}{\varepsilon^2}\Delta_y. \tag{6}$$

Moreover, because one can consider the viscous-thermal effects to occur at the micro-scale, it is necessary to rescale the viscosity and conductivity coefficients appearing in the momentum balance (2a) and energy equations (3) by ε^2 (Ref. 23)

$$\rho_0 i\omega \mathbf{u} = -\nabla p + \varepsilon^2[(\lambda + \mu)\nabla(\nabla \cdot \mathbf{u}) + \mu \Delta \mathbf{u}], \tag{7}$$

$$\rho_0 i\omega C_p \tau = i\omega p + \varepsilon^2 K \Delta \tau. \tag{8}$$

With this rescaling, MAM allows for orderly solution of the seeking variables.

A. Micro-scale boundary value problems

We start with the mass balance equation (2b), combined with the state equation (1b). Updated using the expansions (5) and (6), a single multi-scaled relationship results,

$$i\omega\left[\frac{(p^0+\varepsilon p^1+\cdots)}{P_0}-\frac{(\tau^0+\varepsilon\tau^1+\cdots)}{T_0}\right]$$
$$=-\left(\nabla_x+\frac{1}{\varepsilon}\nabla_y\right)\cdot(\mathbf{u}^0+\varepsilon\mathbf{u}^1+\cdots). \quad (9)$$

Separating scales, the corresponding equation at the highest order ε^{-1} implies that the fluid velocity can be considered as locally incompressible,

$$\nabla_y\cdot\mathbf{u}^0=0, \quad (10)$$

while at order ε^0,

$$i\omega\left[\frac{p^0}{P_0}-\frac{\tau^0}{T_0}\right]=-\nabla_x\cdot\mathbf{u}^0-\nabla_y\cdot\mathbf{u}^1. \quad (11)$$

A similar procedure carried out on the momentum balance Eq. (7) yields

$$\rho_0 i\omega(\mathbf{u}^0+\varepsilon\mathbf{u}^1+\cdots)$$
$$=-\left(\nabla_x+\frac{1}{\varepsilon}\nabla_y\right)(p^0+\varepsilon p^1+\cdots)$$
$$+\varepsilon^2\left\{\mu\left(\Delta_x+\frac{2}{\varepsilon}\Delta_{xy}+\frac{1}{\varepsilon^2}\Delta_y\right)(\mathbf{u}^0+\varepsilon\mathbf{u}^1+\cdots)\right.$$
$$+(\lambda+\mu)\left(\nabla_x+\frac{1}{\varepsilon}\nabla_y\right)\left(\left(\nabla_x+\frac{1}{\varepsilon}\nabla_y\right)\right.$$
$$\left.\left.\cdot(\mathbf{u}^0+\varepsilon\mathbf{u}^1+\cdots)\right)\right\}. \quad (12)$$

Separating orders, order ε^{-1} yields the relationship

$$\nabla_y p^0=0\Rightarrow p^0(x,y)=p^0(x), \quad (13)$$

which implies that the macro-scale pressure is constant at the micro-scale. At order ε^0,

$$\rho_0 i\omega\mathbf{u}^0=-\nabla_y p^1-\nabla_x p^0+\mu\Delta_y\mathbf{u}^0. \quad (14)$$

For the energy equation (8), the procedure results in the multi-scale equation,

$$\rho_0 i\omega C_p(\tau^0+\varepsilon\tau^1+\cdots)=i\omega(p^0+\varepsilon p^1+\cdots)$$
$$+\varepsilon^2\left[K\left(\Delta_x+\frac{2}{\varepsilon}\Delta_{xy}+\frac{1}{\varepsilon^2}\Delta_y\right)\right.$$
$$\left.\times(\tau^0+\varepsilon\tau^1+\cdots)\right]. \quad (15)$$

Unlike in the other scaled equations, the highest order identified in Eq. (15) is ε^0 and thus a single relationship is obtained (macro-scale),

$$\rho_0 i\omega C_p\tau^0=i\omega p^0+K\Delta_y\tau^0. \quad (16)$$

Following the methodology introduced by Lafarge *et al.*,[26] we assume, at a given frequency, the appropriate solution forms of the micro-scale velocity ($\mathbf{u}^0$) and pressure (p^1) distributions as linearly related to $-\nabla_x p^0$, while the micro-scale temperature (τ^0) distribution as linearly related to $i\omega p^0$, as follows:

$$\mathbf{u}^0(x,y)=-\frac{\mathbf{k}(y,\omega)}{\mu}\cdot\nabla_x p^0(x), \quad (17a)$$

$$p^1(x,y)=-\boldsymbol{\pi}(y,\omega)\cdot\nabla_x p^0(x)+\hat{p}^1(x), \quad (17b)$$

and

$$\tau^0(x,y)=\frac{k'(y,\omega)}{K}i\omega p^0(x). \quad (18)$$

Note that $\mathbf{k}$ and k' denote the micro-scale dynamic viscous and thermal permeability functions, and also note that the pressure (p^1) can be expressed in terms of its deviatoric part ($\hat{p}^1$) and zero mean value ($\boldsymbol{\pi}$) on the UFC. Substituting Eqs. (17) and (18) into Eqs. (10), (14), and (16) yields the following decoupled set of PDEs suitable for incorporation into a commercial finite element code.

For momentum equation with no-slip boundaries,

$$i\omega\frac{\rho_0}{\mu}\mathbf{k}+\nabla_y\boldsymbol{\pi}-\Delta_y\mathbf{k}=\mathbf{I} \quad \text{in } \Omega_f$$

$$\nabla_y\cdot\mathbf{k}=0 \quad \text{in } \Omega_f$$

$$\mathbf{k}=0 \text{ on } \Gamma \quad (\mathbf{k} \text{ and } \boldsymbol{\pi}:\Omega-\text{periodic}), \quad (19)$$

where $\mathbf{I}$ denotes a 3×3 identity matrix composed of three unit vectors ($\mathbf{e}$) directed along a global coordinate system.

For energy equation with isothermal boundaries,

$$\frac{\rho_0\ \mathrm{Pr}}{\mu}i\omega\rho_0 k'-\Delta_y k'=1 \quad \text{in } \Omega_f$$

$$k'=0 \text{ on } \Gamma(k':\Omega-\text{periodic}), \quad (20)$$

where $\mathrm{Pr}=\mu C_p/K$ denotes the Prandtl number.

Here Ω, Ω_f, and Γ represent the UFC volume, the fluid-filled pore volume, and the fluid-solid interface, respectively. For a given frequency, we implement Eqs. (19) and (20) in the PDE module provided by COMSOL MULTIPHYSICS®, utilizing the Parametric solver. We then apply no-slip isothermal and translational periodic boundary conditions upon the *xyz*-directions to an irreducible UFC consisting of a fluid-filled interstitial space between given packed spheres (Fig. 2). Finally, we employ a COMSOL script to iteratively compute the micro-scale dynamic viscous and thermal permeability functions in a frequency range of interest.

B. Macro-scale effective parameters in direct and hybrid numerical approaches

Following solution of the micro-scale equations (Sec. II A), one can obtain the following effective density ($\boldsymbol{\rho}_{\mathrm{eff}}$) and compressibility ($\chi_{\mathrm{eff}}$) parameters via averaging the solutions over the UFC:

$$i\omega\boldsymbol{\rho}_{\mathrm{eff}}\cdot\langle\mathbf{u}^0\rangle=-\nabla_x p^0, \quad (21a)$$

and

$$i\omega\chi_{\mathrm{eff}}p^0=-\nabla_x\cdot\langle\mathbf{u}^0\rangle, \quad (21b)$$

with $\langle\cdot\rangle=\int_\Omega\cdot d\Omega/\Omega_f$. As mentioned before, we introduce two numerical procedures in this section. One uses the MAM

procedure directly and requires explicit computation at each frequency of interest. The other is a hybrid approach, which uses alternatively derived physical parameters as input data in an explicit analytical model developed by Pride *et al.*[9] and Lafarge.[10] The hybrid approach significantly decreases the computational burden as compared to the direct approach.

1. Direct numerical approach

Taking the volume average of Eq. (17a) over the UFC, multiplying by the porosity ϕ $(=\Omega_f/\Omega)$ and using the definition (21a), one can easily determine the macro-scale effective density parameter for general periodic porous media as follows:

$$\boldsymbol{\rho}_{\text{eff}} = \frac{\mu\phi}{i\omega}\hat{\mathbf{k}}^{-1}, \tag{22}$$

where $\hat{\mathbf{k}}(=\phi\langle\mathbf{k}\rangle)$ denotes the dynamic viscous permeability parameter. Applying the same procedure as above with Eqs. (11) and (18), together with the divergence theorem, one obtains the effective compressibility parameter,

$$i\omega\frac{p^0}{P_0}\left[1 - i\omega\frac{P_0}{\phi K T_0}\hat{k}'\right] = -\nabla_x\cdot\langle\mathbf{u}^0\rangle. \tag{23}$$

In arriving at Eq. (23), we have used the fact that the derivative of a periodic function integrated over one period is zero, i.e., $\langle\nabla_y\cdot f(x,y)\rangle=0$. In Eq. (23), $\hat{k}'(=\phi\langle k'\rangle)$ represents the dynamic thermal permeability parameter. Introducing the specific heat ratio $\gamma(=C_P/C_V)$ and the Prandtl number into Eq. (1a) and comparing Eq. (23) with Eq. (21b), one can associate the macro-scale effective compressibility parameter with the dynamic thermal permeability parameter,

$$\chi_{\text{eff}} = \frac{1}{\gamma P_0}\left[\gamma - (\gamma-1)\frac{\rho_0}{\mu}\frac{\text{Pr}\, i\omega\hat{k}'}{\phi}\right]. \tag{24}$$

Therefore, taking into account explicitly the micro-scale physics and geometry of the porous media at the microscopic level, we have obtained the dynamic effective density and compressibility tensors [Eqs. (22) and (24)], which ultimately determine the acoustic absorption properties at the macroscopic level.

2. Hybrid numerical approach

The direct approach detailed above requires significant computation at each frequency and can suffer from convergence issues. Recently, an alternative, analytic-based approach has been introduced[15–18] (termed herein the hybrid numerical approach). The main advantage of this approach is the need to solve only three *static* problems instead of Eqs. (19) and (20): Stokes, heat, and electrical conduction problems. Moreover, unlike the direct numerical approach, the statically estimated parameters can be reduced from tensor-based values to scalars by referring to their own symmetry properties in static regime.[27] These then serve as input data in explicit expressions for frequency-dependent effective density and compressibility variables proposed by Pride *et al.*[9] and Lafarge,[10] respectively. These statically estimated

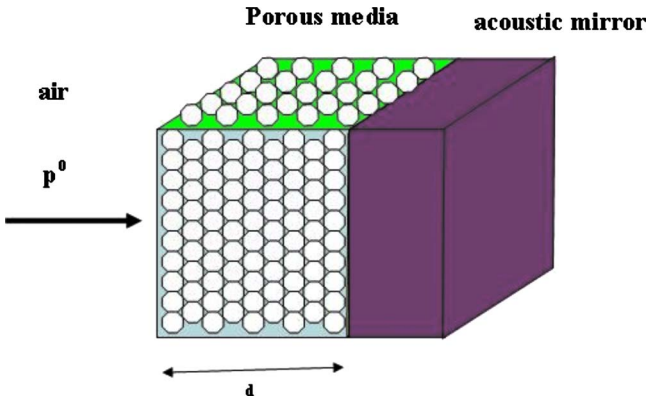


FIG. 1. (Color online) Acoustic absorption propagation model.

parameters are taken to be the static viscous and thermal permeabilities $(\hat{k}_0,\hat{k}_0')$, the static viscous and thermal tortuosities (α_0,α_0'), the tortuosity factor (α_∞), and the viscous and thermal characteristic lengths (Λ,Λ'), which are automatically computed as additional outcomes of the direct numerical procedure. Here subscripts 0 and ∞ represent quantities evaluated at $\omega=0$ and the high-frequency limit case, respectively.

Substituting a zero angular frequency into Eqs. (19) and (20), two static permeabilities, two static tortuosities, and the thermal characteristic lengths are simply calculated by the following standard definitions:[7,10]

$$\hat{\mathbf{k}}_0 = \phi\langle\mathbf{k}_0\rangle = \hat{k}_0, \quad \hat{k}_0' = \phi\langle k_0'\rangle, \quad \alpha_0 = \frac{\langle\mathbf{k}_0^2\rangle}{\langle\mathbf{k}_0\rangle^2},$$

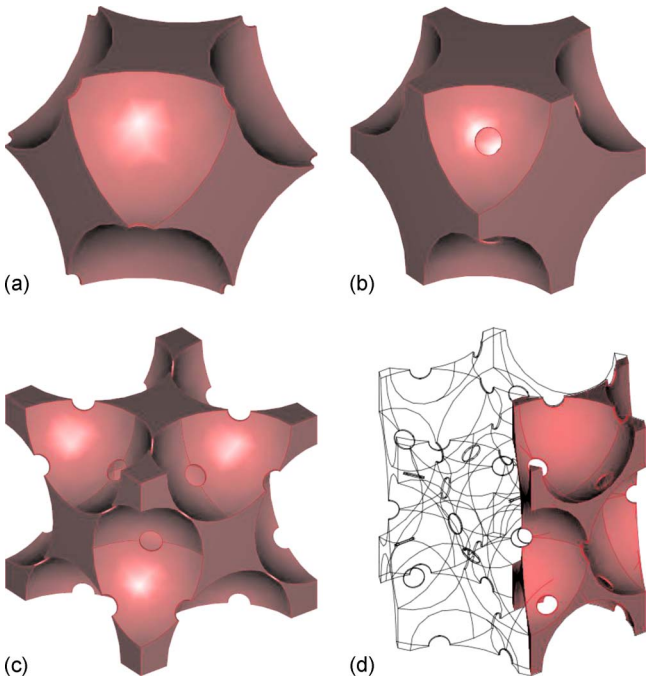


FIG. 2. (Color online) 3D UFC geometries with solder joints. (a) sc periodic model, (b) bcc periodic model, (c) fcc periodic model, and (d) hcp periodic model.

TABLE I. Fluid material properties used in all studies.

ρ_0 (kg/m^3)	T_0 (K)	P_0 (Pa)	μ (kg ms^{-1})	Pr	γ
1.293	300	10^5	1.72×10^{-5}	0.715	1.4

$$\alpha_0' = \frac{\langle k_0'^2 \rangle}{\langle k_0' \rangle^2}, \quad \text{and} \quad \Lambda' = 2\frac{\int\int\int_\Omega d\Omega}{\int\int_\Gamma d\Gamma}. \tag{25}$$

Two parameters defining the tortuosity factor and viscous characteristic length,

$$\boldsymbol{\alpha}_\infty^{-1} = \langle \mathbf{E}\cdot\mathbf{e}\rangle = \alpha_\infty \quad \text{and} \quad \Lambda = 2\frac{\int\int\int_\Omega \mathbf{E}^2 d\Omega}{\int\int_\Gamma \mathbf{E}^2 d\Gamma}, \tag{26}$$

respectively, are obtained by solving the following electrical conduction problem of the porous medium:[14,17,18,28]

$$\nabla_y\cdot\mathbf{E} = 0 \quad \text{with} \quad \mathbf{E} = -\nabla_y\varphi + \mathbf{e} \quad \text{in} \quad \Omega_f,$$

$$\mathbf{E}\cdot\mathbf{n} = 0 \quad \text{on} \quad \Gamma(\varphi{:}\Omega - \text{periodic}), \tag{27}$$

where $\mathbf{E}_j$ represents the scaled electric field, φ represents the deviatoric part of the electric potential, and $\mathbf{n}$ represents the unit outward normal vector from the pore region. Finally, based on the estimated values from Eqs. (25) and (26), the following effective density of Pride *et al.*[9] and Lafarge's effective compressibility[10] are provided for any given frequency:

$$\rho_{\text{eff}} = \rho_0\alpha_\infty\left[1 + \frac{f(\zeta)}{i\zeta}\right] \quad \text{and}$$

$$\chi_{\text{eff}} = \frac{1}{\gamma P_0}\left\{\gamma - (\gamma-1)\left[1 + \frac{f'(\zeta')}{i\zeta'}\right]^{-1}\right\}. \tag{28}$$

With

$$f(\zeta) = 1 - q + q(1 + i\zeta M/2q^2)^{1/2},$$

$$f'(\zeta') = 1 - q' + q'(1 + i\zeta M'/2q'^2)^{1/2}, \tag{29}$$

$$\zeta = \omega\rho_0\hat{k}_0\alpha_\infty/\mu\phi, \quad \zeta' = \omega\rho_0\hat{k}_0' \ \text{Pr}/\mu\phi,$$

$$M = 8\hat{k}_0\alpha_\infty/\Lambda^2\phi, \quad M' = 8\hat{k}_0'/\Lambda'^2\phi, \tag{30}$$

and

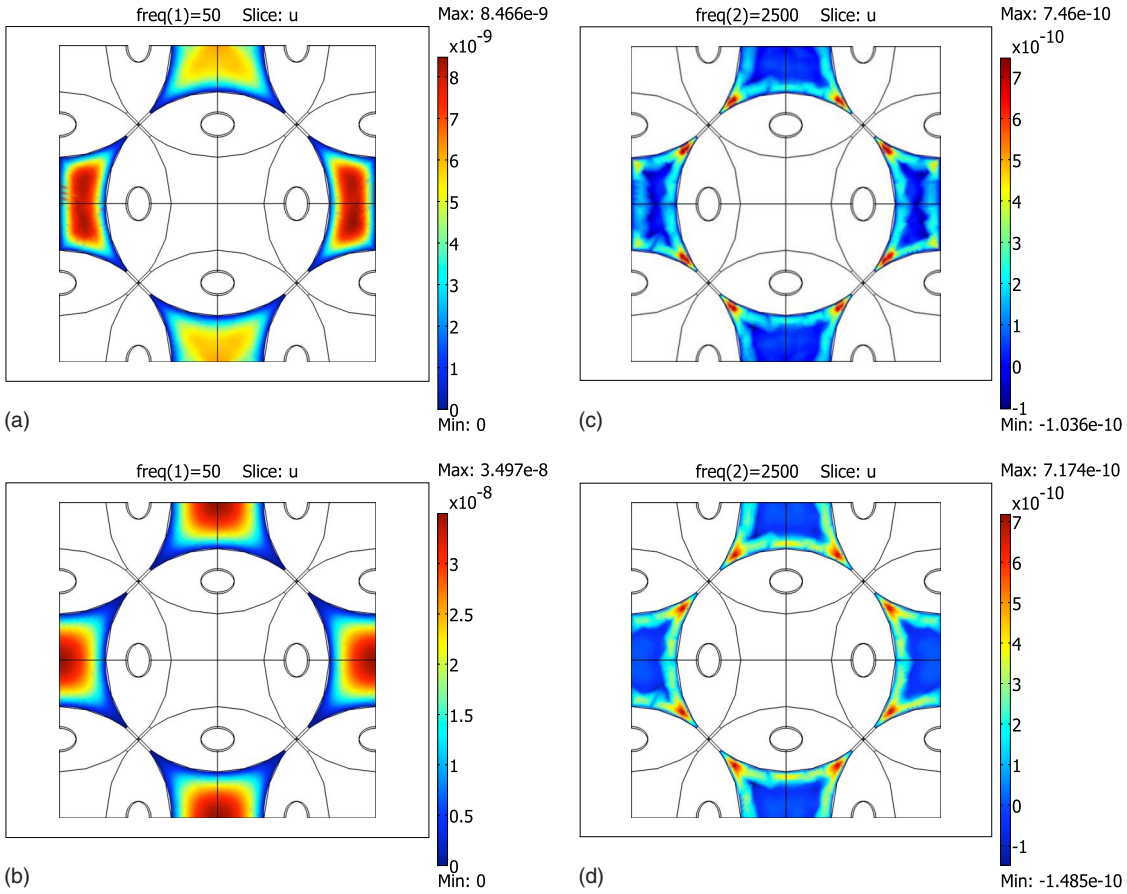


FIG. 3. (Color online) Real parts of the dynamic viscous and thermal permeability distributions (k_{11} and k') at [(a) and (b)] 50 Hz and [(c) and (d)] 2500 Hz.

$$q = M/4(\alpha_0/\alpha_\infty - 1), \quad q' = M'/4(\alpha_0' - 1), \tag{31}$$

where f and f' represent dimensionless viscous and thermal shape functions of each angular frequency (ζ, ζ'), and M and M' represent dimensionless viscous and thermal shape factors, respectively, while q and q' represent dimensionless supplementary parameters.

C. Acoustic absorption calculations based on two effective parameters

To characterize the acoustic absorbing effect on a harmonic plane wave propagating through the porous media, the wave number (κ) and effective impedance (Z_{eff}) must be defined before computing the absorption coefficient. Let us consider 3D acoustic absorption propagation model for the porous structure with thickness d (Fig. 1).

Without loss of generality, we assume that the macroscale pressure variation (p^0) has the following harmonic form:

$$p^0 = \hat{p}^0 \exp(-i\boldsymbol{\kappa}\cdot\mathbf{x}) = \hat{p}^0 \exp[-i(\kappa\boldsymbol{\zeta})\cdot\mathbf{x}], \tag{32}$$

where $\boldsymbol{\kappa}$ denotes the wave vector, which is implicitly dependent on the propagation direction vector $(\boldsymbol{\zeta})$. Substituting the above expression into the definitions of the effective parameters [Eqs. (21a) and (21b)], one can derive the following dispersion equation in the given unit vectors:

$$\omega^2 = \boldsymbol{\kappa}\cdot\left(\frac{\boldsymbol{\rho}_{\text{eff}}^{-1}}{\chi_{\text{eff}}}\right)\cdot\boldsymbol{\kappa} = \kappa^2 c_{\text{eff}}^2 \quad \text{with}$$

$$c_{\text{eff}} = [(\boldsymbol{\zeta}\cdot\boldsymbol{\rho}_{\text{eff}}^{-1}\cdot\boldsymbol{\zeta})/\chi_{\text{eff}}]^{1/2}, \tag{33}$$

where c_{eff} denotes the effective phase speed of the propagating wave. Therefore, the complex wave number and the effective impedance, which are the characteristics of the dissipation phenomena, are usually defined

$$\kappa = \frac{\omega}{c_{\text{eff}}} \quad \text{and} \quad Z_{\text{eff}} = \frac{1}{\phi}\sqrt{\frac{(\boldsymbol{\zeta}\cdot\boldsymbol{\rho}_{\text{eff}}\cdot\boldsymbol{\zeta})}{\chi_{\text{eff}}}}. \tag{34}$$

To calculate acoustic absorption coefficient α, the reflection coefficient r is introduced in terms of the effective impedance (Z_{eff}) and air impedance (Z_0).[5,24,25] Following the cited references, we briefly mention their final results

$$r = \frac{H^* Z_{\text{eff}} - Z_0}{H^* Z_{\text{eff}} + Z_0} \quad \text{with} \quad Z_0 = \rho_0(\gamma P_0/\rho_0)^{1/2} \quad \text{and}$$

$$H = (1 + e^{-2i\omega d/c_{\text{eff}}})/(1 - e^{-2i\omega d/c_{\text{eff}}}), \tag{35}$$

where H is the transfer function over the material thickness d. Finally, one can calculate the acoustic absorption coefficient of the porous media as

$$\alpha = 1 - |r|^2. \tag{36}$$

III. RESULTS AND DISCUSSION

To assess the validity of the described MAM process and our numerical implementation, we first calculated effective parameters using both the direct and hybrid numerical approaches, and then estimated acoustic absorption coefficient curves for the four sphere packing types considered (sc, bcc, fcc, and hcp). For each packing type, Fig. 2 depicts the irreducible unit cells consisting of the fluid-fluid interstitial space between packed spheres. In the implementation of the

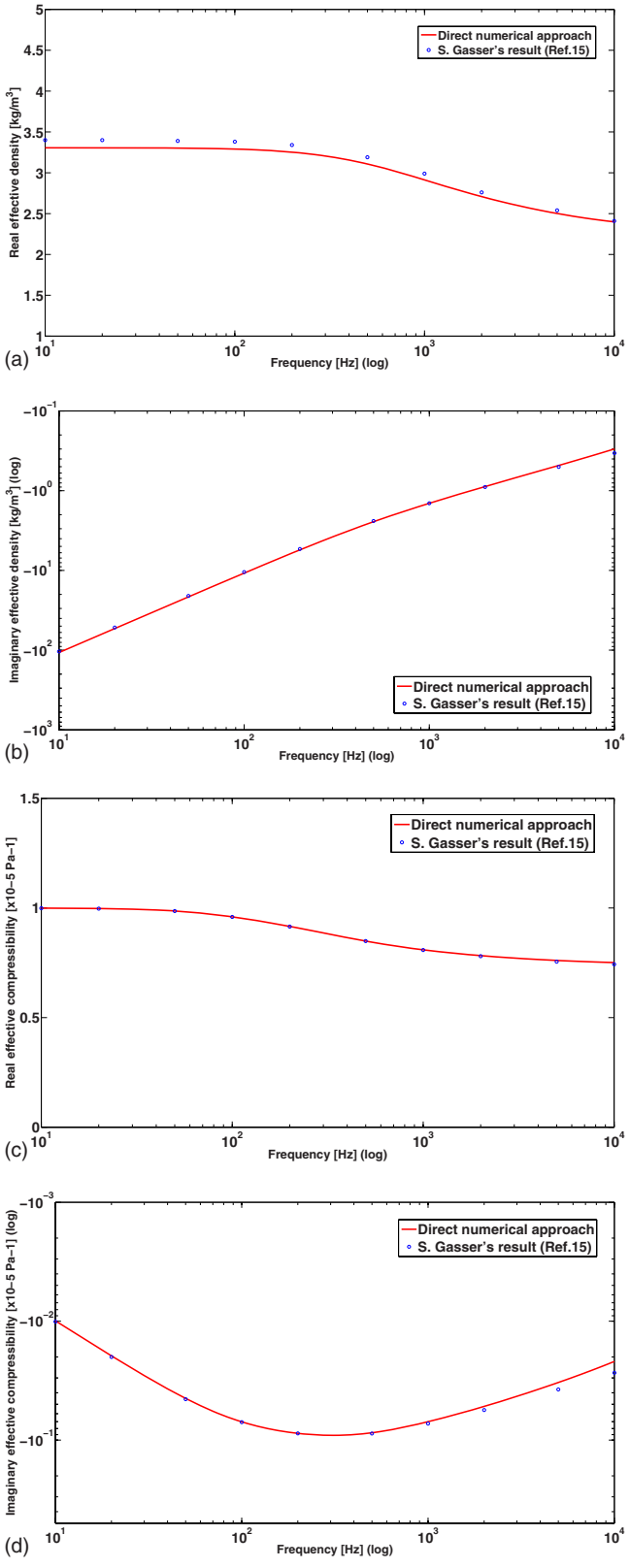


FIG. 4. (Color online) (a) Real and (b) imaginary parts of the effective density (ρ_{eff11}) versus frequency, and (c) real and (d) imaginary parts of the effective compressibility (χ_{eff}) versus frequency.

TABLE II. Geometric and acoustic results for four sphere stackings considered.

	sc			bcc		fcc			hcp
	Ref. 11	Ref. 14	This paper	Ref. 11	This paper	Ref. 11	Ref. 16	This paper	This paper
ϕ	0.48	0.48	0.48	0.32	0.32	0.26	0.26	0.26	0.26
k_0 (kg/m^3)	1.01×10^{-8}	0.984×10^{-9}	$1.02\cdot10^{-8}$	2.01×10^{-9}	1.97×10^{-9}	6.95×10^{-10}	6.83×10^{-10}	6.70×10^{-10}	6.42×10^{-10}
α_0	NA	NA	2.02	NA	2.15	NA	2.63	2.49	2.44
α_∞	1.38	1.48	1.40	1.47	1.48	1.61	1.66	1.65	1.65
Λ (m)	3.24×10^{-4}	NA	3.69×10^{-4}	1.92×10^{-4}	2.34×10^{-4}	1.24×10^{-4}	1.64×10^{-4}	1.59×10^{-4}	1.59×10^{-4}
k_0' (Pa^{-1})	NA	NA	2.46×10^{-8}	NA	3.81×10^{-9}	NA	2.74×10^{-9}	2.70×10^{-9}	2.72×10^{-9}
α_0'	NA	NA	1.43	NA	1.35	NA	1.85	1.85	1.89
Λ' (m)	NA	NA	6.24×10^{-4}	NA	3.25×10^{-4}	NA	2.49×10^{-4}	2.47×10^{-4}	2.47×10^{-4}

approaches, each UFC is subject to translational periodicity boundary conditions in each of the three directions present. Table I provides the coefficients used to obtain the numerical solutions.

As mentioned in Ref. 16, a difficulty is met when one tries to validate the numerical results calculated in both direct and hybrid numerical manners, because most of the published values are limited to a dynamic viscous problem perspective and do not fully provide their results in terms of Eqs. (22) and (24) [or Eqs. (25) and (26)]. Fortunately, all direct and hybrid numerical cases for a fcc sphere packing are directly provided by Gasser[15] and Gasser *et al.*[16] Notably, Gasser[15] documented that computed acoustic properties exhibit good agreement with several experimental results. For these reasons, we first choose the numerical results from Ref. 15 for the validation process of our direct numerical approach, while we compare our hybrid numerical results with Ref. 16, together with Chapman and Higdon[11] for three cubic lattices (sc, bcc, and fcc) and Zhou and Sheng.[14] Moreover, to directly compare our numerical results with those presented in the literature,[15,16] the radius of the sphere is chosen to be 1 mm. As in the cited studies, we also include a solder joint at each sphere contact point, where the solder radius is

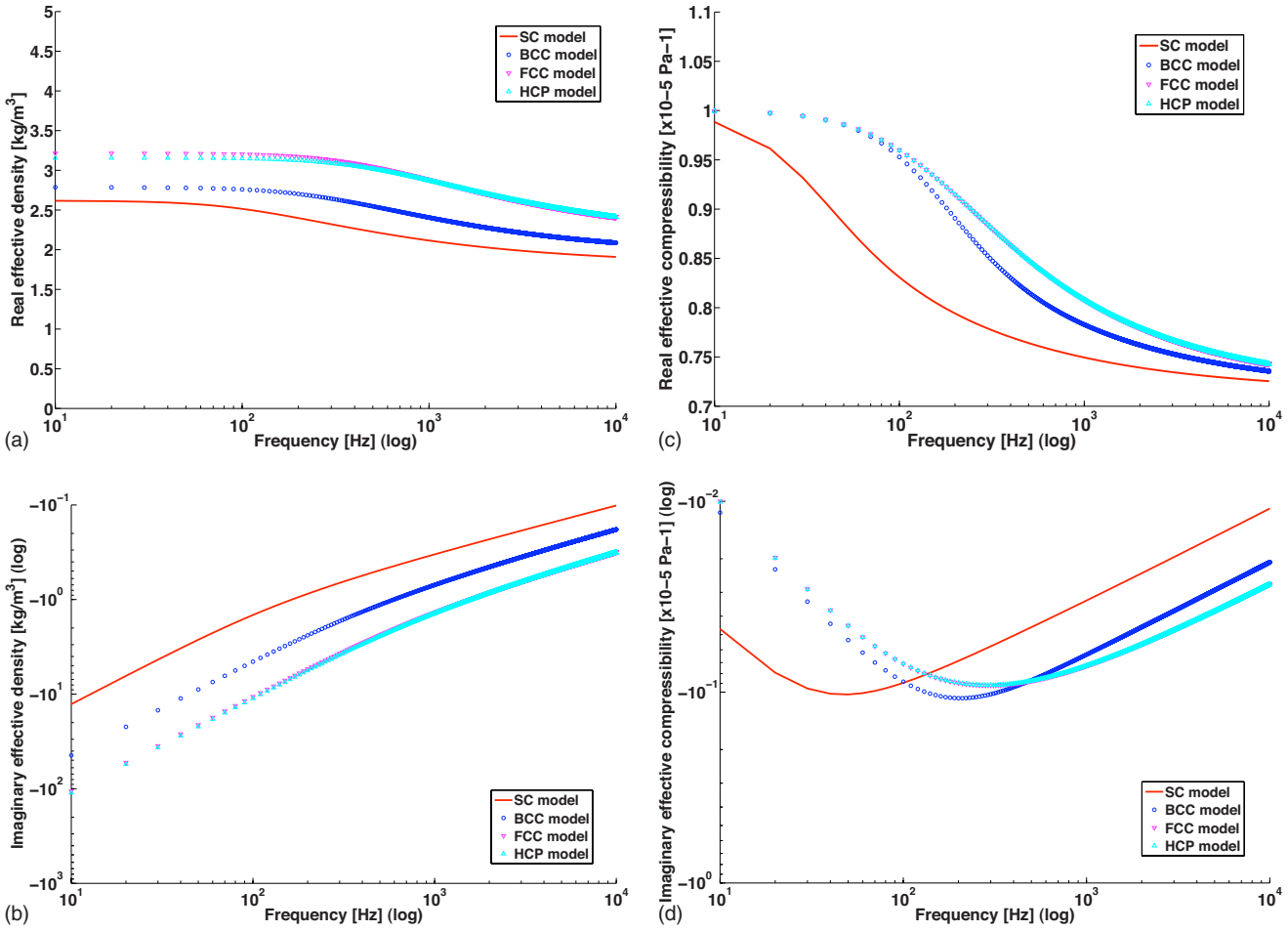


FIG. 5. (Color online) (a) Real and (b) imaginary parts of the effective density (ρ_{eff}) versus frequency, and (c) real and (d) imaginary parts of effective compressibility (χ_{eff}) versus frequency (hybrid numerical approach).

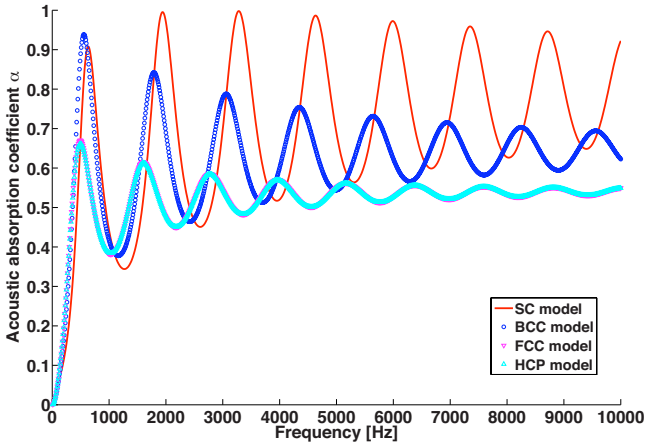


FIG. 6. (Color online) Acoustic absorption coefficients versus frequency.

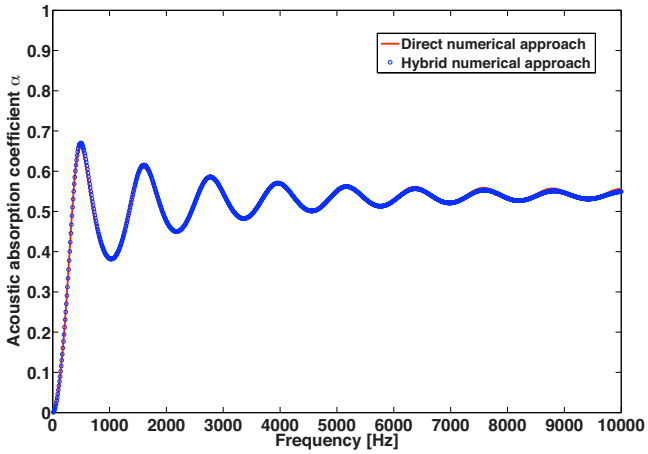


FIG. 7. (Color online) Comparison of acoustic absorption coefficients versus frequency for the direct and hybrid numerical approaches.

150 μm. We only consider results for propagation along the x-axis; however, the formulation is general enough to consider any wave vector.

A. Comparison with direct numerical results of the literature

We first consider the direct numerical process. The micro-scale solutions of the dynamic viscous and thermal equations are calculated from Eqs. (19) and (20). Figure 3 provides some examples of micro-scale dynamic viscous and thermal permeability distributions at 50 Hz (low-frequency regime) and 2500 Hz (high-frequency regime) in the y-z plane.

These figures illustrate that at low-frequency, viscous dissipation and thermal conduction are responsively distributed along the boundary of each sphere, while at the high-frequency, dense fluid acceleration and heat exchange phenomena occur in the vicinity of solder joints. Moreover, we iteratively compute the effective density (22) and compressibility (24) parameters in a frequency range of interest (here, from 10 to 10 000 Hz), as shown in Fig. 4. For validation purposes, numerical points provided in the literature[15] are also plotted. As can be seen in the subfigures, the frequency-dependent effective parameters obtained from our approach are in very good agreement with those provided by Ref. 15 over the entire range of frequencies considered.

B. Comparison with hybrid numerical results of the literature

The results compared with those available in the literature are shown in Table II. As expected, these results show good agreement between those published and our numerically computed results. However, concerning viscous characteristic length, there is a disagreement. As mentioned in Ref. 16, this disagreement can be explained by the presence of solder joints. Using the estimated numerical parameters in Table II as the input data for Eq. (28), we obtain the effective density and compressibility parameters in a frequency range of interest (here, from 10 to 10 000 Hz), as shown in Fig. 5. Figure 6 illustrates the acoustic absorption coefficients for four sphere packings with a fixed material thickness d =0.1 m. As expected, the general shapes of the acoustic absorption curves for the fcc and hcp models are nearly identical since they have similar closed-packed structures (see Table II). The sc model has the largest static permeability as compared to the other packing arrangements due to large wetted areas, and as a result, demonstrates the most potential for acoustic absorption.

C. Comparison between direct and hybrid numerical results

We next compare our results for frequency-dependent acoustic absorption, in the FCC sphere packing case, using both the direct and hybrid numerical approaches. As can be seen in Fig. 7, the frequency-dependent acoustic absorption coefficients obtained from both approaches are in very good agreement in the frequency range of interest. However, as mentioned before, the hybrid numerical approach offers several advantages in terms of computational efficiency and convergence stability.

IV. CONCLUSION

In conclusion, we have presented a consistent and general approach for numerically computing frequency-dependent effective density and compressibility parameters for periodic porous materials using both direct and hybrid numerical approaches. These parameters ultimately characterize the acoustic absorption behavior of the absorbing material. The formulation is cast in a form suitable for incorporation into a COTS finite element analysis package, and as such, avoids the complication of working with and coupling multiple in-house codes. Comparisons of results generated in this study with published numerical data demonstrate very good agreement over all frequencies considered.

ACKNOWLEDGMENT

The authors gratefully acknowledge the EADS N.A. Foundation's support of this work.

[1]G. Kirchhoff, "Über den einfluß der wärmeleitung in einem gase auf die schallbewegungv (On the influence of heat conduction in a gas on sound propagation)," Ann. Phys. Chem. **134**, 177–193 (1868).
[2]K. Attenborough, "Acoustical characteristics of porous materials," Phys.

Rep. **82**, 179–227 (1982).
[3]R. F. Lambert, "Propagation of sound in highly porous open cells elastic forms," J. Acoust. Soc. Am. **73**, 1131–1138 (1983).
[4]K. Attenborough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1987).
[5]J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier Applied Science, Amsterdam, 1993).
[6]J. G. Berryman, "Comparison of upscaling methods in poroelasticity and its generalizations," J. Eng. Mech. **131**, 928–936 (2005).
[7]D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[8]Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air-saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).
[9]S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces of porous medium acoustics," Phys. Rev. B **47**, 4964–4978 (1993).
[10]D. Lafarge, "Propagation du son dans les matériaux poreux à structure rigide saturés par un fluide viscothermique (Sound propagation in rigid porous media saturated by a visothermal fluid)," Ph.D. thesis, Université du Maine, France (1993).
[11]A. M. Chapman and J. J. L. Higdon, "Oscillatory Stokes flow in periodic porous media," Phys. Fluids A **4**, 2099–2116 (1992).
[12]L. Borne, "Harmonic Stokes flow through periodic porous media: A 3D boundary element method," J. Comput. Phys. **99**, 214–232 (1992).
[13]K. Schladitz, S. Peters, D. Reinel-Bitzer, A. Wiegmann, and J. Ohser, "Design of acoustic trim based on geometric modeling and flow simulation for non-woven," Comput. Mater. Sci. **38**, 56–66 (2006).
[14]M. Y. Zhou and P. Sheng, "First-principles calculations of dynamic permeability in porous media," Phys. Rev. B **39**, 12027–12039 (1989).
[15]S. Gasser, "Etude des propriétés acoustiques et mécaniques d'un matériau métallique poreux modèle à base de spheres creuses de nickel (Study of the acoustical and mechanical properties of a model porous metallic material made of hollow nickel sphere)," Ph.D. thesis, INPG, France (2003).
[16]S. Gasser, F. Paun, and Y. Bréchet, "Absorptive properties of rigid porous media: Application to face centered cubic sphere packing," J. Acoust. Soc. Am. **117**, 2090–2099 (2005).
[17]C. Perrot, F. Chevillotte, and R. Panneton, "Dynamic viscous permeability of an open-cell aluminum foam: Computations versus experiments," J. Appl. Phys. **103**, 024909 (2008).
[18]C. Perrot, F. Chevillotte, and R. Panneton, "Bottom-up approach for microstructure optimization of sound absorbing materials," J. Acoust. Soc. Am. **124**, 940–948 (2008).
[19]E. Sanchez-Palencia, *Non-Homogeneous Media and Vibration Theory*, Lecture Notes in Physics Vol. No. **127** (Springer-Verlag, Berlin, 1980), Pt. II.
[20]A. Bensoussan, J. L. Lions, and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures* (North-Holland, Amsterdam, 1978).
[21]J. L. Auriault, "Dynamic behaviour of a porous media saturated by a Newtonian fluid," Int. J. Eng. Sci. **18**, 775–785 (1980).
[22]J. L. Auriault, L. Borne, and R. Chambon, "Dynamics of porous saturated media, checking of the generalized law of Darcy," J. Acoust. Soc. Am. **77**, 1641–1650 (1985).
[23]C. Boutin, P. Royer, and J. L. Auriault, "Acoustic absorption of porous surfacing with dual porosity," Int. J. Solids Struct. **35**, 4709–4737 (1998).
[24]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949).
[25]N. Auffray, R. Bouchet, and Y. Bréchet, "Modèles d'acoustique phénoménologiques—Application à la conception de matériau sur mesure (Phenomenological implementation of acoustic models in the design of tailor material)," 18th Congrès Français de Mécanique, Grenoble (2007).
[26]D. Lafarge, P. Lemarinier, and J. F. Allard, "Dynamic compressibility of air in porous structures at audible frequencies," J. Acoust. Soc. Am. **102**, 1995–2006 (1997).
[27]S. Torquato, "Relationship between permeability and diffusion-controlled trapping constant of porous media," Phys. Rev. Lett. **64**, 2644–2646 (1990).
[28]R. J. S. Brown, "Connection between formation factor for electrical resistivity and fluid-solid coupling factor in Biot equations for acoustic waves in fluid-filled porous media," Geophysics **45**, 1269–1275 (1980).

Predicting the sound insulation of single leaf walls: Extension of Cremer's model

John L. Davy[a)]
School of Applied Sciences, RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, Australia

(Received 2 February 2009; revised 8 July 2009; accepted 27 July 2009)

In his 1942 paper on the sound insulation of single leaf walls, Cremer [(1942). Akust. Z. **7**, 81–104] made a number of approximations in order to show the general trend of sound insulation above the critical frequency. Cremer realized that these approximations limited the application of his theory to frequencies greater than twice the critical frequency. This paper removes most of Cremer's approximations so that the revised theory can be used down to the critical frequency. The revised theory is used as a correction to the diffuse field limp panel mass law below the critical frequency by setting the nonexistent coincidence angle to 90°. The diffuse field limp panel mass law for a finite size wall is derived without recourse to a limiting angle by following the average diffuse field single sided radiation efficiency approach. The shear wave correction derived by Heckl and Donner [(1985). Rundfunktech Mitt. **29**, 287–291] is applied to the revised theory in order to cover the case of thicker walls. The revised theory predicts the general trend of the experimental data, although the agreement is usually worse at low frequencies and depends on the value of damping loss factor used in the region of and above the critical frequency.
 [DOI: 10.1121/1.3206582]

PACS number(s): 43.55.Rg, 43.55.Ti, 43.40.Rj, 43.20.Rz [LMW] Pages: 1871–1877

I. INTRODUCTION

Cremer's (1942) theory for the sound insulation of thin walls above the critical frequency is still in widespread use today. A number of other authors have also derived methods for predicting the sound insulation of thin walls. Important contributions include those of Crocker and Price (1969), Sewell (1970), Leppington *et al.* (1987), and Villot *et al.* (2001).

A critical examination of Cremer's (1942) theory shows that Cremer (1942) made a number of approximations which limit the application of his theory to frequencies which are greater than twice the critical frequency. Cremer (1942) stated this limitation in his paper. The last of Cremer's (1942) approximations was to assume that the single side radiation efficiency of a panel above the critical frequency was unity. This enables Cremer's (1942) theory to be used at and just below the critical frequency, although most of the approximations are obviously incorrect in this frequency region. It is necessary to use Cremer's (1942) approximate theory at and just below the critical frequency because Sewell's (1970) approximate correction factor for just below the critical frequency predicts infinite transmission at the critical frequency.

In this paper, most of Cremer's (1942) approximations are removed and the single sided radiation efficiency of an infinite panel is replaced with that for a finite panel. This produces a theory which is more exact and which can be used down to the critical frequency.

Below the critical frequency, the need to use a limiting angle is avoided by following the average diffuse field single sided radiation efficiency approach. Instead of using Sewell's (1970) approximate correction factor for just below the critical frequency, the more exact version of Cremer's (1942) theory for above the critical frequency, which is developed in this paper, is added to the theory below the critical frequency. This is made possible by setting the coincidence angle, which does not exist below the critical frequency, equal to 90°. A major advantage of this approach is that there is only a very slight discontinuity at the critical frequency.

In order to account for the transition from bending waves to shear waves which occurs in thick panels at high frequencies, this paper follows Ljunggren's (1991) approach and uses the shear wave correction factor developed by Heckl and Donner (1985).

II. ABOVE THE CRITICAL FREQUENCY

The sound transmission coefficient $\tau(\theta)$ of a wall is the ratio of the sound energy transmitted by the wall to the sound energy incident upon the wall. For an infinite, isotropic, uniform thickness plane wall the sound transmission coefficient of a plane wave depends on the angle θ between the direction of propagation of the incident plane wave and the normal to the plane of the wall. To evaluate the diffuse field sound transmission coefficient τ_d it is necessary to average the plane wave sound transmission coefficient $\tau(\theta)$ with appropriate weighting across all angles of incidence [Cremer (1942), Eq. (4.9)],

$$\tau_d = 2\int_0^{\pi/2} \tau(\theta)\cos\,\theta\,\sin\,\theta d\theta. \quad (1)$$

The cos θ term is the cross-sectional area of the plane sound wave that is incident on a unit area of the wall at an angle of incidence of θ to the normal to the wall. The sin θ term is due to the fact that the annulus of solid angle between θ and

a)Author to whom correspondence should be addressed. Electronic mail: john.davy@rmit.edu.au. Also at CSIRO Materials Science and Engineering, P.O. Box 56, Highett, Victoria 3190, Australia.

$\theta+\Delta\theta$ is $2\pi \sin\theta\Delta\theta$. The 2 term is a normalization factor which arises from the fact that τ_d must be 1 when $\tau(\theta)$ is 1 for all values of θ. Equation (1) can be rewritten in a number of forms. Use will be made of the following form [Cremer (1942), Eq. (4.9)]:

$$\tau_d = \int_0^1 \tau(\theta) d(\cos^2\theta). \tag{2}$$

For a thin plane wall with the properties described above Cremer's (1942) Eq. (4.8) is

$$\tau(\theta) = \frac{1}{\left|1 + \frac{Z(\theta)\cos\theta}{2\rho_0 c}\right|^2}, \tag{3}$$

where $Z(\theta)$ is the bending wave impedance of the wall and $\rho_0 c$ is the characteristic impedance of air, being the product of the ambient density ρ_0 and the speed of sound in air c. For an infinite panel, the single sided radiation efficiency $\sigma(\theta)$ is

$$\sigma(\theta) = \frac{1}{\cos\theta}. \tag{4}$$

Thus Eqs. (1) and (3) can be rewritten as

$$\tau_d = 2\int_0^{\pi/2} \frac{\tau(\theta)}{\sigma(\theta)} \sin\theta d\theta \tag{5}$$

and

$$\tau(\theta) = \frac{1}{\left|1 + \frac{Z(\theta)}{2\rho_0 c\sigma(\theta)}\right|^2}. \tag{6}$$

Cremer's (1942) Eq. (9.2) can be written as

$$Z(\theta) = m\omega[j(1 - r^2\sin^4\theta) + \eta r^2 \sin^4\theta], \tag{7}$$

where

$$r = \frac{\omega}{\omega_c} = \frac{f}{f_c}. \tag{8}$$

$f_c(\omega_c)$ is the (angular) critical frequency of the panel, $f(\omega)$ is the (angular) frequency of the sound, and m is the mass per unit area of the single leaf panel. η is the total damping loss factor of the single leaf panel which is equal to the sum of the internal damping loss factor of the panel η_{int}, the damping loss factor η_{edge} due to the transmission of vibrational energy from the panel to its surrounding elements at its edges and twice (to take account of both sides of the panel) its single sided radiation loss factor η_{rad}. Thus

$$\eta = \eta_{\text{int}} + \eta_{\text{edge}} + 2\eta_{\text{rad}}. \tag{9}$$

The single sided radiation loss factor is related to the single sided radiation efficiency by

$$\eta_{\text{rad}} = \frac{\sigma(\theta)\rho_0 c}{m\omega}. \tag{10}$$

Annex C of ISO 15712-1:2005(E) (ISO, 2005) gives guidance on the calculation of the total damping loss factor. For the laboratory situation of a panel with a surface density m of less than 800 kg/m^2, it says that

$$\eta = \eta_{\text{int}} + \eta_{\text{edge}} = \eta_{\text{int}} + \frac{m}{485\sqrt{f}}, \tag{11}$$

where η_{int} is the internal loss factor of the panel material and can normally be taken as 0.01. Equation (11) ignores the radiation loss factor because it is usually insignificant.

Inserting Eq. (7) into Eq. (6) produces Cremer's (1942) Eq. (9.3) for the sound transmission coefficient $\tau(\theta)$ of a single leaf panel as a function of angle of incidence θ.

$$\tau(\theta) = \frac{1}{\left(1 + \frac{a\eta r^2 \sin^4\theta}{\sigma(\theta)}\right)^2 + \left(\frac{a}{\sigma(\theta)}\right)^2 (1 - r^2\sin^4\theta)^2}, \tag{12}$$

where

$$a = \frac{\omega m}{2\rho_0 c} = \frac{\pi f m}{\rho_0 c}. \tag{13}$$

For a given frequency which is greater than or equal to the critical frequency, the maximum value of Eq. (12) occurs at the coincidence angle θ_c where

$$\sin^2\theta_c = \frac{1}{r} = \frac{\omega_c}{\omega} = \frac{f_c}{f}. \tag{14}$$

For values of θ which are close to θ_c, Eq. (12) can be approximated by setting most of the values of θ which occur in Eq. (12) equal to θ_c. This gives

$$\tau(\theta) \approx \frac{1}{\left(1 + \frac{a\eta r^2 \sin^4\theta_c}{\sigma(\theta_c)}\right)^2 + \left(\frac{a}{\sigma(\theta_c)}\right)^2 (1 - r\sin^2\theta)^2(1 + r\sin^2\theta_c)^2} = \frac{1}{\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)^2 + \left(\frac{2ar}{\sigma(\theta_c)}\right)^2 \left(\frac{1}{r} - \sin^2\theta\right)^2}. \tag{15}$$

Put

$$x = \cos^2\theta. \tag{16}$$

Then

$$\sin^2\theta = 1 - x, \tag{17}$$

and

$$\tau(\theta) = \frac{1}{\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)^2 + \left(\frac{2ar}{\sigma(\theta_c)}\right)^2 \left(x + \frac{1}{r} - 1\right)^2}. \tag{18}$$

Putting Eq. (18) into Eq. (2) gives the diffuse field sound transmission coefficient as

$$\tau_d = \int_0^1 \frac{dx}{\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)^2 + \left(\frac{2ar}{\sigma(\theta_c)}\right)^2 \left(x + \frac{1}{r} - 1\right)^2}. \tag{19}$$

Put

$$y = x + \frac{1}{r} - 1, \tag{20}$$

then

$$dy = dx, \tag{21}$$

and

$$\tau_d = \int_{1/r-1}^{1/r} \frac{dy}{\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)^2 + \left(\frac{2ar}{\sigma(\theta_c)}\right)^2 y^2}. \tag{22}$$

This paper departs from Cremer (1942) by not approximating by extending the range of integration to cover from $-\infty$ to ∞. Using integral number 2.124.1 on page 60 of Gradshteyn and Ryzhik (1965),

$$\tau_d = \left[\frac{1}{\frac{2ar}{\sigma(\theta_c)}\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)} \arctan\left(\frac{2arx}{\sigma(\theta_c) + a\eta}\right) \right]_{1/r-1}^{1/r}. \tag{23}$$

Thus

$$\tau_d = \frac{\sigma^2(\theta_c)}{2ar(\sigma(\theta_c) + a\eta)} \left\{ \arctan\left[\frac{2a}{\sigma(\theta_c) + a\eta}\right] - \arctan\left[\frac{2a(1-r)}{\sigma(\theta_c) + a\eta}\right] \right\}, \tag{24}$$

or

$$\tau_d = \frac{\left(\frac{\sigma(\theta_c)}{a}\right)^2}{2r\left(\frac{\sigma(\theta_c)}{a} + \eta\right)} \left\{ \arctan\left[\frac{2}{\frac{\sigma(\theta_c)}{a} + \eta}\right] - \arctan\left[\frac{2(1-r)}{\frac{\sigma(\theta_c)}{a} + \eta}\right] \right\}. \tag{25}$$

This paper will use Eq. (24) or Eq. (25). However, for comparison Cremer's (1942) approximations will now be derived. The integrand in Eq. (22) is a maximum when $y=0$. If $a\eta/\sigma(\theta) \gg 1$, which is usually the case, the integrand is half its maximum value when $|y| = \eta/2r$. Since η is usually very much less than 1 and r is greater than or equal to 1 if the frequency is greater than or equal to the critical frequency, the values of y where the integrand is significantly different from zero usually lie well inside the integral limits from $1/r-1$ to $1/r$. Because of this Cremer (1942) approximated the integral in Eq. (22) by extending the limits of integration from $-\infty$ to $+\infty$.

$$\tau_d \approx \int_{-\infty}^{\infty} \frac{dy}{\left(1 + \frac{a\eta}{\sigma(\theta_c)}\right)^2 + \left(\frac{2ar}{\sigma(\theta_c)}\right)^2 y^2}. \tag{26}$$

With this approximation Eqs. (24) and (25) become

$$\tau_d \approx \frac{\pi\sigma^2(\theta_c)}{2ar(\sigma(\theta_c) + a\eta)}, \tag{27}$$

and

$$\tau_d \approx \frac{\pi\left(\frac{\sigma(\theta_c)}{a}\right)^2}{2r\left(\frac{\sigma(\theta_c)}{a} + \eta\right)}. \tag{28}$$

Equation (27) is the same as Cremer's (1942) Eq. (9.6). Cremer (1942) also assumed the usual case of $a\eta/\sigma \gg 1$, which gives

$$\tau_d \approx \frac{1}{a^2}\frac{\pi}{2\eta}\frac{\sigma^2(\theta_c)}{r}. \tag{29}$$

Because Cremer (1942) assumed an infinite panel above the critical frequency, he also used the radiation efficiency of free bending waves above the critical frequency for an infinite panel. This is because the wavelength of the forced waves at coincidence is equal to the free bending wavelength.

$$\sigma(\theta_c) = \frac{1}{\sqrt{1 - \frac{1}{r}}} = \frac{1}{\sqrt{1 - \frac{\omega_c}{\omega}}} = \frac{1}{\sqrt{1 - \frac{f_c}{f}}} = \frac{1}{\cos\,\theta_c}. \tag{30}$$

Substituting Eq. (30) into Eq. (29) gives Cremer's (1942) Eq. (9.8).

$$\tau_d \approx \frac{1}{a^2}\frac{\pi}{2\eta}\frac{1}{r-1}. \tag{31}$$

When $r>2$, Cremer (1942) approximated Eq. (31) as

$$\tau_d \approx \frac{1}{a^2}\frac{\pi}{2\eta}\frac{1}{r}. \tag{32}$$

This last approximation of Cremer (1942) is equivalent to assuming that the radiation efficiency of a panel above its critical frequency is equal to unity. Equation (32) is a version of Cremer's (1942) Eq. (9.10).

At the critical frequency, $r=1$ and the lower limit of integration in Eq. (22) is zero. Because of the symmetrical nature of the integrand about zero, Cremer's (1942) extension of the limits of integration produces a value of the transmission coefficient at the critical frequency which is more than twice as large as the value before the extension of the limits of integration. Thus in this paper, Eq. (24) or Eq. (25)

will be used instead of Cremer's (1942) further approximations which are given by Eqs. (27)–(29), Eq. (31), or Eq. (32).

Because Eq. (30) gives an infinite value for the radiation efficiency at the critical frequency, this paper uses Davy's (2004) theory to calculate the radiation efficiency. The equations given here are an updated version of those in Davy (2004). First the cosine of the coincidence angle is calculated. Since Eq. (24) or Eq. (25) is going to be used as a correction term below the critical frequency, the cosine of the coincidence angle is set to zero for frequencies below the critical frequency.

$$g = \begin{cases} \cos\ \theta_c = \sqrt{1 - \dfrac{\omega_c}{\omega}} = \sqrt{1 - \dfrac{f_c}{f}} & \text{if } \omega \geq \omega_c \\ 0 & \text{if } \omega < \omega_c. \end{cases} \tag{33}$$

Then three empirical constants are defined. These empirical constants are chosen to make Eq. (39) agree as closely as possible with numerical calculations by Sato (1973). Sato's numerical results appear as Fig. 1.3.2, Table B.1, and Fig. B.2 in Rindel (1975). The three empirical constants are $n=2$, $w=1.3$, and $\beta=0.124$.

The length of the side of an equivalent square panel is defined.

$$2a = \frac{4S}{U}, \tag{34}$$

where S is the area of the panel and U is the perimeter of the panel.

Some intermediate values are calculated using the wave number $k=\omega/c=2\pi f/c$.

$$p = \begin{cases} w\sqrt{\dfrac{\pi}{2ka}} & \text{if } w\sqrt{\dfrac{\pi}{2ka}} \leq 1 \\ 1 & \text{if } w\sqrt{\dfrac{\pi}{2ka}} > 1. \end{cases} \tag{35}$$

$$h = \frac{1}{\dfrac{2}{3}\sqrt{\dfrac{2ka}{\pi}} - \beta}. \tag{36}$$

$$\alpha = \frac{h}{p} - 1. \tag{37}$$

$$q = \frac{2\pi}{k^2 S}. \tag{38}$$

Finally the radiation efficiency is calculated.

$$\sigma(\theta_c) = \begin{cases} \dfrac{1}{\sqrt[n]{g^n + q^n}} & \text{if } 1 \geq g \geq p \\ \dfrac{1}{\sqrt[n]{(h - \alpha g)^n + q^n}} & \text{if } p > g \geq 0. \end{cases} \tag{39}$$

III. BELOW THE CRITICAL FREQUENCY

The sound transmission coefficient below the critical frequency is calculated using the average diffuse field single sided radiation efficiency approach. Bending stiffness is ignored by setting r equal to zero in Eq. (7). Bending stiffness will be included later on by adding in Eq. (24) or Eq. (25). Equation (7) becomes

$$Z(\theta) = jm\omega. \tag{40}$$

Substituting Eq. (40) into Eq. (6), and assuming that the second term in the modulus brackets is much greater than 1, gives

$$\tau(\theta) = \left(\frac{2\rho_0 c\sigma(\theta)}{m\omega}\right)^2 = \left(\frac{\sigma(\theta)}{a}\right)^2. \tag{41}$$

Substituting Eq. (41) into Eq. (5) gives

$$\tau_d = \frac{2}{a^2}\int_0^{\pi/2} \sigma(\theta)\sin\ \theta d\theta = \frac{2\langle\sigma\rangle}{a^2}, \tag{42}$$

where

$$\langle\sigma\rangle = \int_0^{\pi/2} \sigma(\theta)\sin\ \theta d\theta. \tag{43}$$

Substituting Eq. (39) in Eq. (43) gives

$$\langle\sigma\rangle = \ln\left(\frac{1 + \sqrt{1+q^2}}{p + \sqrt{p^2+q^2}}\right) + \frac{1}{\alpha}\ln\left(\frac{h + \sqrt{h^2+q^2}}{p + \sqrt{p^2+q^2}}\right). \tag{44}$$

An earlier version of this equation is given in Davy (2004).

To include the effects of bending stiffness, the sound transmission coefficient below the critical frequency is calculated as the sum of Eq. (42) and Eq. (24) or Eq. (25). The use of Eq. (24) or Eq. (25) as the correction for bending stiffness below the critical frequency is a new approach introduced by this paper. Above the critical frequency, only Eq. (24) or Eq. (25) is used. Equation (42) can be included immediately above the critical frequency, but it causes problems if used a long way above the critical frequency.

Because of the rapid variation in the sound transmission coefficient with frequency near the critical frequency, if the measurement frequency band includes the critical frequency, the sound transmission coefficient is averaged over the measurement frequency band. In this paper for comparison with measurements in the third octave band which included the critical frequency, the theoretical values were averaged over the three frequencies which were $2^{-1/9}$, 1, and $2^{1/9}$ times the center frequency of the third octave band.

IV. SHEAR WAVE CORRECTION

To account for the transition from bending waves to shear waves, Heckl and Donner's correction (Ljunggren, 1991) is applied to Eqs. (24) and (25). According to Ljunggren (1991),

> "the Mindlin plate theory was applied to the case of an infinite wall by Heckl and Donner in the same way as the simple bending wave theory was applied in Cremer's paper from 1942.

In these equations, k_M is the wave number at coincidence, that is, the "corrected" wave number of the free bending wave [see for example, Cremer *et al.* (1988), p. 109], k_T is the wave number of a hypothetical, corrected shear wave,

$$k_T^2 = \frac{\omega^2 \rho}{G^*}, \tag{45}$$

where G^* is the shear modulus, modified to account for shear distribution, and

$$k_S^2 = k_T^2 + k_L^2. \tag{46}$$

The wave number k_B is … the wave number of the free bending wave according to the thin plate theory. ρ is the density of the plate material."

k_L is the wave number of the quasilongitudinal wave. G^* is calculated from the following expressions [Magrab (1979), p. 281]:

$$G^* = \frac{G}{\chi}, \tag{47}$$

$$\chi = \left(\frac{1+\upsilon}{0.87+1.12\upsilon} \right)^2, \tag{48}$$

where G is the shear modulus and υ is Poisson's ratio.

$$A = \left[1 + \frac{\delta^2}{12} \left(\frac{k_M^2 k_T^2}{k_L^2} - k_T^2 \right) \right]^2, \tag{49}$$

$$B = 1 - \frac{(k_T \delta)^2}{12} + \frac{k_M^2 k_S^2}{k_B^4}, \tag{50}$$

$$C = \sqrt{1 - \frac{(k_T \delta)^2}{12} + \frac{k_S^4}{4k_B^4}}, \tag{51}$$

where δ is the thickness of the panel. Equations (24) and (25) are multiplied by the correction factor

$$\frac{A}{BC}. \tag{52}$$

k_M^2 is the positive solution of the following quadratic equation (Cremer *et al.*, 1988):

$$\frac{E}{(1-\upsilon^2)\rho} \frac{\delta^2}{12} k_M^4 - \omega^2 \left(1 + \frac{2\chi}{1-\upsilon} \right) \frac{\delta^2}{12} k_M^2 - \omega^2 = 0, \tag{53}$$

where E is Young's modulus.

The uncorrected shear modulus is

$$G = \frac{E}{2(1+\upsilon)}. \tag{54}$$

k_B^4 is given by

$$k_B^4 = \frac{12\omega^2 \rho (1-\upsilon^2)}{E\delta^2}. \tag{55}$$

k_L^2 is given by

$$k_L^2 = \frac{\omega^2 \rho (1-\upsilon^2)}{E}. \tag{56}$$

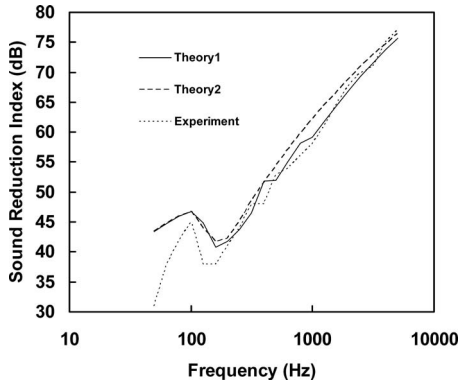


FIG. 1. The sound insulation of 200 mm thick concrete. The experimental results are from Ljunggren (1991). The theory1 curve is calculated using the *in-situ* measured total damping loss factor. The theory2 curve is calculated using the total damping loss factor given by Eq. (11).

m, ρ, and δ are related by $m = \rho\delta$. The angular critical frequency is given by

$$\omega_c = \frac{c^2}{\delta} \sqrt{\frac{12\rho(1-\upsilon^2)}{E}}. \tag{57}$$

This equation can be inverted to calculate Young's modulus from the angular critical frequency. If Poisson's ratio is not known, assume that it is 0.3. Most materials have Poisson's ratios that are close to 0.3.

V. COMPARISON WITH EXPERIMENT

At and above the critical frequency, the theoretical values were calculated using Eq. (24) or Eq. (25) multiplied by Eq. (52). Below the critical frequency Eq. (44) was also added to obtain the final theoretical result. Figure 1 compares the experimental sound insulation of a 200 mm thick concrete wall (Ljunggren, 1991) with the theory described in this paper. There are two theoretical curves. The theory1 curve is calculated using the *in-situ* measured total damping loss factor, while the theory2 curve is calculated using the total damping loss factor given by Eq. (11). The theoretical calculations assume that the density is 2300 $\mathrm{kg/m^3}$, Young's modulus is 1.36×10^{10} Pa, and Poisson's ratio is 0.3. The specimen size is 4.08×3.08 $\mathrm{m^2}$. Below 180 Hz, the experimental measurements are substantially lower than the theoretical results. It should be noted that the experimental uncertainties are larger in this frequency region and that Ljunggren (1991) did not attempt to compare theory and experiment below 315 Hz. Above 180 Hz, not surprisingly, the theory1 results, which are calculated using the measured total damping loss factor, are in better agreement with the experimental results than the theory2 results which use predicted total damping loss factors. Nevertheless, the theory2 results still give a good estimate of the general trend of the experimental results.

Figure 2 compares the theory with the measured sound insulation of a single layer of 13 mm gypsum plaster board. The experimental results were measured by the National Research Council of Canada (NRCC) (Northwood, 1968; Dupree, 1981). One of the measurements is on a wall with no studs, while the other two measurements have wooden studs

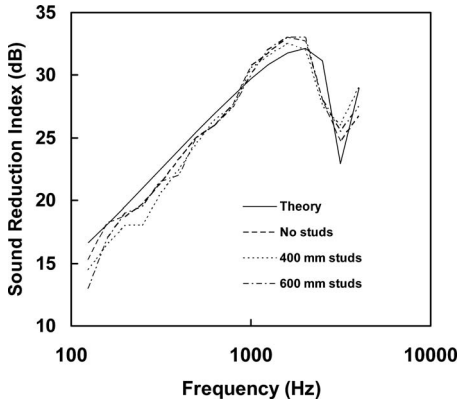


FIG. 2. The sound insulation of 13 mm gypsum plasterboard without studs and with studs spaced at 400 and 600 mm centers. The experimental results were measured by the NRCC (Northwood, 1968; Dupree, 1981).

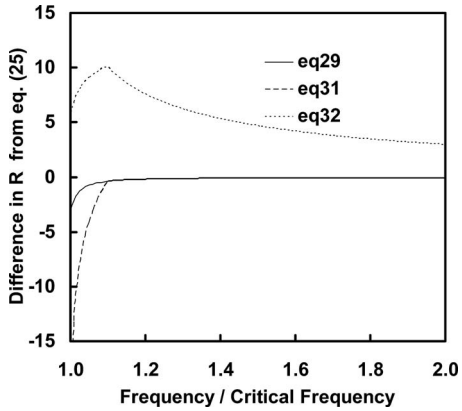


FIG. 4. The difference in decibels of the sound reduction index R of 6 mm monolithic glass calculated using Eq. (29), Eq. (31), or Eq. (32) compared to that calculated using Eq. (25).

spaced at 400 and 600 mm, respectively. The no stud results are the average of three separate measurements. The theoretical calculations assume that the density is 770 kg/m^3, Young's modulus is 1.85×10^9 Pa, Poisson's ratio is 0.3, and the total damping loss factor is 0.04. This value of the total damping loss factor is chosen to give the best agreement between theory and experiment. It is slightly greater than the usually accepted range for gypsum plaster board from 0.01 to 0.03. The specimen size used in the calculations is 3.05×2.44 m^2. The experimental results show that the wooden studs do not make any significant difference, while the theory slightly but significantly overestimates the experimental results in the lower frequency range.

Figure 3 compares theory with three measurements of the sound insulation of 6 mm monolithic glass by Monsanto, Pilkington, and the NRCC (Quirt, 1981, 1982). The theoretical calculations assume that the density is 2500 kg/m^3, Young's modulus is 6.5×10^{10} Pa, Poisson's ratio is 0.22, and the total damping loss factor is 0.05. Again, this value of the total damping loss factor is chosen to give the best agreement between theory and experiment. A specimen size measuring 1.8×1.2 m^2 is assumed for the calculations. Although the total damping loss factor seems high for glass, it should be noted that Cremer (1942) assumed a damping loss factor of 0.1. The author's experience is that it is often necessary to use higher damping loss factors than would be measured directly in order to make theoretical predictions and experimental measurements of diffuse field sound insulation and the directivity of sound insulation agree. When the scatter of the experimental results is taken into account the agreement between theory and experiment is good.

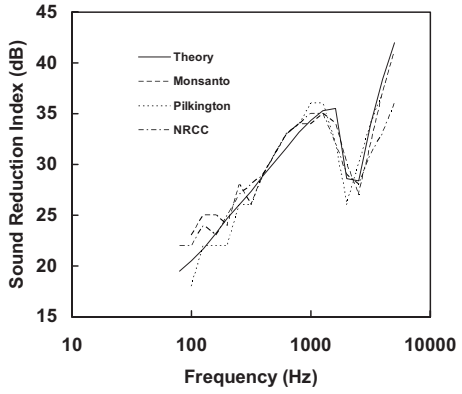


FIG. 3. The sound insulation of 6 mm monolithic glass. The experimental results were measured by Monsanto, Pilkington, and the NRCC (Quirt, 1981, 1982).

Figure 4 shows the difference in decibels of the sound reduction index R calculated using Eq. (29), Eq. (31), or Eq. (32) compared to that calculated using Eq. (25). These calculations are for the specimen shown in Fig. 3 since the results depend on the properties of the specimen. At the critical frequency, the result calculated using Eq. (29) is 3 dB less than the result calculated using Eq. (25). The result calculated using Eq. (31) is $-\infty$ dB at the critical frequency.

VI. CONCLUSIONS

This paper develops a more exact version of Cremer's (1942) theory of sound insulation which allows Cremer's (1942) theory to be used down to the critical frequency and as a correction below the critical frequency. The approach adopted in this paper avoids large discontinuities in the region of the critical frequency. This more exact version of the theory agrees reasonably well with the experimental data presented, although there are some systematic differences. The use of the single sided forced radiation efficiency for a finite sized panel below the critical frequency avoids the need to use a limiting angle of integration. Applying Heckl and Donner's correction (Ljunggren, 1991) for the transition from bending to shear waves works well for the case of a thicker 200 mm concrete wall.

Cremer, L. (**1942**). "Theorie der schalldämmung wände bei schrägem einfall," Akust. Z. **7**, 81–104. Most of this article has been republished with an English language summary in: Northwood, T. D. (**1977**). "Theory of the sound attenuation of thin walls with oblique incidence," in *Architectural Acoustics*, Benchmark Papers in Acoustics Vol. 10, edited by T. D. Northwood (Dowden, Hutchinson and Ross, Stroudsburg, PA), pp. 367–399.

Cremer, L., Heckl, M., and Ungar, E. E. (**1988**). *Structure-Borne Sound: Structural Vibrations and Sound Radiation at Audio Frequencies*, 2nd ed. (Springer-Verlag, Berlin).

Crocker, M. J., and Price, A. J. (**1969**). "Sound transmission using statistical energy analysis," J. Sound Vib. **9**, 469–486.

Davy, J. L. (**2004**). "The radiation efficiency of finite size flat panels," in *Acoustics 2004, Transportation Noise and Vibration—The New Millen-*

nium, Proceedings of the Annual Conference of the Australian Acoustical Society, Gold Coast, Australia, 3–5 November, edited by M. J. Mee, R. J. Hooker, and I. D. M. Hillock (Australian Acoustical Society, Castlemaine, VIC, Australia), pp. 555–560.

Dupree, R. B., **1981**, "Catalog of STC and IIC ratings for wall and floor/ceiling assemblies—With TL and ISPL data plots," Office of Noise Control, California Department of Health Services, Berkeley, CA.

Gradshteyn, I. S., and Ryzhik, I. M., **1965**, in *Table of Integrals, Series, and Products*, prepared by Yu. V. Geronimus and M. Yu. Tseytlin, translated from Russian by Scripta Technica Inc., 4th ed., edited by A. Jeffrey (Academic, New York).

Heckl, M., and Donner, U. [(**1985**). "Schalldämmung dicker Wände (Sound insulation of thick walls)," Rundfunktech Mitt. **29**, 287–291.

ISO, **2005**, "ISO 15712-1:2005(E) Building acoustics—Estimation of acoustic performance of buildings from the performance of elements—Part 1: Airborne sound insulation between rooms," International Organisation for Standardization, Geneva, Switzerland.

Leppington, F. G., Heron, K. H., Broadbent, E. G., and Mead, S. M. (**1987**). "Resonant and nonresonant acoustic properties of elastic panels. II. The transmission problem," Proc. R. Soc. London, Ser. A **412**, 309–337.

Ljunggren, S. (**1991**). "Airborne sound insulation of thick walls," J. Acoust. Soc. Am. **89**, 2338–2345.

Magrab, E. B. (**1979**). *Vibrations of Elastic Structural Members* (Sijthoff & Noordhoff, Alphen aan der Rijn, The Netherlands).

Northwood, T. D. (**1968**). "Transmission loss of plasterboard walls," Building Research Note, BRN-66, Division of Building Research, National Research Council of Canada, Ottawa, ON, Canada.

Quirt, J. D. (**1981**). "Measurements of the sound transmission loss of windows," Building Research Note, BRN-172, Division of Building Research, National Research Council of Canada, Ottawa, ON, Canada.

Quirt, J. D. (**1982**). "Sound transmission through windows I. Single and double glazing," J. Acoust. Soc. Am. **72**, 834–844.

Rindel, J. H. (**1975**). "Transmission of traffic noise through windows—Influence of incident angle on sound insulation in theory and experiment," Technical University of Denmark Report No. 9, The Acoustics Laboratory, Technical University of Denmark, Lyngby, Denmark.

Sato, H.. (**1973**). "On the mechanism of outdoor noise transmission through walls and windows—A modification of infinite wall theory with respect to radiation of transmitted wave," J. Acoust. Soc. Jpn. **29**, 509–516.

Sewell, E. C. (**1970**). "Transmission of reverberant sound through a single-leaf partition surrounded by an infinite rigid baffle," J. Sound Vib. **12**, 21–32.

Villot, M., Guigou, C., and Gagliardini, L. (**2001**). "Predicting the acoustical radiation of finite size multi-layered structures by applying spatial windowing on infinite structures," J. Sound Vib. **245**, 433–455.

Relation between derived-band auditory brainstem response latencies and behavioral frequency selectivity

Olaf Strelcyk, Dimitrios Christoforidis, 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 6 July 2009; accepted 9 July 2009)

Derived-band click-evoked auditory brainstem responses (ABRs) were obtained for normal-hearing (NH) and sensorineurally hearing-impaired (HI) listeners. The latencies extracted from these responses, as a function of derived-band center frequency and click level, served as objective estimates of cochlear response times. For the same listeners, auditory-filter bandwidths at 2 kHz were estimated using a behavioral notched-noise masking paradigm. Generally, shorter derived-band latencies were observed for the HI than for the NH listeners. Only at low click sensation levels, prolonged latencies were obtained for some of the HI listeners. The behavioral auditory-filter bandwidths accounted for the across-listener variability in the ABR latencies: Cochlear response time decreased with increasing filter bandwidth, consistent with linear-system theory. The results link cochlear response time and frequency selectivity in human listeners and offer a window to better understand how hearing impairment affects the spatiotemporal cochlear response pattern. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203310]

PACS number(s): 43.64.Ri, 43.64.Kc, 43.66.Sr, 43.66.Dc [BLM] Pages: 1878–1888

I. INTRODUCTION

The cochlea responds to sound with a displacement wave that propagates on the basilar membrane (BM) from base to apex (e.g., Ruggero, 1994; Robles and Ruggero, 2001). This traveling wave reaches maximum amplitude at a particular point before slowing down and decaying rapidly. The lower the frequency of a sound, the further its wave propagates down the cochlea. The corresponding propagation delay has been referred to as cochlear response time (CRT) (e.g., Don *et al.*, 1993).

Several studies have suggested that the extraction of spatiotemporal information, i.e., the combination of phase-locked auditory-nerve (AN) responses and systematic frequency-dependent delays along the cochlea (associated with the traveling wave), may be important in the context of pitch perception (e.g., Loeb *et al.*, 1983; Shamma and Klein, 2000), localization (e.g., Shamma *et al.*, 1989), and speech formant extraction (e.g., Deng and Geisler, 1987). It has been proposed that a distorted spatiotemporal BM response might be, at least partly, responsible for deficits in the processing of temporal-fine-structure information in hearing-impaired (HI) listeners (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. Hence, it is important to gain a better understanding of how hearing impairment affects the spatiotemporal behavior of the auditory periphery. In this study, CRT was investigated as it is an important component of the spatiotemporal response. Changes in CRT, due to cochlear hearing impairment, may result in distortions in the spatiotemporal response pattern.

The CRT can be considered as the sum of a cochlear transport time and a filter build-up time. It has been shown that concepts of linear-system theory apply to some extent to BM responses (e.g., Goldstein *et al.*, 1971; Geisler and Sinex, 1982; Recio and Rhode, 2000). In such a linear framework, the transport time corresponds to the signal-front delay in an auditory filter (Ruggero, 1980), while the filter build-up time corresponds to the duration from response onset to the time when the center of gravity (Goldstein *et al.*, 1971; Ruggero, 1994) or peak amplitude (Geisler and Sinex, 1983) of the BM response is reached (see also Ruggero and Temchin, 2007). Don *et al.* (1998) suggested that the filter build-up time mainly reflects the delay that is introduced by the cochlear amplifier sharpening the BM tuning (e.g., Robles and Ruggero, 2001).

CRTs have been studied extensively in normal-hearing (NH) listeners, using noninvasive methods such as measurements of compound action potentials (e.g., Eggermont, 1976), stimulus-evoked otoacoustic emissions (OAEs) (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; Don *et al.*, 1993).

Apart from studies on Ménière's disease (e.g., Eggermont, 1979; Rutten, 1986; Donaldson and Ruth, 1996), where changes in CRT are supposed to reflect changes in cochlear transport time due to endolymphatic hydrops (Thornton and Farrell, 1991), only few studies examined CRT in HI listeners. Donaldson and Ruth (1996) measured derived-band ABRs in HI listeners and found no alterations as a consequence of hearing loss (in the group without Ménière's disease). In contrast, Don *et al.* (1998), using a similar method, reported a tendency toward shorter response latencies with increasing hearing loss. Since hearing loss is often related to reduced frequency selectivity in HI listeners (e.g., Tyler *et al.*, 1983; Moore, 1996; Baker and Rosen, 2002), Don *et al.* (1998) suggested that the shorter latencies

 0001-4966/2009/126(4)/1878/11/$25.00

reflected increased auditory-filter bandwidths, consistent with the uncertainty principle of Fourier analysis (Papoulis, 1962).

Animal studies provided empirical evidence for such a relation between CRT and frequency selectivity (e.g., Goldstein *et al.*, 1971; Geisler and Sinex, 1983). However, in humans, this relation has not yet been demonstrated directly. Shera *et al.* (2002) estimated CRTs (in terms of BM group delays) from stimulus-frequency OAEs in humans. Based on relations between CRT and auditory-filter bandwidth from animal data, they predicted human filter bandwidths. The average bandwidths as a function of frequency were consistent with behavioral bandwidth estimates obtained in a forward-masking paradigm. However, the method by which Shera *et al.* (2002) obtained their CRT estimates is a matter of debate, as it relies on assumptions of how stimulus-frequency OAE delays relate to BM delays (Ruggero and Temchin, 2005). Furthermore, Ruggero and Temchin (2005) questioned the use of filter bandwidths obtained in forward masking as measures of cochlear frequency tuning in the above study.

In the present study, CRTs were estimated from derived-band ABRs of NH and HI listeners, as a function of frequency and level. Possible alterations in CRT, linked with cochlear hearing impairment, were examined. In order to study explicitly the relation between CRT and frequency selectivity across the individual listeners, behavioral estimates of filter bandwidth were obtained at a frequency of 2 kHz, using a notched-noise simultaneous-masking paradigm (e.g., Patterson and Nimmo-Smith, 1980). Specifically, it was examined whether individual filter bandwidths could account for the observed across-listener variability in CRTs. It was expected that the across-listener variability within the group of HI listeners would provide valuable information when investigating the relation between CRT and frequency selectivity.

II. AUDITORY BRAINSTEM RESPONSES

A. Method

1. Listeners

The five NH listeners were aged between 23 and 25 years and had audiometric thresholds better than 20 dB hearing level (HL) (ISO 389-8, 2004) at all octave frequencies from 125 to 8000 Hz and from 750 to 6000 Hz. Only female listeners participated in the present study to avoid gender-related differences in the ABR latencies (e.g., Don *et al.*, 1993). The 12 HI listeners were aged between 42 and 80 years (median: 55). Their better ears in terms of audiometric thresholds were chosen for further testing, so that effects of cross-ear listening could be ruled out. Their audiograms are shown in Fig. 1. All HI listeners showed sloping audiograms with a mild-to-moderate hearing loss at high frequencies. This is reflected in the mean audiogram, indicated by the dotted curve in Fig. 1. Audiograms were "normal" (thresholds $\lesssim$20 dB HL) up to 1 kHz in seven of the HI ears, while one audiogram reflected a mild hearing loss and four reflected a moderate hearing loss of up to 50 dB HL at these frequencies. The sensorineural origin of the hearing losses

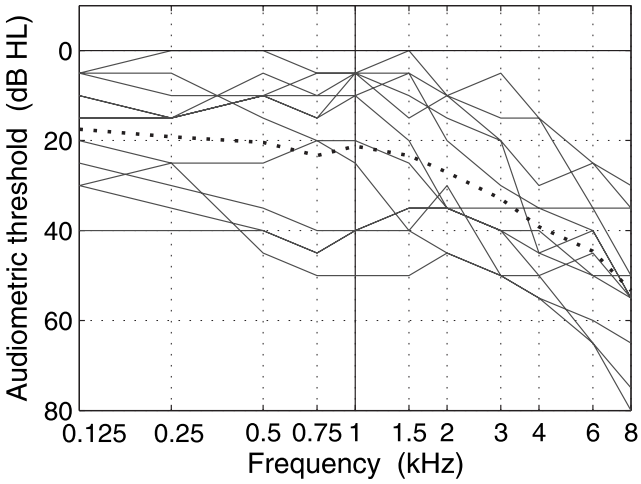


FIG. 1. Audiograms of the 12 HI listeners. The dotted curve shows the arithmetic mean.

was established by means of otoscopy, bone-conduction, and acoustic-reflex measurements. There were no indications of Ménière's disease such as episodic vertigo, fluctuating hearing loss, or sensation of fullness. ABRs to 100-dB peak-to-peak equivalent sound-pressure-level (ppe SPL) clicks were measured in both ears of all HI listeners. Since the interaural wave-*V* delay, the interaural wave-*V* amplitude difference, and the monaural wave *I*-*V* interpeak delays were within normal range for all ears, there was no indication of eighth-nerve tumors or brainstem lesions (Don and Kwong, 2002).

2. Stimuli

Rarefaction clicks were produced by applying 83-μs rectangular pulses (generated in MATLAB®) to an Etymotic Research ER-2 insert earphone. In all measurements, click trains with a click repetition rate of 45 Hz were used. The clicks were presented monaurally at five equally spaced levels ranging from 16 dB sensation level (SL) (for one HI listener a lower level of 11 dB was used) to the upper fixed level of 93 dB ppe SPL. The number of intermediate levels was reduced if a small individual dynamic range would have resulted in a level spacing smaller than 5 dB. 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). In this way, click and noise stimuli were mixed acoustically. The spectrum level of the high-pass noise maskers was identical to that of the broadband pink noise, for which pilot measurements with the NH listeners indicated that a level of 91 dB SPL was sufficient to mask the ABR to the

93-dB ppe SPL clicks. This is comparable to the noise level used by Don *et al.* (1998). For the lower click levels, the noise was attenuated with the click to maintain a fixed click-to-noise ratio.

The 93-dB ppe SPL clicks were presented in quiet (unmasked) and in the five high-pass noise maskers. The frequency bands between 1 and 2 kHz as well as 2 and 4 kHz have previously been found to yield the most salient derived-band ABRs (e.g., Eggermont and Don, 1980). Therefore, in order to save measurement time, only the conditions with 1-, 2-, and 4-kHz high-pass maskers were measured for the lower click levels.

Perceptual click thresholds were measured for 500-ms click trains using a three-interval, three-alternative, forced-choice (3I-3AFC) task with a one-up two-down rule, tracking the 71%-correct point on the psychometric function (Levitt, 1971). The final threshold was estimated as the arithmetic mean over three runs.

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 (zero-phase shift forward-backward filtering) was applied. Weighted averaging, as discussed in Elberling and Wahlgreen (1985) and in Don and Elberling (1994), was used for estimation of the auditory evoked potential. 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 24.9 (19.5, 31.9) nV, averaged across all listeners and conditions, with the values in parentheses representing the range of the individual results. The background noise level did not differ significantly between the NH and HI listeners ($p > 0.1$) and was not correlated with absolute hearing thresholds ($p > 0.2$).

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 in adjacent high-pass maskers, were obtained and the corresponding wave-*I* and wave-*V* latencies were extracted. The center frequencies of the derived bands were computed as the geometric mean 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 arbitrarily chosen as the upper

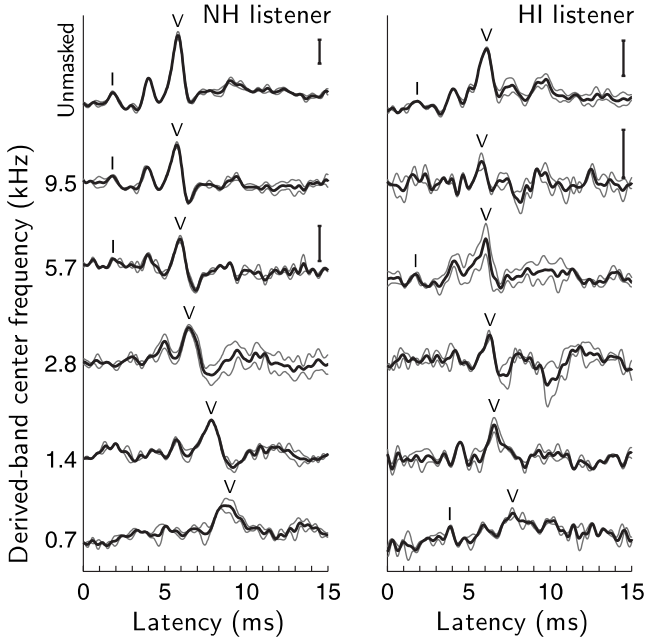


FIG. 2. Examples of unmasked and derived-band ABRs to 93-dB ppe SPL clicks from one NH listener (left) and one HI listener (right). Two replications (gray) and their average (black) are shown. Waves *I* and *V* are indicated by the corresponding symbols. The bars to the right represent 200 nV. If no bar is shown, the nearest bar above holds.

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 2 illustrates a series of derived-band ABRs to 93-dB ppe SPL clicks from one NH listener (left) and one HI listener (right). Waves *I* and *V* are indicated. As can be seen, wave *I*s could not be identified for all derived bands. For both listeners, wave-*V* latencies increased with decreasing derived-band center frequency. However, the increase was larger for the NH listener than for the HI listener: The latter showed shorter latencies in the three lowest derived bands than the NH listener.

For the further analysis of the wave-*V* latencies, in terms of nonlinear statistical modeling, the following latency model was adapted from Neely *et al.* (1988):

$$\tau(f,i) = a + bc^{0.93-i} f^{-d}, \tag{1}$$

where f represents the derived-band center frequency, normalized to 1 kHz, i represents click stimulus level, normalized to 100 dB ppe SPL, and a, b, c, and d are fitting constants. This parametrization slightly deviates from that in Neely *et al.* (1988). In contrast to their reference level of 0 dB, a reference level of 93 dB was chosen here, so that the parameter b reflects the CRT at a frequency of 1 kHz and a click level of 93 dB. The model parameter a represents an asymptotic delay, which is independent of frequency and level. To a first approximation, it reflects the post-cochlear contributions to the wave-*V* latencies, i.e., synaptic and neural delays. A possible shortcoming of this parametrization will be discussed in Sec. II B 3.

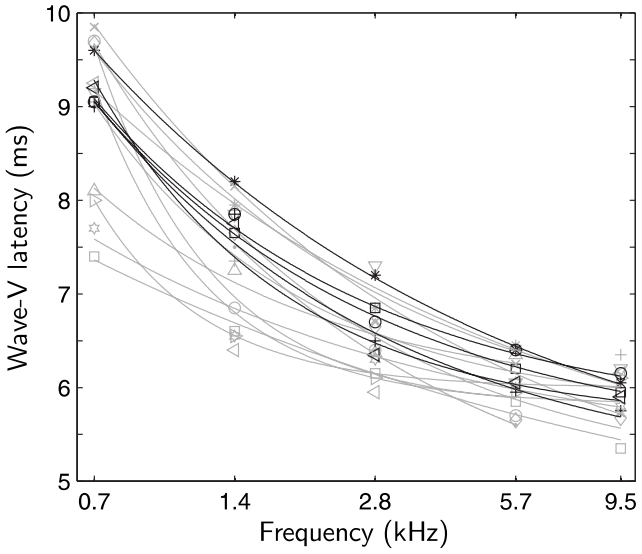


FIG. 3. Measured derived-band wave-*V* latencies in response to the 93-dB ppe SPL clicks, as a function of derived-band center frequency, for the NH (black symbols) and HI listeners (gray symbols). The curves show individual fits, according to the latency model in Eq. (1).

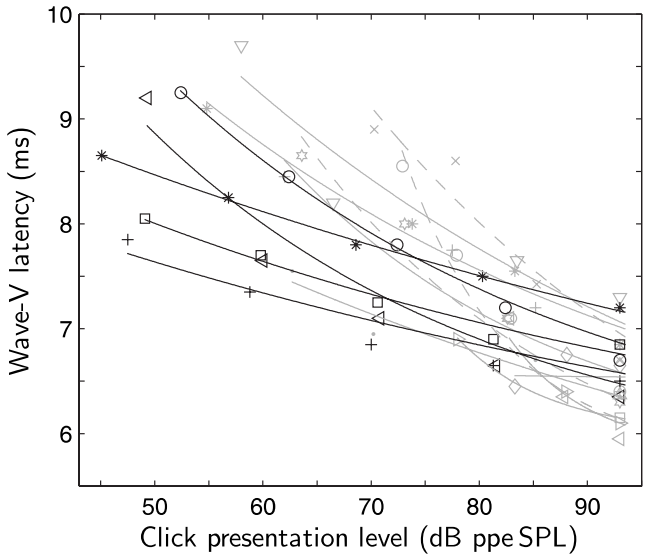


FIG. 4. Measured derived-band wave-*V* latencies as a function of absolute click level for the 2.8-kHz derived band, for the NH (black symbols) and HI listeners (gray symbols). The curves show individual model fits.

B. Results and discussion

1. Wave-V latency as a function of frequency

Figure 3 shows the measured (symbols) and fitted (curves) wave-*V* latencies at 93-dB ppe SPL as a function of frequency for the NH (black) and the HI (gray) listeners. 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 0.17 (SD 0.09) ms, averaged across all listeners (SD denotes the standard deviation).

Latencies decreased with increasing frequency. The latencies for the NH listeners were consistent with previous reports in literature (e.g., Eggermont and Don, 1980; Don *et al.*, 1993), both in absolute terms as well as in frequency dependence. For 8 of the 12 HI listeners, shorter latencies were observed than for the NH listeners (at 1.4 and 2.8 kHz). A similar trend was reported previously by Don *et al.* (1998). Also, the latency differences between wave *V*s of the 1.4- and 2.8-kHz derived bands were smaller for the HI than for the NH listeners ($p=0.05$). The across-listener variability was larger among the HI listeners than the NH listeners, and the variability among the HI listeners was larger at the lower than at the higher frequencies, in terms of the wave-*V* latencies. At the lowest frequency of 700 Hz, a wave-*V* latency could not be identified for one of the HI listeners, whereas at the highest frequency of 9500 Hz, latencies could not be identified for four of the HI listeners.

2. Wave-V latency as a function of click level

Figure 4 shows the measured (symbols) and fitted (curves) wave-*V* latencies as a function of click level in the 2.8-kHz derived band. The results obtained for the 1.4-kHz derived band followed the same trends and are therefore not shown. Latencies decreased with increasing click level, consistent with the results from Eggermont and Don (1980) and Don *et al.* (1993). At high click levels (above about 80 dB ppe SPL), the HI listeners (gray curves) showed a trend to shorter latencies than the NH listeners (black curves), consistent with the trend in the data from Fig. 3, observed at 93 dB. At lower click levels, however, some of the HI listeners showed longer latencies than the NH listeners. This finding will be discussed further below.

To accommodate the repeated-measures design, the statistical analyses were carried out using nonlinear mixed-effect models (NMEMs) (Lindstrom and Bates, 1990; Pinheiro and Bates, 2000), as implemented in S-PLUS®. The between-listener variability that was not explained in terms of the fixed effect listener group was accounted for in terms of listener-specific random effects. An analysis of variance (ANOVA) was performed based on a NMEM, which followed the latency model given in Eq. (1). The ANOVA confirmed the significance of derived-band frequency $[F(1,148)=58.1, p<0.0001]$ and click level $[F(1,148)=62.6, p<0.0001]$ on wave-*V* latency. Also the effect of listener group was significant ($p<0.001$), with a smaller parameter b (reference CRT at 1 kHz and 93 dB) and a steeper level slope c for the HI than for the NH listeners (on average, slopes were increased by a factor of 2, independent of derived-band frequency). Since the parameter b varied significantly across listeners ($p<0.0001$), it was modeled as listener-specific random effect. The frequency slope d did not differ significantly across listeners ($p>0.1$). The estimated mean parameters were $a=4.7$ ms, $b=3.4$ ms, $c=5.2$, and $d=0.50$.

For some of the HI listeners, longer latencies were observed than for the NH listeners (Fig. 4), for click presentation levels below about 80 dB ppe SPL. This cannot be explained in terms of cochlear filter bandwidth since the observed longer latencies would imply an implausible better-than-normal frequency tuning for the HI listeners. Outer hair cell (OHC) damage might not only affect the frequency tuning and sensitivity of the BM (e.g., Evans and Harrison, 1976; Liberman and Dodds, 1984) but also its local stiffness. Donaldson and Ruth (1996) and Don *et al.* (1998) suggested

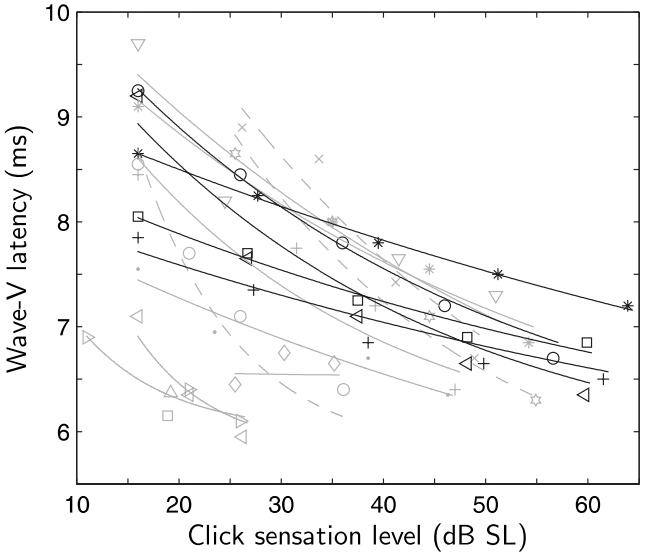


FIG. 5. Same as Fig. 4, but as a function of click sensation level (click level re perceptual click threshold).

that a loss of OHCs could result in decreased stiffness of the cochlear partition. Decreased stiffness would result in longer transport times and consequently in prolonged wave-*V* latencies. However, since the transport time itself reflects a "passive" BM property (Ruggero and Temchin, 2007), such changes in transport time would be expected to be independent of stimulus level. Hence, this explanation is unlikely to account for the fact that steeper slopes of the latency-level curves were observed for the HI than for the NH listeners.

The latency prolongations might be a result of reduced sensitivity reflected in the hearing losses of the HI listeners: They showed significantly higher click thresholds than the NH listeners [NH: 32.7 (SD 2.7) dB ppe SPL; HI: 54.3 (SD 13.5) dB ppe SPL; $F(1,15)=12.3, p=0.003$].[1] Indeed, when latencies are plotted as a function of the click sensation level, as shown in Fig. 5, the latencies for the HI listeners fall well into or below the range of the NH listeners. Reduced BM sensitivity would result in decreased input levels to the inner hair cells (IHCs). Recordings from AN-fibers in response to tones have shown that AN first-spike latencies depend on stimulus level (see Heil, 2004). This has been attributed to temporal integration processes in the synapses between IHCs and AN-fibers (Heil and Neubauer, 2003). Hence, increasing synaptic delays with decreasing input level might have been responsible for the abnormally long latencies observed in some of the HI listeners (when considered as a function of absolute SPLs).

Also, damage to or loss of IHCs or AN-fibers might partly account for the latency prolongations. Synaptic delays, for example, would increase more strongly with decreasing stimulus level in less sensitive neurons, i.e., neurons with high spike thresholds (cf. Heil, 2004, Fig. 1). Loss of IHCs or AN-fibers does not necessarily result in elevated pure-tone thresholds (e.g., Schuknecht and Woellner, 1953) and is even less likely to be reflected in thresholds for broadband clicks. Hence, the steep slopes of the latency-level curves for some of the HI listeners might be partly attributable to hidden IHC or AN-fiber losses.

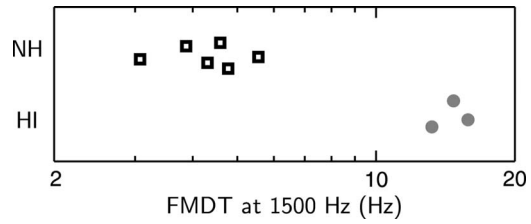


FIG. 6. FMDTs at a carrier frequency of 1500 Hz, for six NH listeners (data taken from Strelcyk and Dau, 2009) and three of the HI listeners.

In order to test the tenability of this hypothesis, detection thresholds for sinusoidal 2-Hz frequency modulation (FMDTs) were measured for three of the HI listeners who showed steep latency-level curves (dashed curves in Figs. 4 and 5). Low-rate FM detection is supposed to be determined by a temporal mechanism processing temporal-fine-structure information (e.g., Moore and Sek, 1996), which is coded in terms of phase locking in AN-fibers. Therefore, it has been suggested that IHC or AN-fiber integrity may be crucial for normal FM detection performance, and that IHC/AN-fiber losses might account for the degraded performance observed in HI listeners (cf. Buss *et al.*, 2004). Under these assumptions, FM detection can be regarded as a measure of IHC/AN-fiber integrity. The FMDTs were measured using a carrier frequency of 1500 Hz, for which NH reference data were available from a previous study (Strelcyk and Dau, 2009; a detailed method description can be found therein). As illustrated in Fig. 6, the three tested HI listeners (gray bullets) showed markedly increased FMDTs compared with the six NH listeners (black squares) of Strelcyk and Dau (2009). This outcome is consistent with the hypothesis that steeply sloping latency-level curves might, at least partly, be linked to loss of IHCs or AN-fibers.

3. Estimation of cochlear response time

In addition to CRTs, wave-*V* latencies reflect the delay introduced by the IHC-AN synapses and a central conduction time to the point in the brainstem, which is responsible for the wave-*V* peak activity. Central conduction time has been shown to be independent of frequency (Don and Eggermont, 1978; Ponton *et al.*, 1992; Don *et al.*, 1993) and click level (Eggermont and Don, 1980). Synaptic time delays are independent of frequency (cf. Ruggero, 1992), but they may depend on stimulus level, as discussed above.

In order to estimate central conduction times, wave *I*-*V* interpeak delays were extracted from the derived-band responses (cf. Fig. 2). When interpeak delays were available for several derived bands and click levels, the mean value was taken as an estimate of the central conduction time in the individual listener. The obtained average conduction time was 4.1 (SD 0.2) ms, with no significant difference between listener groups ($p>0.2$).[2] For the synaptic delay, a constant value of 0.8 ms was assumed (Eggermont, 1979). The resulting average value of 4.9 ms for the sum $\tau_{IV+0.8}$ of conduction time and synaptic delay was consistent with previous estimates of the post-cochlear delay in literature (e.g., Neely *et al.*, 1988; Donaldson and Ruth, 1993). The difference between the wave-*V* latency and the $\tau_{IV+0.8}$ delay was taken as

an estimate of the CRT (cf. Don *et al.*, 1998). Therefore, the individual $\tau_{IV+0.8}$ delay was substituted for the asymptotic delay a in model (1).[3] This yielded the new latency model

$$\tau(f,i) = \tau_{IV+0.8} + bc^{0.93-i}f^{-d}. \quad (2)$$

A new NMEM, based on this model, confirmed the results of the previous NMEM, in terms of parameter estimates as well as of significance of derived-band frequency, click level, and listener group. Also in terms of goodness of fit the two models were roughly equivalent, yielding equally good descriptions of the latency data.

Synaptic delays might not be constant, as assumed above, but increase with decreasing click level. In this case, the CRT estimates [second summand in Eq. (2)] would partly reflect the level-dependent synaptic delays and CRTs would be overestimated at low click levels. In the following, the relation between CRTs and behavioral estimates of frequency selectivity will be explored, in order to test if the across-listener variability in CRTs could be attributed to differences in frequency selectivity. While decreased frequency selectivity is expected to result in decreased CRTs, decreased sensitivity (and thus potentially decreased input levels at the synapse) should result in increased CRT estimates. Hence, the assumption of a constant synaptic delay is conservative with regard to the hypothesized relation between CRT and frequency selectivity.

III. FREQUENCY SELECTIVITY

A. Method

1. Listeners

The notched-noise masking measurements were performed by the same listeners who participated in the ABR measurements.

2. Stimuli and procedures

Auditory-filter shapes at 2 kHz were determined for the ABR test-ears using a notched-noise masking paradigm (cf. Patterson and Nimmo-Smith, 1980). Rosen *et al.* (1998) presented evidence that auditory-filter shapes are output driven. Under the assumption of the power-spectrum model (cf. Patterson and Moore, 1986) that a constant signal-to-noise ratio at the output of the auditory filter is required for detection, a constant masker *spectrum* level results in different filter-output levels for the different notch conditions at threshold, while a constant signal level results in constant output levels (to the extent that the signal level remains unchanged under filtering). Therefore, here, in order to obtain a faithful filter estimate, the signal level was kept constant while the masker level was varied adaptively.

The 2-kHz target tones of 440-ms duration were temporally centered in the 550-ms noise maskers. Maskers and tones were gated with 50-ms raised-cosine ramps. The noise was generated in the spectral domain as fixed-amplitude random-phase noise. Five symmetric ($\delta f/f_0$: 0.0, 0.1, 0.2, 0.3, and 0.4) and two asymmetric notch conditions ($\delta f/f_0$: 0.2|0.4 and 0.4|0.2) were used, where δf denotes the spacing between the inner noise edges and the signal frequency f_0. The outside edges of the noise maskers were fixed at $\pm 0.8 f_0$. The tones were presented at a fixed level of 40 dB SPL for the NH listeners. For some of the HI listeners, a level of 40 dB would have resulted in a sensation level of less than 15 dB. In order to obtain reliable filter estimates, in these cases, the tone level was increased to ensure a minimum sensation level of 15 dB.[4] The average tone level for the HI listeners was 47 (SD 8) dB.

A 3I-3AFC weighted up-down method (Kaernbach, 1991) was applied to track the 75%-correct point on the psychometric function. A run was terminated after 14 reversals. The threshold was defined as the arithmetic mean of all masker levels following the fourth reversal. Following a training run for each notch condition, the threshold was estimated as the average over three runs. If the SD of these three runs exceeded 1 dB, one or two additional runs were taken and the average of all was used.

3. 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 48 kHz. The stimuli were presented in a double-walled sound-attenuating booth via Sennheiser HD580 headphones. Calibrations were done using an ear simulator (IEC 60318-1 and -2, 1998; B&K 4153 with flat plate) and, prior to playing, 128-tap linear-phase FIR equalization filters were applied, rendering the headphone frequency response flat.

4. Filter fitting

A nonlinear minimization routine was implemented in MATLAB® to find the best-fitting rounded-exponential (roex) filter in the least-squares sense, assuming that the signal was detected using the filter with the maximum signal-to-noise ratio at its output. The roex(p,r) filter model (Patterson *et al.*, 1982) and a more complex variant, the roex(p,w,t,p) model as used by Oxenham and Shera (2003), were considered. At the low-frequency side, the filter shape $W(f)$ of the roex(p,w,t,p) filter is defined by

$$W(f) = (1-w)(1+pg)\exp(-pg) + w(1+pg/t)\exp(-pg/t), \quad (3)$$

where g represents the deviation from the center frequency as a proportion of the center frequency, p determines the passband-slope of the filter, t determines the factor by which the tail-slope is shallower than the passband-slope, and w determines the relative weights of the two slopes. The high-frequency side of the filter is described by a single slope

$$W(f) = (1+pg)\exp(-pg), \quad (4)$$

which is independent of the low-frequency side.

The equivalent rectangular bandwidth (ERB) and the 10-dB bandwidth were computed as measures of filter tuning. However, because they yielded very similar results, for ease of comparison, only the ERB results will be discussed further. Also, the uncertainty of the ERB was estimated via bootstrapping: Based on the empirical standard errors of the individual notched-noise thresholds, for each listener, a large number of threshold curves were resampled. Subsequently,

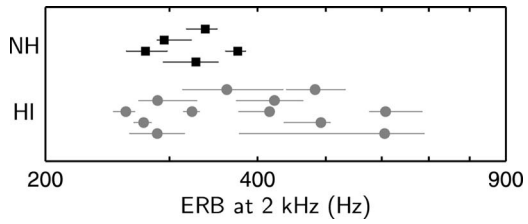


FIG. 7. ERB of the roex(p,w,t,p) filter estimates at 2 kHz for the NH (black squares) and HI (gray bullets) listeners. The errorbars represent the 15th and 85th percentiles, which were estimated via bootstrapping; see text for details.

auditory filters were fitted to these threshold replications. The resulting bandwidth distribution yielded a confidence interval for the ERB.

B. Results and discussion

The roex(p,w,t,p) filter model provided a good description of the individual notched-noise threshold data, with a residual rms fitting error of 0.5 (SD 0.3) dB, averaged across all listeners. The rms fitting error for the simpler roex(p,r) filter model was on average larger by a factor of 1.3 (1.0, 1.6); the numbers in brackets represent the 15th and 85th percentiles, respectively. Therefore, only the results for the roex(p,w,t,p) model will be discussed in the following. However, the pattern of results and conclusions would remain unchanged if the roex(p,r)-results were considered instead.

Figure 7 shows the estimated ERBs for the NH (black squares) and HI (gray bullets) listeners. Although the HI listeners showed, on average, larger bandwidths than the NH listeners by a factor of 1.2 (0.9, 1.8) (15th and 85th percentiles), the difference in bandwidths between the two groups was not significant ($p=0.19$). This was due to the large spread of results within the group of HI listeners and the fact that six of the HI listeners showed bandwidths within the range of the NH listeners. For one of the HI listeners, the uncertainty of the estimated ERB was considerably larger than for the other listeners. This was due to the small range of masked thresholds (8 dB) across the different notch conditions for this listener, which rendered the filter estimate less precise (cf. Tyler *et al.*, 1984).

Within the group of HI listeners, the ERB at 2 kHz was significantly correlated with the individual hearing threshold at this frequency (Pearson correlation and two-tailed p value: $r=0.65$, $p=0.02$). Here, the hearing threshold was estimated by means of a 3I-3AFC method with a 1-dB stepsize. A similar correlation was observed when the pure-tone average (PTA) threshold at 0.5, 1, 2, and 4 kHz was considered instead ($r=0.59$). The finding of a correlation between frequency selectivity and hearing threshold is consistent with previous reports in literature (e.g., Tyler *et al.*, 1983; Moore, 1996). Typically, the correlations are less distinct for hearing losses below 30–40 dB HL (as in the present study) than for more severe losses (see Baker and Rosen, 2002).

The mean ERB for the NH listeners was 322 (SD 38) Hz. This value is larger than the value of 241 Hz predicted by the ERB function given in Glasberg and Moore (1990). Baker and Rosen (2002) found good agreement between their NH mean ERB (for 40-dB SPL 2-kHz tones) and the prediction by Glasberg and Moore (1990). The discrepancy, observed here, may be due to variability within the NH population and the particular subset of NH listeners chosen in the present study. In view of the comparison between the HI and the NH listeners, broader filter bandwidths for the NH listeners would have resulted in more conservative estimates of significance. This may explain why no significant bandwidth difference was found between the NH and HI listeners.

IV. RELATION BETWEEN CRT AND FREQUENCY SELECTIVITY

Figure 8 (left column) shows the objective ABR-based estimates of CRTs, for the five derived bands (at 93 dB), as a function of the behaviorally derived ERBs at 2 kHz. The black squares indicate the results for the NH listeners while the results for the HI listeners are shown with the gray bullets. CRTs were estimated by subtracting the delay $\tau_{IV+0.8}$ from the measured wave-V latencies, as discussed in Sec. II B 3. The dotted regression lines were obtained by means of least trimmed squares robust regression. It can be seen that CRT decreased with increasing filter bandwidth, consistent with the uncertainty principle. This was significant at all frequencies ($p<0.05$), but the correlations were strongest at 1.4, 2.8, and 5.7 kHz ($p<0.01$). The observed correlations remained largely unchanged when the results of the NH listeners were excluded and only the results of the HI listeners were considered. Also, in Fig. 8 (left column), it can be seen that the inclusion of the HI listeners was crucial in order to study the relation between CRT and auditory-filter bandwidth. While the HI listeners provided a relatively large span of bandwidths, the variability among the NH listeners alone would have been too small. At the frequencies of 1.4, 2.8, and 5.7 kHz, not only CRTs but also the wave-V latencies (not shown) were significantly correlated with the ERB ($r\sim-0.57, p<0.05$). However, the correlations were stronger for the corresponding CRT estimates. The latency difference between wave Vs of the 1.4- and 2.8-kHz derived bands was not significantly correlated with the ERB at 2 kHz ($p=0.09$).

Since the ERB was correlated with the individual hearing threshold, the correlations between CRT and ERB could have reflected an effect of hearing threshold on CRT rather than an effect of filter bandwidth *per se*. Therefore, the correlations between CRTs and hearing thresholds, as shown in Fig. 8 (right column), were examined. As suggested by Don *et al.* (1998), the locally weighted pure-tone average (PTA_w) was taken as a predictor for the effect of hearing loss on the derived-band ABR. In computing the PTA_w, the audiometric threshold for the pure tone closest to the derived-band center frequency was given twice the weight of the thresholds for the two adjacent pure tones (at the highest derived-band frequency of 9.5 kHz, the audiometric threshold at 8 kHz was taken instead). As can be seen in Fig. 8, for all frequencies, the correlations between CRTs and the PTA_w (right column) were weaker than the corresponding correlations between

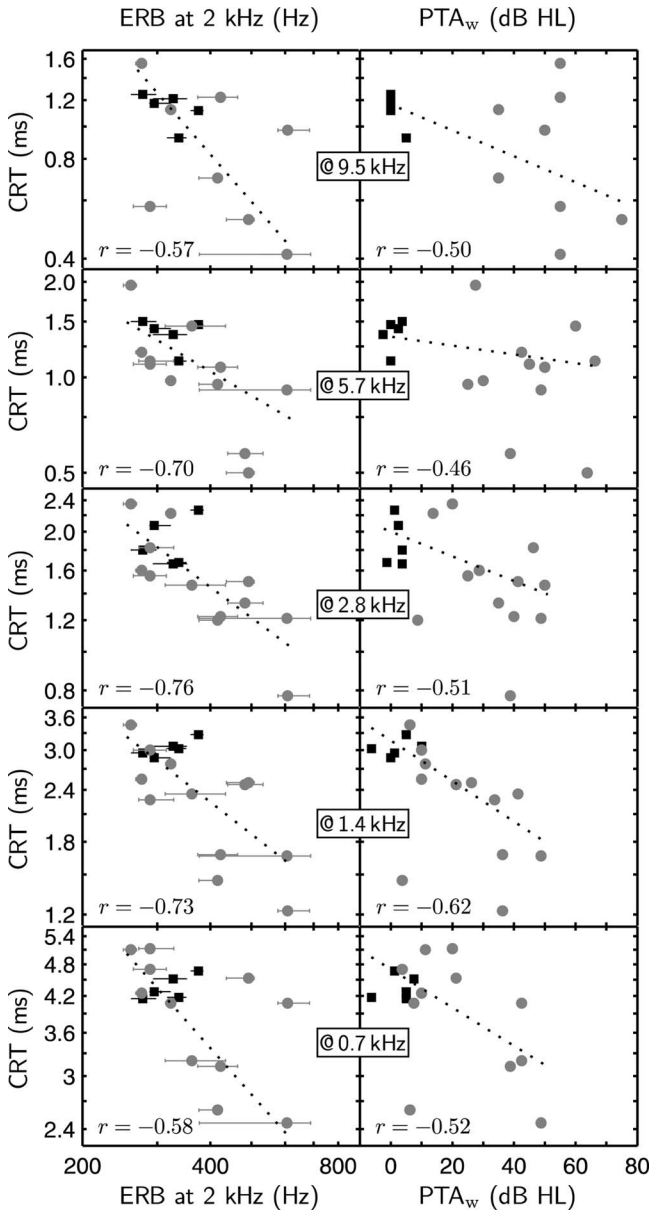


FIG. 8. CRTs for the derived-band center frequencies of 0.7, 1.4, 2.8, 5.7, and 9.5 kHz, as a function of the ERB at 2 kHz (left column) and the PTA$_w$ (right column). CRTs were estimated by subtracting the individual $\tau_{IV+0.8}$ delays from the 93-dB derived-band wave-*V* latencies. Black squares indicate the results for the NH listeners while the results for the HI listeners are represented by gray bullets. The regression lines were obtained by means of robust regression. The errorbars to the ERBs represent the 15th and 85th percentiles, estimated via bootstrapping; see text for details.

CRTs and ERBs (left column). This suggests that the across-listener trend of decreasing CRT with increasing ERB indeed reflected an effect of filter bandwidth *per se*.

In order to test if the bandwidth estimates at 2 kHz could account for the across-listener variability in all latency data, including the latencies obtained at the lower click levels, the NMEM (from Sec. II B) was extended. In addition to the significant effects of derived-band frequency and click level, the filter bandwidth in terms of the ERB was included, following a power-law dependence with exponent e as new model parameter

$$\tau(f,i) = \tau_{IV+0.8} + bc^{0.93-i} f^{-d} \mathrm{ERB}^{-e}. \quad (5)$$

The ERB was found to be highly significant [$F(1,149) = 40.6, p < 0.0001$], with an estimated value of 0.6 for the exponent e. This confirmed the trend that was observed for the 93-dB latencies (Fig. 8, left column): Listeners with broader auditory filters at 2 kHz showed shorter derived-band latencies. Due to the inclusion of the ERB, the effect of listener group on the parameter b (CRT at 1 kHz and 93 dB) was no longer significant. However, the level slope c was still significantly steeper for the HI than for the NH listeners ($p < 0.001$).

Also here, it was tested whether the predictive power of the ERB (for estimating CRT) was due to its correlation with the hearing threshold rather than an effect of filter bandwidth *per se*. The significance of the absolute hearing thresholds (in terms of the PTA and the PTA$_w$) as well as the individual click thresholds was examined in a type-III ANOVA. This was done by allowing the CRT-reference parameter b in the NMEM to be an exponential function of one of the above factors, analogous to the latency-level dependence in Eq. (5). For example, in case of the PTA, this yielded $b = \exp(b_1 + b_2 \mathrm{PTA})$. However, none of the threshold measures reached significance ($p > 0.15$), while the effect of bandwidth remained significant. Again, this indicates that the variability in CRTs across listeners reflected changes in bandwidth and not sensitivity. Furthermore, in contrast to changes in filter bandwidth, it is unclear how reduced sensitivity, reflected in elevated absolute thresholds, could account for the observed shorter CRTs in the HI listeners. Changes in (unmasked) wave *I-V* intervals linked to hearing loss have been reported previously (e.g., Elberling and Parbo, 1987; Boettcher, 2002). However, these can be ruled out here as an explanation, since individual *I-V* delays were used for CRT estimation. Other effects, such as changes in synaptic delay with decreasing sensation level (see Sec. II B 2), would be expected to result in prolonged but not shortened CRTs as observed here. Hence, the most appropriate conclusion seems to be that the changes in CRT reflected changes in filter bandwidth. Also, no significant effect of age was found ($p > 0.1$).

The above results did not depend on the choice of $\tau_{IV+0.8}$ as estimate of the asymptotic delay in the NMEM latency model, Eq. (5). The same pattern of results was obtained using a model with free asymptotic-delay parameter a. Furthermore, in the above NMEM, all measured latencies (at all derived-band frequencies and click levels) were included. However, the results did not depend on a particular subset of the data: They were confirmed separately for the 93-dB data (fixed level) as well as the 1.4- and 2.8-kHz derived-band data (fixed frequency).

V. OVERALL DISCUSSION

Shorter derived-band latencies were generally observed for the HI than for the NH listeners. Only for low click presentation levels, prolonged latencies were obtained for some of the HI listeners. This seemed to be attributable to the low click sensation levels for these listeners. Behaviorally derived auditory-filter bandwidths at 2 kHz accounted for

part of the across-listener variability in the ABR latencies, consistent with the expectation that CRT decreases with increasing filter bandwidth.

It is a matter of debate whether behavioral bandwidth estimates obtained in simultaneous masking, as used in the present study, are "representative" of cochlear frequency tuning, or whether a nonsimultaneous (e.g., forward-masking) paradigm should be preferred (e.g., Moore and O'Loughlin, 1986; Shera *et al.*, 2002; Oxenham and Shera, 2003; Ruggero and Temchin, 2005). However, the simultaneous-masking paradigm might bear a closer relationship to CRTs estimated from derived-band ABRs, which are obtained in simultaneous high-pass masking. The observed relation between CRT and ERB is consistent with this hypothesis.

Auditory-filter shapes and the corresponding ERBs were measured at a single frequency and tone level. Therefore, the bandwidth results could not account for the frequency and level dependence of the ABR latencies. Nevertheless, strong correlations were observed between ERBs at 2 kHz and CRTs at the adjacent derived-band frequencies of 1.4 and 2.8 kHz (based on the 93-dB ABRs, Fig. 8). It remains to be seen, whether such correlations between ERB and CRT also hold at other frequencies. If this was the case, it might be possible to predict individual frequency selectivity from derived-band ABR measurements. Similar attempts have been made based on measurements of OAEs (e.g., Moleti and Sisto, 2003). Prediction of ERBs from ABR latencies could serve as an alternative to time-consuming objective tuning-curve measures of frequency selectivity in human listeners, based on masking functions of compound action potentials or ABRs (e.g., Klein and Mills, 1981; Harrison, 1984; Markessis *et al.*, 2009).

ACKNOWLEDGMENTS

We thank Dr. James Harte for technical support and valuable scientific discussions about various aspects of this project. We also thank Brenda L. Lonsbury-Martin, Bob Burkard, and one anonymous reviewer for their helpful comments on an earlier version of this manuscript. O.S. would like to thank Sabine Caminade and Dr. Gürkan Sin for valuable discussions regarding error estimation in nonlinear filter fitting. We are grateful to the listeners for their participation in testing. Part of this work was supported by the Danish Research Foundation, the Danish Graduate School SNAK "Sense organs, neural networks, behavior, and communication," and the Oticon Foundation.

[1]The click thresholds for the NH listeners are lower than the corresponding reference threshold of 43.2 dB ppe SPL given by Richter and Fedtke (2005). This 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.

[2]Wave *I-V* interpeak delays were also extracted from the unmasked click responses. As for the derived-band wave *I-V* delays, results did not differ significantly between listener groups ($p>0.2$). For the HI listeners, the unmasked *I-V* delays were slightly shorter than the derived-band *I-V* delays, on average by 0.1 ms. However, this difference was not significant ($p>0.2$). Neither the unmasked nor derived-band *I-V* delays were significantly correlated with absolute hearing thresholds or age ($p>0.1$).

[3]Alternatively, instead of using the $\tau_{IV+0.8}$ delay, the asymptotic delay a in Eq. (1) could have been estimated directly from the individual wave-*V* latencies. However, this was problematic since wave-*V* latencies at 9.5 kHz (and 93 dB) were missing for four of the HI listeners. Even if these latencies had been available, it is questionable that the upper frequency limit of 9.5 kHz was sufficiently high to estimate the asymptotic delay. Subtraction of the wave *I-V* delay from the observed 9.5-kHz wave-*V* latency at 93 dB yielded on average a remaining delay of 1.9 (SD 0.1) ms for the NH and 1.7 (SD 0.4) ms for the HI listeners. It seems that this remainder cannot be solely accounted for by a synaptic delay, for which a value of about 0.8 ms is commonly assumed at comparably high stimulus levels (e.g., Eggermont, 1979; Robles and Ruggero, 2001). In parts, the remainder may reflect a finite CRT at the frequency of 9.5 kHz and the level of 93 dB.

[4]The tone level of 40 dB SPL was chosen for the NH listeners since higher levels would have resulted in uncomfortably loud noise levels at masked threshold for some of the listeners. Alternatively, a constant low sensation level could have been chosen for all listeners. However, absolute sensitivity and frequency selectivity may be considered as representing parallel rather than sequential processes in the auditory periphery. Therefore, it seems more appropriate to compare filter bandwidths at similar SPLs rather than SLs. For another reason, choosing the same *low* sensation levels for the NH listeners as for the HI listeners would be problematic: Since filters broaden with increasing stimulus level (e.g., Baker and Rosen, 2002), much lower SPLs for the NH group than for some of the HI listeners (as would be the case if, for example, a common low sensation level of 15 dB had been chosen) would be least conservative in view of the group difference between NH and HI listeners.

Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330–1339.
Boettcher, F. A. (**2002**). "Presbyacusis and the auditory brainstem response," J. Speech Lang. Hear. Res. **45**, 1249–1261.
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.
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., and Kwong, B. (**2002**). "Auditory brainstem response: Differential diagnosis," in *Handbook of Clinical Audiology*, edited by J. Katz (Lippincott Williams & Wilkins, Philadelphia, PA), 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.
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.
Donaldson, G. S., and Ruth, R. A. (**1996**). "Derived-band auditory brainstem response estimates of traveling wave velocity in humans: II. Subjects with noise-induced hearing loss and Meniere's disease," J. Speech Hear. Res. **39**, 534–545.
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. (**1979**). "Narrow-band AP latencies in normal and recruiting human ears," J. Acoust. Soc. Am. **65**, 463–470.
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 Parbo, J. (**1987**). "Reference data for ABRs in retrocochlear diagnosis," Scand. Audiol. **16**, 49–55.
Elberling, C., and Wahlgreen, O. (**1985**). "Estimation of auditory brainstem

response, ABR, by means of Bayesian inference," Scand. Audiol. **14**, 89–96.
Evans, E. F., and Harrison, R. V. (**1976**). "Correlation between cochlear outer hair cell damage and deterioration of cochlear nerve tuning properties in the guinea-pig," J. Physiol. (London) **256**, 43P–44P.
Geisler, C. D., and Sinex, D. G. (**1982**). "Responses of primary auditory fibers to brief tone bursts," J. Acoust. Soc. Am. **72**, 781–794.
Geisler, C. D., and Sinex, D. G. (**1983**). "Comparison of click responses of primary auditory fibers with minimum-phase predictions," J. Acoust. Soc. Am. **73**, 1671–1675.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Goldstein, J., Baer, T., and Kiang, N. (**1971**). "A theoretical treatment of latency, group delay and tuning characteristics for auditory-nerve responses to clicks and tones," in *The Physiology of the Auditory System*, edited by M. Sachs (National Educational Consultants, Baltimore, MD), pp. 133–141.
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.
Harrison, R. V. (**1984**). "Objective measures of cochlear frequency selectivity in animals and in man. A review," Acta Neurol. Belg. **84**, 213–232.
Heil, P. (**2004**). "First-spike latency of auditory neurons revisited," Curr. Opin. Neurobiol. **14**, 461–467.
Heil, P., and Neubauer, H. (**2003**). "A unifying basis of auditory thresholds based on temporal summation," Proc. Natl. Acad. Sci. U.S.A. **100**, 6151–6156.
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.
Kaernbach, C. (**1991**). "Simple adaptive testing with the weighted up-down method," Percept. Psychophys. **49**, 227–229.
Klein, A. J., and Mills, J. H. (**1981**). "Physiological (waves I and V) and psychophysical tuning curves in human subjects," J. Acoust. Soc. Am. **69**, 760–768.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Liberman, M. C., and Dodds, L. W. (**1984**). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," Hear. Res. **16**, 55–74.
Lindstrom, M. L., and Bates, D. M. (**1990**). "Nonlinear mixed effects models for repeated measures data," Biometrics **46**, 673–687.
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.
Markessis, E., Poncelet, L., Colin, C., Coppens, A., Hoonhorst, I., Kadhim, H., and Deltenre, P. (**2009**). "Frequency tuning curves derived from auditory steady state evoked potentials: A proof-of-concept study," Ear Hear. **30**, 43–53.
Moleti, A., and Sisto, R. (**2003**). "Objective estimates of cochlear tuning by otoacoustic emission analysis," J. Acoust. Soc. Am. **113**, 423–429.
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., and O'Loughlin, B. J. (**1986**). "The use of nonsimultaneous masking to measure frequency selectivity and suppression," *Frequency Selectivity in Hearing* (Academic, London).
Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.
Moore, B. C. J., and Skrodzka, E. (**2002**). "Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation," J. Acoust. Soc. Am. **111**, 327–335.
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.
Oxenham, A. J., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.
Papoulis, A. (**1962**). *The Fourier Integral and Its Applications* (McGraw-Hill, New York).
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.
Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London).
Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter asymmetry," J. Acoust. Soc. Am. **67**, 229–245.
Patterson, R. D., 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.
Pinheiro, J., and Bates, D. (**2000**). *Mixed-Effects Models in S and S-PLUS* (Springer-Verlag, New York).
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.
Recio, A., and Rhode, W. S. (**2000**). "Basilar membrane responses to broadband stimuli," J. Acoust. Soc. Am. **108**, 2281–2298.
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.
Rosen, S., Baker, R. J., and Darling, A. (**1998**). "Auditory filter nonlinearity at 2 kHz in normal hearing listeners," J. Acoust. Soc. Am. **103**, 2539–2550.
Ruggero, M. A. (**1980**). "Systematic errors in indirect estimates of basilar membrane travel times," J. Acoust. Soc. Am. **67**, 707–710.
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 Temchin, A. N. (**2005**). "Unexceptional sharpness of frequency tuning in the human cochlea," Proc. Natl. Acad. Sci. U.S.A. **102**, 18614–18619.
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.
Rutten, W. L. (**1986**). "The influence of cochlear hearing loss and probe tone level on compound action potential tuning curves in humans," Hear. Res. **21**, 195–204.
Schuknecht, H. F., and Woellner, R. C. (**1953**). "Hearing losses following partial section of the cochlear nerve," Laryngoscope **63**, 441–465.
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.
Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.
Strelcyk, O., and Dau, T. (**2009**). "Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing," J. Acoust. Soc. Am. **125**, 3328–3345.
Thornton, A. R., and Farrell, G. (**1991**). "Apparent travelling wave velocity

changes in cases of endolymphatic hydrops," Scand. Audiol. **20**, 13–18.
Tognola, G., Grandori, F., and Ravazzani, P. (**1997**). "Time-frequency distributions of click-evoked otoacoustic emissions," Hear. Res. **106**, 112–122.
Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C., and Patterson, R. D. (**1984**). "Auditory filter asymmetry in the hearing impaired," J. Acoust. Soc. Am. **76**, 1363–1368.
Tyler, R. S., Wood, E. J., and Fernandes, M. (**1983**). "Frequency resolution and discrimination of constant and dynamic tones in normal and hearing-impaired listeners," J. Acoust. Soc. Am. **74**, 1190–1199.

Diotic and dichotic detection with reproducible chimeric stimuli

Sean A. Davidson
Department of Biomedical and Chemical Engineering, Institute for Sensory Research, Syracuse University, 621 Skytop Road, Syracuse, New York 13244

Robert H. Gilkey
Department of Psychology, Wright State University, Dayton, Ohio 45435 and Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433

H. Steven Colburn
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Department of Biomedical and Chemical Engineering, Department of Electrical Engineering and Computer Science, and The Institute for Sensory Research, Syracuse University, 621 Skytop Road, Syracuse, New York 13244

(Received 3 January 2008; revised 20 July 2009; accepted 21 July 2009)

Subject responses were measured for individual narrow-band reproducible stimuli in a low-frequency tone-in-noise detection task. Both N_0S_0 and N_0S_π conditions were examined. The goal of the experiment was to determine the relative importance of envelope and fine-structure cues. Therefore, chimeric stimuli were generated by recombining envelopes and fine structures from different reproducible stimuli. Detection judgments for noise-alone or tone-plus-noise stimuli that had common envelopes but different fine structures or common fine structures but different envelopes were compared. The results showed similar patterns of responses to stimuli that shared envelopes, indicating the importance of envelope cues; however, fine-structure cues were also shown to be important. The relative weight assigned to envelope and fine-structure cues varied across subjects and across interaural conditions. The results also indicated that envelope and fine-structure information are not processed independently. Implications for monaural and binaural models of masking are discussed.
 [DOI: 10.1121/1.3203996]

PACS number(s): 43.66.Dc, 43.66.Rq, 43.66.Ba [RLF] Pages: 1889–1905

I. INTRODUCTION

Fletcher (1940) suggested that tone-in-noise masking was directly related to the total stimulus energy in a narrow frequency region (the critical band) surrounding the tonal signal. Most subsequent research on diotic or monaural tone-in-noise masking also supports the idea that subjects base their decisions, at least in part, on the differences in energy between the stimulus on signal-plus-noise and noise-alone trials. Nevertheless, a wide variety of findings indicate that other stimulus cues also influence listeners' detection judgments (e.g., Ahumada and Lovell, 1971; Hall *et al.*, 1984; Neff and Callaghan, 1988; Richards, 1992; Colburn *et al.*, 1997; Davidson *et al.*, 2006). Specifically, several researchers have argued for the importance of fluctuations in the temporal envelope or the temporal fine structure of the wave form (e.g., Richards, 1992; Isabelle, 1995; Bernstein and Trahiotis, 1996; Carney *et al.*, 2002). A variety of psychophysical models for detection have been developed that rely on envelope (e.g., Dau *et al.*, 1996a, 1996b; Eddins and Barber, 1998) or on fine structure (e.g., Moore, 1975). Moreover, a number of researchers using a variety of approaches have provided evidence that envelope and fine structure are, or could be, separately processed in the auditory system (e.g., van de Par and Kohlrausch, 1997; van de Par and Kohlrausch, 1998; Kohlrausch *et al.*, 1997; Eddins and Barber, 1998; Breebaart *et al.*, 1999; Bernstein and Trahiotis, 2002; Smith *et al.*, 2002; Joris, 2003; and Zeng *et al.*, 2004). It has long been known that auditory-nerve responses phase lock to both individual cycles and to the envelopes of low-frequency stimuli (Kiang *et al.*, 1965) and only to the envelopes of high-frequency stimuli (Joris and Yin, 1992; Kay, 1982).

The questions of whether envelope-based or fine-structure-based decision variables can really be processed separately at low frequencies, and if so which dominates the detection process, remain. Unfortunately, because energy, envelope, and fine structure tend to co-vary in randomly generated stimuli, it is difficult to evaluate separately their impact on detection judgments. That is, adding a tone to a

[a)]Author to whom correspondence should be addressed. Present address: Department of Biomedical Engineering, Department of Neurobiology and Anatomy, University of Rochester, 601 Elmwood Ave., P.O. Box 603, Rochester, NY 14642. Electronic mail: laurel.carney@rochester.edu

narrow-band noise wave form tends to increase its energy, smooth its envelope, and reduce variation in the frequency of its fine structure.

One approach that has been successfully used to evaluate the role of envelope and fine-structure cues in other contexts involves the use of chimeras. Chimeras are stimuli formed by combining the envelope from one stimulus with the fine structure from another. Smith *et al.* (2002) tested speech recognition and sound localization using various chimeras and suggested that speech identification appeared to be based on envelope, whereas sound localization appeared to be based on fine structure. Zeng *et al.* (2004) refuted the latter result using chimeras with directionally conflicting interaural-time differences (ITDs, embedded in the fine structure) and interaural-level differences (ILDs, embedded in the envelope). The approach used in the study presented here differs from these efforts in that envelope and fine-structure cues were not systematically put in opposition. Instead, the envelopes and fine structures were chosen independently (within the bandwidth constraints discussed below), so that in any given wave form they could be in agreement or in disagreement in terms of their influence on a subject's probability of responding "tone present," and the subject could use either cue or both cues.

The goal of the study presented here was to evaluate the relative importance of envelope and fine-structure cues in detection judgments for both noise-alone and tone-plus-noise stimuli (i.e., for both hits and false alarms) for a task involving detection of low-frequency tones in narrow-band noise. The approach was straightforward and is described here in general terms, with reference to the N_0S_0 wave forms; a more explicit description is provided below in Sec. II. The Hilbert transform was used to separate the envelope and fine structure of two reproducible wave forms [see Fig. 1(a), left]. The envelopes and fine structures from the two different wave forms were then multiplied to yield two new wave forms, the chimeras [Fig. 1(a), right]. If detection judgments were *solely* determined by envelope cues, then wave forms with the same envelopes should result in the same judgments even if their fine structures differed. Conversely, if detection judgments were solely determined by fine-structure cues, then wave forms with the same fine structures should result in the same detection judgments even if their envelopes differed.

Both N_0S_0 and N_0S_π cases were studied. In the N_0S_0 case, the noise-alone (N) and diotic tone-plus-noise (T+N) wave forms were adjusted to the same overall level, so that overall energy differences were *not* a viable cue for the N_0S_0 detection task. This experimental approach was intended to force listeners to rely on temporal information for the detection task, either in the form of the envelope or fine structure. Across-wave-form level equalization was not performed for the N_0S_π case because normalization of the energy of wave forms with tones added in different phases could have introduced overall level differences and thus potential ILD cues. However, in the N_0S_π case, very small energy differences between wave forms were created by adding threshold-level tones to the noise (see Sec. II).

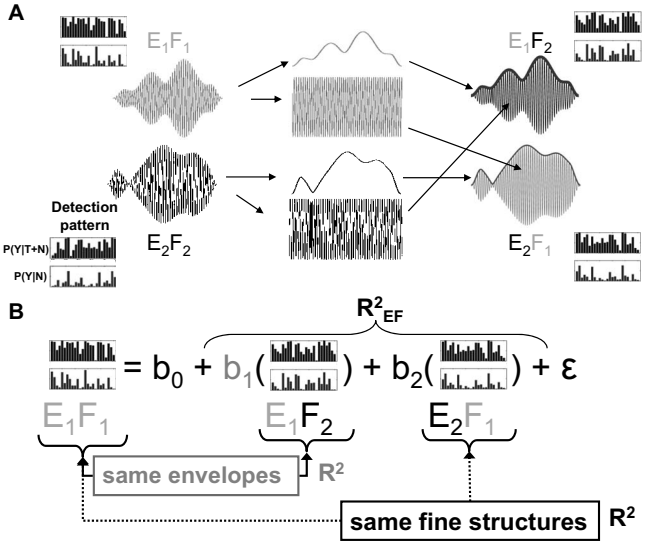


FIG. 1. (A) Schematic illustration of the stimulus-construction procedure. Envelopes (e) and fine structures (f) were separated from the E_1F_1 and E_2F_2 stimulus sets using the Hilbert transform. The envelopes and fine structures were exchanged and recombined to create chimeric stimulus sets E_1F_2 and E_2F_1. Detection patterns are shown to remind the reader that each stimulus wave form illustrated is a single member of an entire set of wave forms. A more detailed description of the stimuli (including distortion-control procedures) is given in the text. (B) Illustration of the multiple regression procedure for the E_1F_1 stimulus sets. Chimeric detection patterns sharing envelopes (E_1 in the example above) and sharing fine structure (F_1 above) were used to predict the detection pattern residuals (see text) for the base line stimulus set (E_1F_1 above). The *b* coefficients represent the slopes of the regression lines used in the multiple regression statistical model and also indicate how strongly the subject weighted the information associated with that cue. The b_0 coefficient is always equal to zero because variability linearly associated with the base line stimulus set not in the model (E_2F_2 above) was removed (see text for details). The ε term represents error variance. R^2 values were computed for envelope (gray), fine structure (black), and a linear combination of envelope and fine structure (R^2_{EF}). If envelope completely dominated the detection process, it was expected that the E_1F_1 and E_1F_2 detection patterns (and the related patterns of residuals) would be the same and the $R^2_{envelope}=1$. If fine structure completely dominated the detection process, it was expected that the E_1F_1 and E_2F_1 detection patterns (and residuals) would be the same and the $R^2_{fine\ structure}=1$.

The same uncertainty about the roles of envelope and fine structure that exist for low-frequency diotic masking also exists for low-frequency dichotic masking. Models based on interaural differences (e.g., Hafter, 1971) can be viewed as recovering ITDs based on fine structure (or perhaps envelopes) and ILDs based on envelope, whereas noise-reduction (e.g., Durlach, 1963) and correlation (e.g., Osman, 1973) models compute energylike statistics based on the entire wave form. In the study presented here, experiments using chimera stimuli were also carried out using N_0S_π reproducible noise wave forms, again with the goal of determining the relative importance of envelope and fine-structure cues in determining detection judgments. Note that common envelopes imply similar ILD distributions in the signals and that common fine structures imply similar fine-structure ITD distributions. At the target frequency of 500 Hz used in this study, fine-structure ITDs tend to dominate detection results (Bernstein and Trahiotis, 1985), although envelope ITDs would still influence the ITD distribution. Models for dichotic detection based strictly on the statistics of ILD cues would predict similar detection judgments for reproducible

wave forms that have the same envelope but different fine structures. In contrast, models based strictly on the statistics of fine-structure ITDs would predict similar detection results for wave forms that have matched fine structures. Thus, the detection results for these wave forms provide a useful test for these classes of models for dichotic detection.

II. METHODS

A. General design

Four related sets of reproducible stimuli were created as described below. Two of the sets contained "base line" stimuli, which were 25 random, narrow-band, noise-alone diotic wave forms, plus both diotic and dichotic tone-plus-noise stimuli created from these 25 noise wave forms using standard techniques (described in detail below).[1] The other two sets contained "chimeras" that were created by combining individual wave form envelopes from one of the base line sets with the individual fine structures from the other base line set. Thus, each wave form in the chimera stimulus sets shared its envelope with a wave form in one of the base line sets and shared its fine structure with a corresponding wave form in the other base line set. The relative dominance of envelope vs fine structure in the detection task was then investigated by making detailed comparisons among the probabilities of "target present" or "yes" (Y) responses for N or T+N stimuli for the four sets of stimuli. Note that the potential influence on subject responses of spectral splatter introduced in the process of combining envelopes and fine structure from different wave forms (Amenta *et al.*, 1987) was minimized by rejecting wave forms that resulted in chimeras with significantly increased bandwidth (see Appendix). Details regarding the construction of the stimuli are discussed below.

Experimental procedures adapted from those of Davidson *et al.* (2006), Evilsizer *et al.* (2002), and Gilkey *et al.* (1985) were used to obtain detection patterns for each set of base line and chimeric stimuli. Detection patterns were defined as the hit rates and false-alarm rates estimated for each of the reproducible noise maskers in a particular group of wave forms; a detection pattern can be visualized as a bar graph of hit and false-alarm rates, plotted as a function of the masker identification numbers [shown in Fig. 1(a)]. Detection patterns were constructed for the probability of Y responses for T+N stimuli [P(Y|T+N), i.e., hits] or for N stimuli [P(Y|N), i.e., false alarms]. Thus, the first probability in each P(Y|N) detection pattern shown in Fig. 1(a) is the probability of a Y response for N wave form 1 in that stimulus set. Similarly, the first probability in each P(Y|T+N) detection pattern is the probability of a Y response for the T+N stimulus created with N wave form 1 in each set. The second probability in each detection pattern is for N or T+N stimuli created with N wave form 2 in each set, etc. Detection patterns for each subject were measured for each of the four sets of stimuli (two base line sets and two chimeric sets) for both N_0S_0 and N_0S_π conditions [note that Fig. 1(a) only shows the four detection patterns for the N_0S_0 condition for one subject]. Analyses of these detection patterns for stimulus sets that had matched envelopes or matched fine structures allowed quantification of the relative contributions of envelope and fine structure to the listeners' decisions. For example, the ability to predict a subject's detection pattern for one stimulus set using that subject's detection pattern for another stimulus set that had the same envelopes (but different fine structures) would suggest that envelope cues dominated the detection results. Similarly, dominance of fine-structure cues would be indicated by the ability to predict the detection pattern based on results for another stimulus set with the same fine structures.

Six subjects, all of whom had previous listening experience, completed the experiment. S3 and S2 were the first and fourth authors of the present paper. Training and testing procedures were performed in a double-walled sound attenuating booth (Acoustic Systems, Austin, TX).

B. Stimuli

The goal of the experiment was to estimate the relative contribution of envelope and fine-structure cues in determining detection judgments when *no detectable overall energy differences* were present. The design also allowed the comparison of judgments across subjects and across interaural conditions (N_0S_0 vs N_0S_π). Generating the stimuli for the experiment is conceptually fairly simple: Create a group of narrow-band reproducible noises and interchange their envelopes and fine structures to produce chimeras. However, in practice, the need to avoid the introduction of unintended detection cues and to present comparable wave forms across subjects and under the two interaural conditions made the stimulus generation process more complicated. For example, combining envelopes and fine structures from different wave forms can produce chimeras that are wider in bandwidth than the original wave forms; therefore, stimulus selection was constrained to control this problem (see Appendix for details).

The same noise-alone (N) wave forms were used for each subject and under both interaural conditions. Tones (T) were added to these N wave forms to produce the T+N stimuli. However, because the tones were added at threshold level and threshold varied across subjects and across interaural conditions, the resulting wave forms differed somewhat across subjects and conditions.

The four sets of reproducible wave forms were created for each subject, as follows (Fig. 2): A narrow-band (50 Hz) N wave form was created as a candidate for the *i*th reproducible stimulus in one of the base line stimulus sets (E_1F_1). Base line N wave forms were created in the frequency domain by adding five frequency components (480, 490, 500, 510, and 520 Hz). The magnitudes of the five components were randomly selected from a Rayleigh distribution, and the phases of the five components were selected from a uniform distribution on the interval [−pi, pi]. The inverse Fourier transform was used to generate the time-domain noise wave forms. All wave forms were 100 ms in duration, with 10-ms $\cos^2$ on/off ramps. Each of the N wave forms was normalized to an overall level of 57 dB SPL (sound pressure level), which corresponds to a 40-dB SPL spectrum level, N_0, for a bandwidth of 50 Hz.

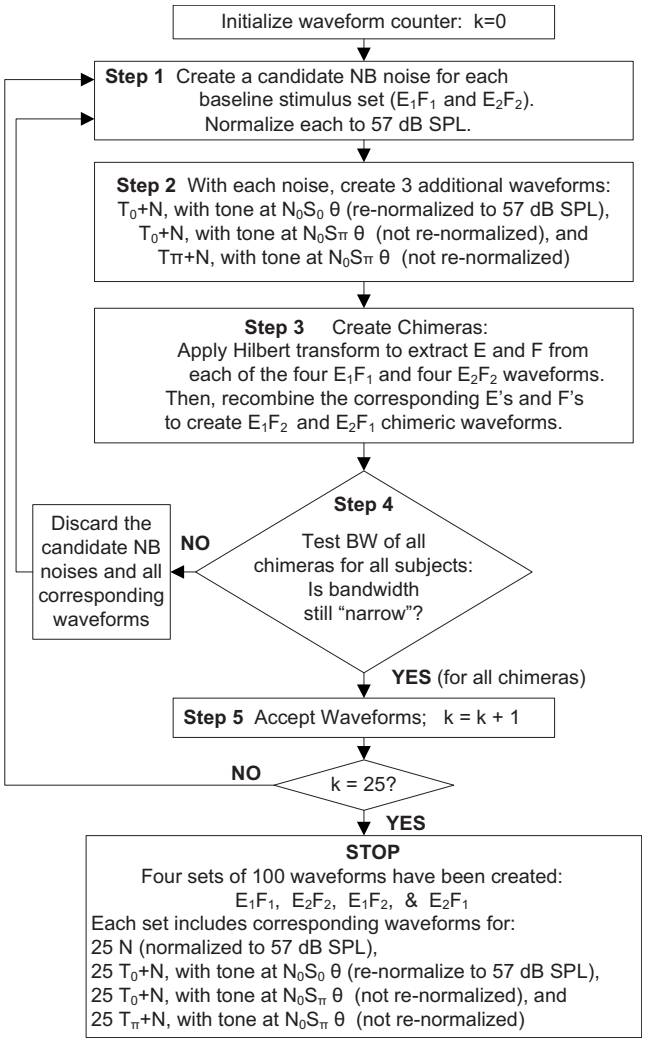


FIG. 2. Flow chart illustrating procedure for stimulus generation. First, two random Gaussian narrow-band noises are created, one for the E_1F_1 base line stimulus set and one for the E_2F_2 set. The noises are normalized to 57 dB SPL, and then tones are added at the appropriate levels and phases for tests of detection in the N_0S_0 and N_0S_π interaural configurations. These wave forms are then used to create chimeras that have the envelopes from one wave form and the fine structure from another, and vice versa [see Fig. 1(a)]. Before accepting the original and chimera N and T+N wave forms into the four sets of wave forms to be used for testing, the bandwidths of the chimeras are checked (see Appendix for details). If all of the chimera wave forms are sufficiently narrow band, then all base line and chimera wave forms are accepted (N, T_0+N for N_0S_0, T_0+N for N_0S_π, and T_π+N for N_0S_π for each of the wave form sets: E_1F_1, E_2F_2, E_1F_2, and E_2F_1). For testing in the N_0S_0 condition, diotic N or T+N stimuli are used based on tone levels at the N_0S_0 thresholds, and all wave forms are normalized to 57 dB SPL. For N_0S_π testing, stimuli are diotic N stimuli on noise-alone trials (normalized to 57 dB SPL); on signal trials, T_0+N is presented to one ear and T_π+N is presented to the other ear (these are not normalized but differ only slightly from 57 dB SPL, see text). For N_0S_π stimuli, tone levels were based on N_0S_π thresholds during training. Abbreviations: NB, narrow band; E, envelope; F, fine structure; θ, detection threshold (during training for each subject); T_0, tone with zero phase re stimulus onset; T_π, tone with π phase re stimulus onset; BW, bandwidth.

Using the same procedure as above, another random noise was created as a candidate for the *i*th stimulus in the other base line stimulus set (E_2F_2). These *i*th N candidate wave forms in each of the base line stimulus sets were then used to create three other wave forms: (1) The *i*th T+N wave form for the N_0S_0 condition was created by adding a 500-Hz tone at 0 phase with respect to stimulus onset (T_0+N). The tone level was set to the listener's N_0S_0 detection threshold, as determined during training. The stimulus was windowed with 10-ms $\cos^2$ ramps and then re-normalized to 57 dB SPL to remove overall level differences as potential cues for discrimination between N and T+N wave forms in the N_0S_0 condition. (2) The *i*th T+N used for one of the ears in the N_0S_π condition was created by adding a 500-Hz tone at 0 phase and windowing with 10-ms $\cos^2$ ramps. The tone level was matched to the subject's N_0S_π detection threshold, determined during training. This T_0+N wave form was not re-normalized to avoid adding undesired interaural level cues in the N_0S_π condition (see below). (3) The *i*th T+N for the opposite ear in the N_0S_π condition was created by adding a 500-Hz tone at π phase (T_π+N) and by windowing with 10-ms $\cos^2$ ramps. Again, the tone level was matched to the subject's N_0S_π detection threshold. This wave form was also not re-normalized. In the N_0S_π condition, the un-normalized T_0+N was presented to one ear, and T_π+N was presented to the other ear. Because the T+N wave forms used for the N_0S_π condition were not normalized, level differences did exist across the stimuli used in the N_0S_π condition; however, the average level difference between N and T+N stimuli under the N_0S_π condition was 0.09 dB, and level varied across T+N wave forms with a standard deviation of 0.7 dB.

Next, each wave form in E_1F_1 base line set and the corresponding wave form in the E_2F_2 base line set (i.e., the *i*th N wave form in the E_1F_1 set with the *i*th N wave form in the E_2F_2 set, the *i*th T_π+N wave form in the E_1F_1 set with the *i*th T_π+N wave form in the E_2F_2 set, etc.) were used to create two chimeric wave forms (one for the E_1F_2 set and one for the E_2F_1 set) as follows: The Hilbert transform was used to compute the envelope and fine structure for each wave form in the base line set, and then the envelopes from the E_1F_1 wave form and fine structures from the E_2F_2 wave form were combined (multiplied) to create the corresponding E_1F_2 chimeric wave form [Fig. 1(a), right]. Similarly, the fine structure from the E_1F_1 wave form and the envelope from the E_2F_2 wave form were combined to create the corresponding E_2F_1 chimeric wave form. The chimeric wave forms were then tested to ensure that they were still narrow band (see Appendix for details). If any of the chimeras failed the bandwidth test, then all of the associated wave forms (i.e., the *i*th N, T_0+N for N_0S_0, T_0+N for N_0S_π, and T_π+N for N_0S_π) in each of the four sets of reproducible stimuli ($E_1F_1, E_2F_2, E_1F_2, E_2F_1$) were discarded, and the process to create the *i*th wave forms in each set was re-initiated (Fig. 2). If all of the chimeras passed the bandwidth test, then the *i*th wave form of each stimulus type was accepted into each of the four stimulus sets, and the procedure moved on to the (i+1)th wave forms. The stimulus generation process was continued until there were 25 wave forms of each type (N, T_0+N for N_0S_0, T_0+N for N_0S_π, and T_π+N for N_0S_π) in each of the four sets ($E_1F_1, E_2F_2, E_1F_2, E_2F_1$).

Note that the ensemble of stimuli was specific to each subject because early in the stimulus generation procedure the tones were added to the base line N wave forms at threshold levels determined for each subject. However, the differences in the tone levels across subjects did not result in sig-

nificant differences in the bandwidths of the chimeras. So, although the 25 E_1F_1 base line N wave forms and the 25 E_2F_2 base line N wave forms were identical for all subjects, the various T+N wave forms differed across subjects, as is true in any study with reproducible maskers (because subjects have different detection thresholds); however, these wave forms were "comparable" across subjects, as explored by cross-subject comparisons in the analyses of the detection results.

Stimuli were created using custom MATLAB software (Mathworks, Natick, MA) and were presented using a TDT System III (Tucker Davis Technologies, Gainesville, FL) RP2 digital-analog converter (48 828 Hz sampling rate, 24 bits/sample) over TDH-39 headphones (Telephonics, Corp., Farmington, NY).

C. Training

Training stages were similar to those described in Davidson *et al.* (2006) and are summarized here. The extensive training paradigm was designed to encourage subjects to develop a consistent detection strategy at threshold-level performance that would remain constant over the duration of the experiment (threshold was defined here for each subject and each interaural condition as the E_S/N_0 value in decibels, where the d' for yes/no testing, $d'_{Y/N}$, was approximately equal to 1).[2] The final testing procedure was a single-interval task without trial-by-trial feedback, but early in training other procedures were used to help subjects learn acoustic cues that could be used to determine the presence of the signal. Three separate training tasks were completed, and each task was progressively more similar to the final testing procedure. The training procedures used approximately 50-Hz bandwidth, 100-ms duration noise wave forms that were generated randomly on each trial (i.e., they were not reproducible stimuli as used in the testing procedure, and they were not chimeras). The training noises contained the same five frequency components as the testing noises. Randomly generated noises were used to prevent any possible learning of reproducible stimuli. Training stimuli were normalized with the same procedures as the testing stimuli; that is, all N_0S_0 N and T+N stimuli and N_0S_π N stimuli were normalized to 57 dB SPL, while N_0S_π T+N stimuli were not re-normalized after addition of the tone.

The following training and testing procedures were conducted under both the N_0S_0 and N_0S_π interaural conditions. In general, subjects received only one type of interaural stimulus condition per session (2–3 h). For S1, S3, and S4, the initial interaural condition was randomized across subjects, and the use of N_0S_0 or N_0S_π stimuli alternated by session. S1, S3, and S4 had relatively small differences between thresholds for the diotic and dichotic conditions, which raised the question as to whether the alternation of interaural stimulus conditions across sessions may have affected their results due to the possible confusion of the diotic and dichotic cues. Therefore, S2, S5, and S6, who were tested later, were trained and tested completely in one interaural condition before moving on to the other conditions. The initial interaural condition was also randomized across this subset of subjects. (As a further test, S3, who initially alternated interaural conditions by session, subsequently repeated the entire experiment but completed the N_0S_π interaural condition first, followed by the N_0S_0 condition. Detection patterns from the two training and testing orders for this subject were highly correlated.) In rare cases, stimuli from both interaural conditions were presented during the same session (e.g., to finish a particular training or testing paradigm). During those sessions, presentation of the individual blocks of stimuli never alternated between the two conditions.

During the first training stage, each subject completed 10–15 tracks in a two-interval two-alternative forced-choice, 2-down/1-up tracking procedure with trial-by-trial feedback to estimate a level where $d'_{2AFC}=0.77$. Each track had a fixed length of 100 trials. The step size was maintained at 4 dB for the first two reversals and dropped to 2 dB thereafter. Thresholds were estimated by averaging tone levels at all but the first four or five reversals in the track such that an even number of reversals was averaged. Subjects were instructed to "select the interval containing the tone" and learned the task based on trial-by-trial feedback.

During the second training stage, a single-interval, fixed-level task was used to familiarize the subject with the task that would be used during testing; however, trial-by-trial feedback was provided to help subjects stabilize their performance near threshold during this training stage. The instructions for the single-interval tasks were to "determine whether the tone was present" on each trial and to click on buttons labeled "tone" and "no tone." Approximately ten blocks, containing 100 trials each, were completed at +3, +1, and −1 dB relative to the threshold established in the two-interval task. The d' values calculated from these blocks were used to estimate the tone level where d' was approximately equal to unity, rounded to 1-dB resolution. Approximately ten blocks were then run at that tone level. Throughout the single-interval training procedures (and the testing procedure described in Sec. II D), d' and bias (β, Macmillan and Creelman, 1991) were monitored. (Note that d' with no subscript refers to d' for yes/no testing, which was used throughout the rest of the study.) If a subject's threshold changed, the tone level was adjusted again with 0.5 or 1-dB resolution until d' returned to unity.

After a stable tone level was established, the trial-by-trial feedback was removed, and subjects completed approximately ten 100-trial blocks without feedback in order to determine whether d' values remained near unity after feedback was removed. In rare cases, tone levels were further adjusted with 0.5- or 1-dB resolution such that $d' \approx 1$. The block length was then increased to 400 trials, and subjects completed five more blocks.

If a listener was noticeably biased (i.e., β departed by more than 15% from unity, with unity indicating an equal probability of responding "tone" or "no tone"), the subject was given verbal feedback after the session to "try and make an equal number of tone and no tone responses." Subjects were informed of the value of β after each block, and they were notified that $\beta<1$ indicated too many "tone" responses, and $\beta>1$ indicated too many "no tone" responses. The val-

ues of d' and β were computed using P(Y|T+N) (the probability of a "yes" response conditional on a T+N trial, or hit rate) and P(Y|N) (the probability of a "yes" response conditional on an N trial, or false-alarm rate).

D. Testing

The testing stage was identical to the final training stage except that the reproducible noises described in Sec. II B were used as stimuli. Before each 400-trial block, 20 practice trials (that did not use reproducible stimuli or chimeras) were presented with feedback. The 20 practice trials were presented with tone levels 2 dB above the tone level used for testing. For each 400-trial block, which included only one interaural condition, the appropriate T+N (25 stimuli) and N (25 stimuli) from each of the four stimulus sets were presented twice each in a randomly interleaved order. A total of 50 blocks were presented to each listener under each interaural condition such that 100 presentations of each T+N and each N wave form were presented at the final tone level.

The narrow-band-noise wave forms used in training were random and did not include chimeric stimuli. As a result, the tone level determined from the training procedure did not necessarily represent the level where $d' \approx 1$ for each subject when using the sets of reproducible noise wave forms. In these cases, the tone level was adjusted in 0.5- or 1-dB steps until $d' \approx 1$ for each subject, and data collection was re-initiated for that subject. In practice, the tone level was adjusted at least once for each listener, which was most likely a consequence of the specific stimuli selected with the distortion-control algorithm (described in the Appendix). Learning was unlikely to occur during this process because the long training procedure with feedback was designed to encourage subjects to establish a fixed decision strategy. Trial-by-trial feedback was never presented while testing with the reproducible noise wave forms. Values of d' and β were computed across the combination of all stimulus wave forms from the four stimulus sets (i.e., E_1F_1, E_1F_2, etc.), and were not monitored within each of the sets. No attempt was made to control for variations in values of d' and β computed for the individual envelope and fine-structure sets of stimuli (e.g., E_1F_1) during the course of the experiment.

III. RESULTS

The analyses of the experimental results are presented below in several sections. First, the reliability of the data is addressed. Next, detection patterns estimated with the base line and chimeric stimuli are compared within subjects to determine the relative contributions of envelope and fine-structure cues used in the detection task. Detection patterns are then compared between subjects to determine if the cues or detection strategies used by the different subjects were similar. Finally, detection patterns are compared between interaural conditions to determine if any similarities in detection cues occurred between the diotic and dichotic conditions. The analyses considered detection patterns constructed from the proportion of "yes" responses to N wave forms [P(Y|N)] and to T+N stimuli [P(Y|T+N)]. For each stimulus set and for each interaural condition, these two detection patterns (each having 25 elements) were also combined into one larger detection pattern (with 50 elements) to create P(Y|W), where W refers to one of the 50 T+N or N stimuli. To compare detection patterns (i.e., P(Y|N), P(Y|T+N), or P(Y|W)), they were first converted to z-scores (i.e., relative to the standard normal distribution),[3] so that the predicted relation between detection patterns was linear. Detection patterns were then compared, both within and across subjects, using regression techniques, as further described below.

TABLE I. Performance and reliability statistics for P(Y|W) under the N_0S_0 interaural condition. One tone level (E_S/N_0 in dB^2) was used for each subject. Overall d' and β were computed using responses to wave forms in all stimulus sets. Individual d' and β values are given for one base line stimulus set and for one chimera stimulus set. Values for the other stimulus sets are available in Davidson (2007). The coefficient of determination between responses from the first and the last half of the trials (r^2_{12}) and the proportion of predictable variance (r^2_{max}) are given for each condition. All r^2 values were significant ($p<0.05$).

	Overall				Per stimulus set		P(Y|W)	
S	E_S/N_0	d'	β	Stimulus set	d'	β	r^2_{12}	r^2_{max}
S1	10	0.87	0.94	E_1F_1	0.95	0.70	0.93	0.97
				E_1F_2	0.76	1.05	0.95	0.98
S2	10	0.88	0.99	E_1F_1	1.01	0.87	0.93	0.97
				E_1F_2	0.63	1.03	0.95	0.97
S3	10	1.02	1.07	E_1F_1	0.86	1.02	0.88	0.94
				E_1F_2	1.01	1.10	0.89	0.94
S4	11	0.96	0.95	E_1F_1	0.85	0.85	0.93	0.96
				E_1F_2	0.95	0.93	0.93	0.97
S5	11	0.86	0.99	E_1F_1	0.51	0.88	0.95	0.97
				E_1F_2	0.68	0.97	0.95	0.97
S6	11.5	0.94	0.97	E_1F_1	0.79	0.81	0.89	0.94
				E_1F_2	1.05	1.04	0.89	0.94
S_{avg}	10.58	0.92	0.98	E_1F_1	0.82	0.85	0.98	0.99
				E_1F_2	0.84	1.02	0.98	0.99

Two conflicting problems arise when using these techniques. On the one hand, when correlating z-scores based on P(Y|W), the value of the correlation coefficient r is a function of d'; that is, as d' goes to infinity (for both of the detection patterns being compared), r goes to 1.0. Thus, correlations of P(Y|W), which include detection patterns for responses to both N and T+N stimuli, are influenced by the value of d', and high r values do not necessarily indicate that there is a relation between the two cases in terms of underlying processing. On the other hand, the approach of analyzing P(Y|T+N) and P(Y|N) results separately means that the range of observed proportions of "yes" responses is almost certainly truncated, forcing an artificial reduction in r. By analyzing all three detection patterns [i.e., P(Y|W), P(Y|N), and P(Y|T+N))], it was possible to evaluate the relations between the full detection patterns (P(Y|W) while safeguarding against artifactually high r values introduced by conditions with higher values of d'.

A. Reliability of the data and detection performance

Tables I–IV show reliability and detection performance statistics for each individual subject and also the average across subjects (S_{avg}) under both the N_0S_0 (Tables I and II)

TABLE II. Reliability statistics for P(Y|T+N) and P(Y|N) under the N_0S_0 condition. The χ^2 statistic, coefficient of determination between responses from the first and the last half of the trials (r_{12}^2), and the proportion of predictable variance (r_{max}^2) are shown. All χ^2 values were significant ($p<0.001$), and all r^2 values were also significant ($p<0.05$).

		P(Y|T+N)			P(Y|N)		
S	Stimulus set	χ^2	r_{12}^2	r_{max}^2	χ^2	r_{12}^2	r_{max}^2
S1	E_1F_1	1371	0.91	0.95	1829	0.91	0.95
	E_1F_2	2198	0.94	0.97	2078	0.95	0.97
S2	E_1F_1	1543	0.89	0.94	1856	0.92	0.96
	E_1F_2	1737	0.94	0.97	1779	0.94	0.97
S3	E_1F_1	669	0.73	0.85	1011	0.85	0.92
	E_1F_2	488	0.61	0.77	1431	0.89	0.94
S4	E_1F_1	1350	0.91	0.95	1340	0.89	0.94
	E_1F_2	940	0.86	0.93	1628	0.92	0.96
S5	E_1F_1	2352	0.95	0.97	3017	0.95	0.98
	E_1F_2	1645	0.90	0.95	2310	0.96	0.98
S6	E_1F_1	1258	0.75	0.86	1645	0.93	0.96
	E_1F_2	1113	0.77	0.87	1620	0.87	0.93
S_{avg}	E_1F_1	4873	0.94	0.97	7659	0.98	0.99
	E_1F_2	3912	0.93	0.96	8530	0.98	0.99

and N_0S_π (Tables III and IV) interaural conditions. These tables only include detailed results for one set of base line stimuli and one set of chimera stimuli; results for the other two sets of wave forms were comparable, as expected, and are available in Davidson (2007). Tables I and III summarize data combined over both N and T+N wave forms [i.e., P(Y|W)]; Tables II and IV separate the N and T+N data [i.e., P(Y|N) and P(Y|T+N)]. The threshold tone level used for the experiment (where $d' \approx 1$) is given in terms of E_S/N_0 for each subject and condition.[2] (Note that because the N_0S_0 stimuli were all normalized to 57 dB SPL, changes in E_S/N_0 do not indicate changes in level between N and T+N stimuli. Also, note that because the level differences were eliminated from the N_0S_0 stimuli, the difference between N_0S_0 thresholds and N_0S_π thresholds is not necessarily comparable to masking level differences reported in other studies.) The actual d' and β values calculated across and within the four stimulus sets are also shown. The training procedure was relatively successful in finding overall d' values near 1 with the possible exception of S1 in the N_0S_π condition ($d'=0.78$ in Table III). For individual sets of N_0S_0 stimuli, d' values ranged from 0.51 to 1.14, and β values ranged from 0.70 to 1.32 (Table I). For individual sets of N_0S_π stimuli, d' values ranged from 0.54 to 1.11, and β values ranged from 0.57 to 1.35 (Table III).

TABLE III. Same as Table I but for the N_0S_π interaural condition. All r^2 values were significant ($p<0.05$).

	Overall				Per stimulus set		P(Y|W)	
S	E_S/N_0	d'	β	Stimulus set	d'	β	r_{12}^2	r_{max}^2
S1	0	0.78	0.91	E_1F_1	1.10	0.57	0.93	0.97
				E_1F_2	0.66	0.97	0.93	0.96
S2	−10	0.97	1.10	E_1F_1	0.85	1.35	0.90	0.95
				E_1F_2	1.09	0.96	0.94	0.97
S3	−17	1.01	0.99	E_1F_1	0.94	1.11	0.89	0.94
				E_1F_2	1.06	0.92	0.86	0.92
S4	−1	0.93	1.00	E_1F_1	0.92	0.85	0.90	0.95
				E_1F_2	0.79	1.03	0.93	0.97
S5	−16.5	0.91	1.02	E_1F_1	0.90	1.24	0.92	0.96
				E_1F_2	1.09	1.11	0.92	0.96
S6	−10	0.96	1.06	E_1F_1	0.87	0.95	0.89	0.95
				E_1F_2	1.08	0.96	0.87	0.93
S_{avg}	−9.08	0.92	1.01	E_1F_1	0.89	1.00	0.98	0.99
				E_1F_2	0.96	0.99	0.97	0.98

TABLE IV. Same as Table II but for the N_0S_π interaural condition. All χ^2 values were significant ($p<0.001$), and all r^2 values were significant ($p<0.05$).

		P(Y|T+N)			P(Y|N)		
S	Stimulus set	χ^2	r_{12}^2	r_{max}^2	χ^2	r_{12}^2	r_{max}^2
S1	E_1F_1	885	0.94	0.97	1621	0.89	0.94
	E_1F_2	1859	0.94	0.97	1938	0.90	0.95
S2	E_1F_1	1283	0.89	0.94	639	0.75	0.86
	E_1F_2	1188	0.91	0.95	970	0.85	0.92
S3	E_1F_1	844	0.81	0.90	530	0.75	0.86
	E_1F_2	909	0.72	0.84	366	0.63	0.78
S4	E_1F_1	921	0.90	0.95	1100	0.82	0.90
	E_1F_2	1390	0.91	0.95	1625	0.91	0.95
S5	E_1F_1	1388	0.89	0.95	623	0.87	0.93
	E_1F_2	1220	0.88	0.94	614	0.74	0.86
S6	E_1F_1	1438	0.87	0.93	490	0.73	0.85
	E_1F_2	1159	0.77	0.87	654	0.72	0.85
S_{avg}	E_1F_1	3089	0.98	0.99	1287	0.87	0.93
	E_1F_2	1953	0.91	0.95	978	0.83	0.91

Tables II and IV include χ^2 statistics, with larger values indicating that variations in the subjects' responses were tied to across-wave-form changes in the reproducible stimuli and not due to chance alone (Siegel and Colburn, 1989). All χ^2 values greatly exceeded the threshold for significance ($\chi_{0.01}^2=43$), demonstrating that the between-wave-form differences in hit rates and the between-wave-form differences in false-alarm rates were reliable. Note that the χ^2 values observed for S_{avg} under the N_0S_π condition (Table IV) were low relative to those under the N_0S_0 condition (Table II), suggesting that although the detection patterns for individual subjects were reliable, there were individual differences that "diluted" the detection patterns when averaged across subjects, particularly under the N_0S_π condition.

Tables I–IV also include squared first-half, last-half correlation coefficients (r_{12}^2). Again these values were high and significant, indicating that the subjects' responses were driven by the stimulus in a manner that was consistent across time. The value of r_{12}^2 is directly related to r_{max}^2, the maximum proportion of predictable variance that can be expected when comparing detection patterns across stimulus sets, interaural conditions, or subjects[4] (also shown in Tables I–IV). The value of r_{max}^2 exceeds the value of r_{12}^2 because the first-half and last-half detection patterns are necessarily based on half as many trials as the overall (i.e., first half and last half combined) detection patterns (50 vs 100 trials).

B. Within-subject comparisons of detection patterns estimated with base line and chimeric stimuli

Recall that the overall logic of the experiment and analysis is fairly straightforward. If subjects used only envelope cues to make detection judgments, then stimulus sets that shared the same envelopes (e.g., E_1F_1 and E_1F_2) should have produced the same detection patterns. If fine structure were the only relevant cue, then stimulus sets with the same fine structures (e.g., E_1F_1 and E_2F_1) should have produced the same detection patterns. If a linear combination of envelope and fine structure was relevant, then it should be possible to combine two detection patterns to predict a third (e.g., E_1F_2 and E_2F_1 could be used to predict E_1F_1). The general strategy was then to predict each of the base line detection patterns (E_1F_1 or E_2F_2) with the chimeric detection patterns (E_2F_1 and E_1F_2) using multiple regression.[5] The multiple regression method used for predicting the detection pattern for the E_1F_1 stimulus set [as shown schematically in Fig. 1(b)] is described here. The method for predicting the E_2F_2 detection pattern is equivalent, but the subscripts "1" and "2" would be exchanged in the description.

This correlation analysis assumes that the E_1F_1 and E_2F_2 stimulus sets are statistically independent of each other and that the detection patterns for these two stimulus sets are also statistically independent. Chance similarities, reflected in nonzero correlations, between the E_1F_1 and E_2F_2 detection patterns could cause misleading correlations between the detection patterns for E_1F_2 and E_2F_1. If this were true, there could be nonzero correlations between response patterns to E_1F_1 and E_2F_2, even in the case in which envelope manipulations had absolutely no effect on responses. Such misleading correlations are referred to as "false" correlations. These false correlations were avoided by considering only the components of the detection patterns that were uncorrelated with the original E_2F_2 pattern. This "partialing-out" approach, described below, automatically excludes correlations across response sets that arise from correlations in the original wave forms. These correlations could be based on similarities in the envelopes or fine structures of the original wave form sets but could also be due to any other response-determining components that are shared by the original wave form sets (i.e., the E_1F_1 and E_2F_2 sets). To avoid this potential problem, any such nonzero correlations were statistically "partialed out" by separately regressing the E_1F_1, E_1F_2, and E_2F_1 z-score detection patterns on the E_2F_2 detection pattern and then using the residuals from each of these regressions in the subsequent analyses. The residuals of the regression of any detection pattern on the E_2F_2 detection pattern are by definition *not* correlated with the E_2F_2 pattern. Thus, by using these residual detection patterns in further analyses, any chance similarity in the variability of the detection patterns to that associated with the E_2F_2 pattern were "blocked" or "removed" from the other three patterns (i.e., the *residual* detection patterns for E_1F_1, E_1F_2, and E_2F_1 were uncorrelated with the detection pattern for E_2F_2). Note that the residuals were computed with separate regressions for the $P(Y|W)$, $P(Y|N)$, and $P(Y|T+N)$ data.

Next, two simple linear regressions were performed to predict the E_1F_1 detection pattern residuals using either the E_2F_1 or E_1F_2 detection pattern residuals as the predictor. These regressions indicated the proportion of variance explained (R^2) by the fine structure (because F_1 was held constant) and by the envelope (because E_1 was held constant), respectively. Next, the E_1F_1 detection pattern residuals were simultaneously regressed on the E_2F_1 and E_1F_2 detection pattern residuals to compute the proportion of variance explained by the multiple regression or a linear combination of both envelope and fine structure. Incremental-F tests (Edwards, 1979) were performed to determine if the proportion of predicted variance in the E_1F_1 detection pattern residuals were significantly increased by incorporating fine-structure information (the E_2F_1 residuals) in addition to envelope alone (the E_1F_2 residuals) or by incorporating envelope information in addition to fine-structure information alone.

Because the findings when predicting E_1F_1 and when predicting E_2F_2 were comparable,[6] the results were combined and are shown as scatter plots in Figs. 3 and 4. In each panel, the detection pattern residuals estimated from responses to chimeric stimuli were used to predict the detection pattern residuals estimated for the E_1F_1 or E_2F_2 stimulus sets. The predictors are plotted on the abscissa of each panel. Envelope-based predictions are always shown with gray squares, and fine-structure-based predictions are shown with black circles. The regression lines and the slopes of the linear regressions [b_E and b_F, see Fig. 1(b)] are shown in each panel. The slope values were computed using the multiple regression procedure (i.e., both envelope and fine structure were predictors) and thus differed slightly from the slopes that would be obtained using either envelope or fine structure individually (as discussed below). Slopes for the individual envelope-based and fine-structure-based predictions (corresponding to the individual envelope and fine structure R^2 values) are not shown. If the fine structure were a perfect predictor of the variance in the detection pattern residuals, the black circles would fall exactly along the diagonal and b_F would equal 1. Conversely, if the envelope were a perfect predictor of variance in the detection pattern residuals, the gray squares would fall exactly along the diagonal and b_E would equal 1.

Three R^2 values are shown in each panel (Figs. 3 and 4). The R_E^2 corresponds to the prediction using only envelope information (i.e., using the detection pattern residuals for the stimulus set with the same envelopes as a predictor). R_F^2 corresponds to the prediction using only fine-structure information (i.e., using the detection pattern residuals for the stimulus set with the same fine structures as a predictor). R_{EF}^2 corresponds to the prediction using a linear combination of envelope and fine-structure information with weights given by b_E and b_F, respectively. Significant R^2 values are denoted with an asterisk. Incremental-F tests (p_{incr}) were used to determine if the addition of envelope as a predictor to a prediction based only on fine structure, or the addition of fine structure as a predictor to a prediction based only on envelope, significantly increased the amount of predictable variance in the detection pattern residuals. Note that the

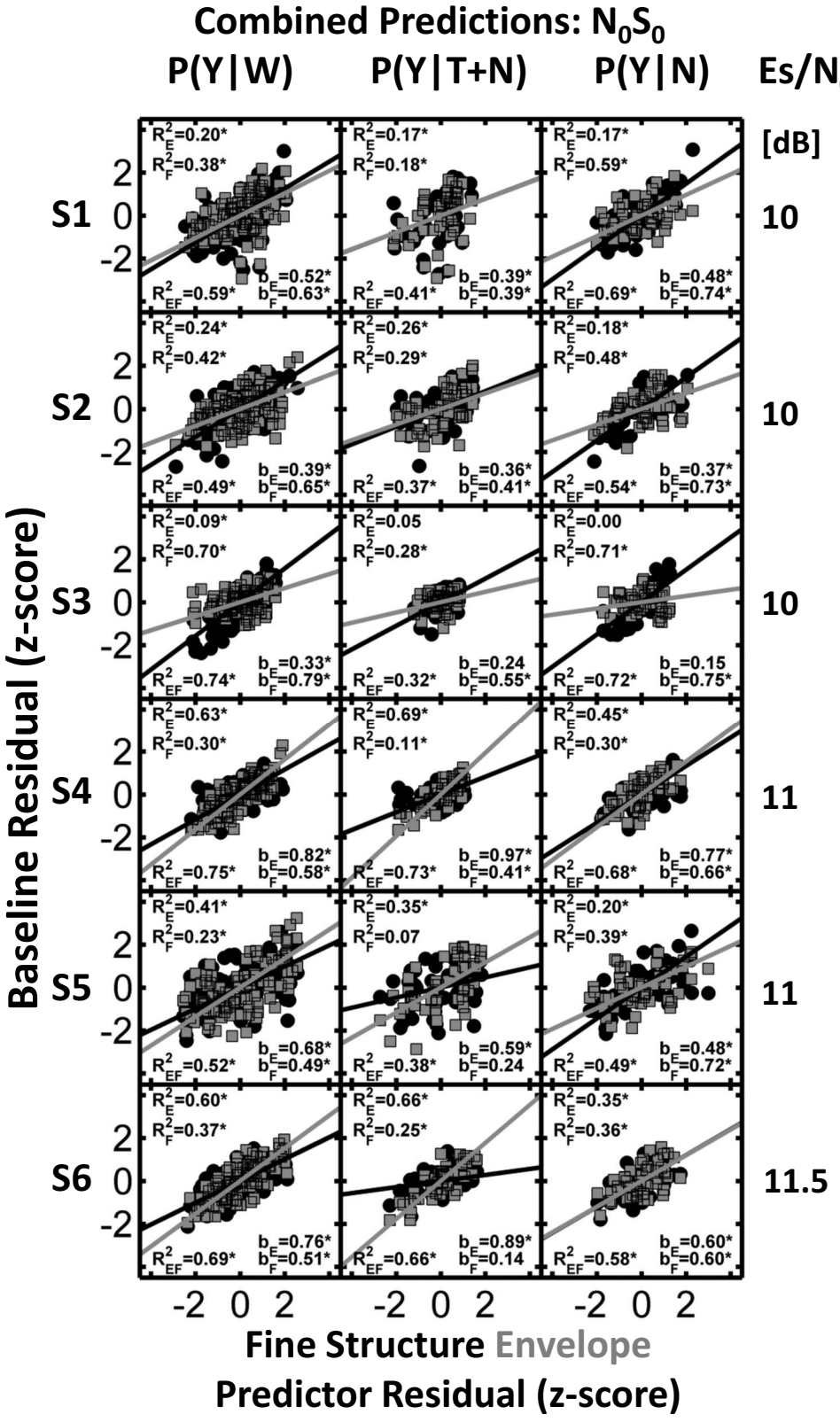


FIG. 3. Combined predictions for E_1F_1 and E_2F_2 for the N_0S_0 conditions. Comparisons between cue conditions for P(Y|W), P(Y|T+N), and P(Y|N) (columns) for six subjects (rows) using z-score residuals to characterize the probabilities. Envelope-based predictions are shown with gray squares, while fine-structure-based predictions are shown with black circles. Relative weights for each cue are shown by the b_E and b_F values, with asterisks indicating significant slopes (i.e., significant incremental-F-test results). R^2_E and R^2_F indicate the proportion of predictable variance based on the individual cues (asterisks indicate significant values). R^2_{EF} corresponds to the proportion of predictable variance using a linear combination of both envelope and fine structure. Signal-to-noise ratio (E_S/N_0 in dB) is shown to the right of each plot. See text for details.

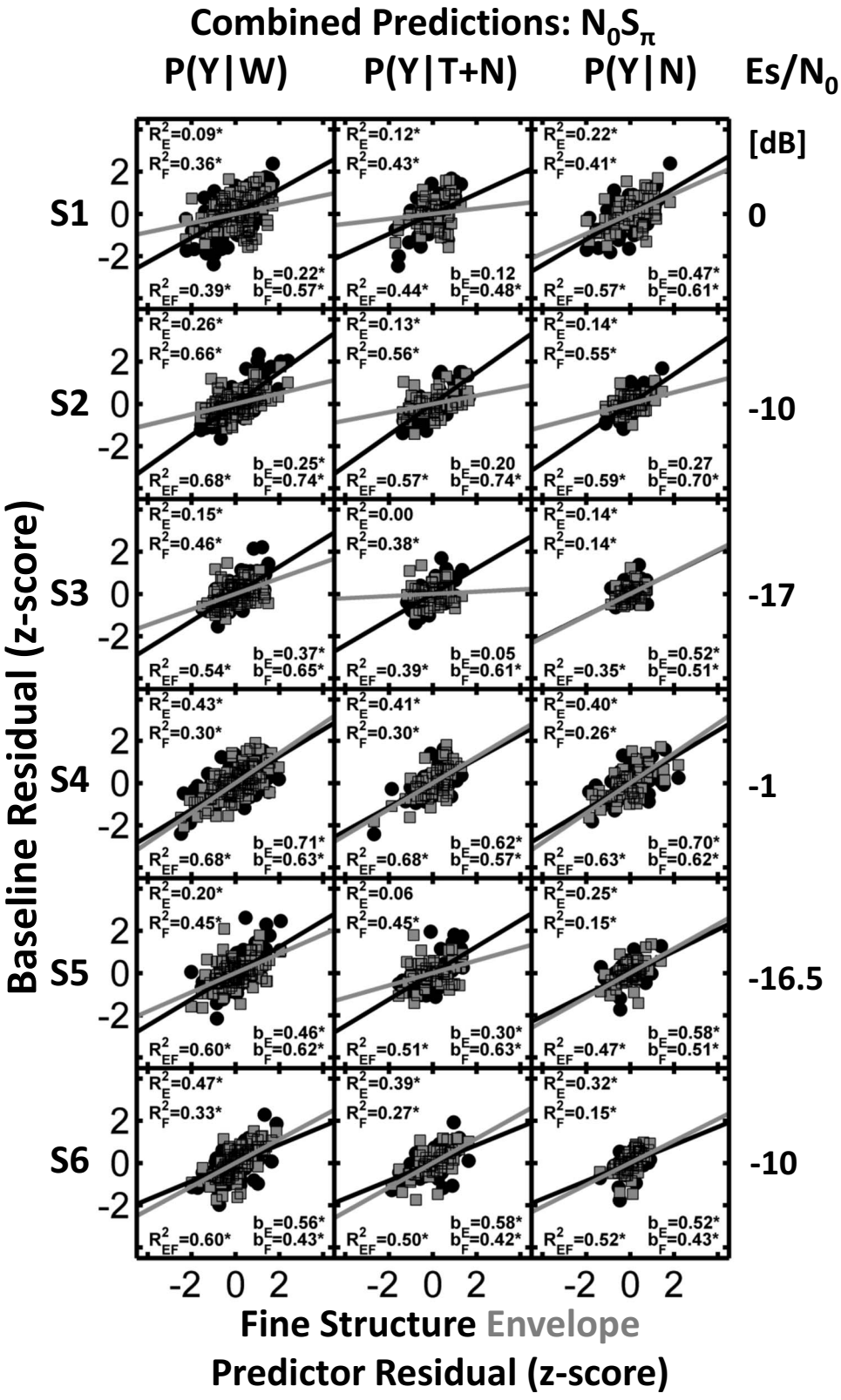


FIG. 4. Same as Fig. 3 except that predictions were made for the N_0S_π conditions.

incremental-F test is equivalent to testing whether b_E or b_F is significantly different from zero; significant values are indicated by asterisks.

1. N_0S_0 *stimuli*

Inspection of Fig. 3 reveals that both envelope and fine structure show, in general, significant correlation with the responses of each listener under the N_0S_0 condition (in all but four cases). The large number of significant p_{incr} values (i.e., significant b values) indicates that for most subjects, both envelope and fine structure contributed unique information that was correlated with the listeners' decision variables. However, there was some intersubject variability in the R^2 values observed in Fig. 3. In previous N_0S_0 detection experi-

ments in which energy was not equalized, subjects' detection patterns were highly correlated with one another (e.g., Evilsizer *et al.*, 2002; and Davidson *et al.*, 2006). These high correlations indicate that the same or very similar decision variables were used by each subject in those studies. Recall that in the N_0S_0 condition of this experiment, overall stimulus levels were equalized to remove the availability of energy as a decision variable. As a result, high intersubject correlations were not necessarily expected (Sec. III C includes a more complete discussion of intersubject correlations), nor was the use of identical decision variables across subjects. In fact, the results shown in Fig. 3 (in addition to the relatively low intersubject correlations described in Sec. III C) suggest the use of different detection strategies by different subjects. The b and R^2 values for subject 3 suggest the dominance of cues related to the fine structure of the stimulus wave forms rather than envelopes of the stimulus wave forms. The remaining subjects used a combination of fine-structure and envelope-related cues, as indicated by the b and R^2 values for envelope predictors with respect to the b and R^2 values for fine-structure predictors.

Although the majority of R^2_{EF} values in Fig. 3 are significant, the values range from 0.32 to 0.75, and all are substantially lower than the estimates of the proportions of predictable variance shown in Tables I and II. That is, these linear multiple regression models do not provide a satisfactory description of these data. On the other hand, visual inspection of the scatter plots in Fig. 3 does not suggest that a nonlinear model would perform substantially better. Implications of these results for modeling are discussed in Sec. III D.

2. N_0S_π stimuli

In the case of dichotic detection, the comparisons of detection results can be interpreted in terms of interaural differences. That is, the distributions of ILDs were similar for stimuli that had matched envelopes, and the distributions of fine-structure ITDs were similar for stimuli that had matched fine structures. Figure 4 shows the results of the same kind of regression analyses for the N_0S_π condition. Note that the threshold tone level under the N_0S_π condition varied widely across subjects. Subjects with similar thresholds also had more similar detection patterns, so the results are described for pairs of subjects with similar N_0S_π thresholds. Subjects S3 and S5 had the lowest threshold tone levels and showed similar trends in terms of envelope and fine-structure predictions. The linear combination of envelope and fine structure failed to predict the majority of the variance in the $P(Y|N)$ data. Predictions for the $P(Y|T+N)$ data indicated a stronger reliance on fine structure, but they failed to predict more than approximately half the variance in the base line detection pattern residuals. Predictions for $P(Y|W)$ indicate that fine-structure information dominated the detection process for these two subjects but also showed a significant contribution of cues based on the envelope. The multiple regression model predicted 54% and 60% of the variance in the base line detection pattern residuals for these subjects. Slightly larger weights (b values) were found in the best fits for cues derived from the fine structure as compared to the weights for fits derived from the envelope.

Subjects S2 and S6 were tested with threshold tone levels about 7 dB higher than subjects S3 and S5. Subject S2 showed consistent dominance of fine-structure-based cues over envelope-based cues. The multiple regression model explained 68% of the variance in the $P(Y|W)$ residuals for this subject. Results for subject S6 indicated a stronger contribution of envelope over fine structure with significant incremental-F tests for both envelope and fine structure. The multiple regression model explained about 60% of the variance in the $P(Y|W)$ residuals for this subject.

Subjects S1 and S4 were tested with the highest threshold tone levels. Subject S1 weighted cues derived from the fine structure more strongly than those derived from the envelope, but the predictions explained only 39% of the variance in the $P(Y|W)$ residuals. Subject S4 used cues derived from both envelope and fine structure, and the multiple regression model was able to explain up to 68% of the variance in the $P(Y|W)$ residuals.

In general, the results for the N_0S_π condition, as for the N_0S_0 condition, were not well described by the statistical model based on a linear combination of envelope and fine-structure information. The model seemed to fit best for the subjects with higher thresholds but, in general, predicted about 40%–70% of the variance in the base line detection pattern residuals, substantially lower than the estimates of predictable variance (r^2_{max}) for the N_0S_π condition, which were 84% or higher.

C. Comparisons between subjects

To compare detection patterns across subjects, the square of Pearson's correlation coefficient (r^2) was computed. Table V shows between-subject r^2 values for the base line and chimeric detection patterns. The between-subject r^2 values were lower for the N_0S_0 condition in this study than in previous studies (Evilsizer *et al.*, 2002; Davidson *et al.*, 2006) and ranged from 0.34 to 0.69 for $P(Y|W)$. The lower between-subject correlations suggest the use of a more diverse set of decision variables across subjects in this experiment than in previous experiments with diotic stimuli, which was likely caused by the lack of a simple energy cue to rely on and by the narrow stimulus bandwidth. Pairs of subjects with the highest between-subject r^2 values did not necessarily share envelope or fine-structure dominance (e.g., S2 and S6 in Fig. 3 and Table V).

Under the N_0S_π condition, between-subject r^2 values were on average lower than those for N_0S_0 condition and ranged from 0.00 to 0.63 for $P(Y|W)$ (Table V). Subject pairs with similar threshold tone levels had more similar detection patterns than those with differing threshold tone levels. Subject pair S1 and S4 had the highest intersubject correlations, and these subjects also had similar and relatively high thresholds, suggesting a dependence of threshold on detection strategy. However, comparing Fig. 4 to Table V for the pairs of subjects with the largest intersubject correlations (and the most similar thresholds) did not reveal any clear

TABLE V. Comparisons between subjects presented in terms of r^2. Correlations were based on responses to all four stimulus sets.

	N_0S_0														
	P(Y|W)					P(Y|T+N)					P(Y|N)				
	S2	S3	S4	S5	S6	S2	S3	S4	S5	S6	S2	S3	S4	S5	S6
S1	0.45[a]	0.52[a]	0.63[a]	0.34[a]	0.50[a]	0.28[a]	0.22[a]	0.36[a]	0.10[a]	0.27[a]	0.36[a]	0.49[a]	0.66[a]	0.34[a]	0.44[a]
S2		0.48[a]	0.57[a]	0.62[a]	0.69[a]		0.27[a]	0.30[a]	0.52[a]	0.60[a]		0.35[a]	0.56[a]	0.53[a]	0.56[a]
S3			0.49[a]	0.40[a]	0.51[a]			0.11[a]	0.21[a]	0.28[a]			0.36[a]	0.28[a]	0.31[a]
S4				0.43[a]	0.65[a]				0.15[a]	0.46[a]				0.44[a]	0.54[a]
S5					0.68[a]					0.44[a]					0.75[a]
	N_0S_π														
	P(Y|W)					P(Y|T+N)					P(Y|N)				
	S2	S3	S4	S5	S6	S2	S3	S4	S5	S6	S2	S3	S4	S5	S6
S1	0.00	0.14[a]	0.63[a]	0.05[a]	0.14[a]	0.13[a]	0.01	0.43[a]	0.01	0.00	0.37[a]	0.00	0.63[a]	0.02	0.05[a]
S2		0.40[a]	0.00	0.54[a]	0.39[a]		0.19[a]	0.18[a]	0.46[a]	0.24[a]		0.00	0.36[a]	0.10[a]	0.00
S3			0.19[a]	0.45[a]	0.50[a]			0.01	0.20[a]	0.24[a]			0.02	0.05[a]	0.05[a]
S4				0.06[a]	0.13[a]				0.02	0.03				0.09[a]	0.01
S5					0.36[a]					0.20[a]					0.00

[a] $p < 0.05$.

pattern of envelope or fine-structure dominance. Evidence for individual differences between subjects in the use of available cues for binaural detection has been described previously (e.g., McFadden *et al.*, 1971).

D. Comparisons between interaural conditions

Table VI shows correlations (in terms of r^2) between N_0S_0 and N_0S_π interaural conditions for P(Y|W) and for P(Y|T+N) and P(Y|N). The subjects with the highest thresholds (S1 and S4) had the highest correlations between detection patterns from the two interaural conditions. Closer inspection of Table VI reveals that the sources of the correlations between the two interaural conditions for these subjects were largely from responses to noise-alone stimuli, P(Y|N). Subjects S1 and S4 show substantial r^2 values (0.93 for both listeners) between P(Y|N) values from the two interaural conditions. Recall that N stimuli in the N_0S_π condition were identical to those for the N_0S_0 condition (but not the T+N stimuli). Such high r^2 values suggest that S1 and S4 may have been attempting to use the same detection strategy for the two interaural conditions; if this strategy were more appropriate for N_0S_0 listening, that would explain their high thresholds for the N_0S_π stimuli. The fact that these subjects still had substantially lower dichotic thresholds suggests that they were not employing a strictly diotic strategy but may have instead been monitoring more than one cue for detection (e.g., one diotic cue and one dichotic cue). Because the dichotic cue would not be present on N trials, responses would be similar to those under the N_0S_0 condition. On T+N trials, both cues might be present, leading to the more modest correlations observed. The subjects with the lowest thresholds (S3 and S5) and intermediate thresholds (S2 and S6) had much lower correlations between detection patterns from the two interaural conditions. The noise-alone intersubject r^2 values have implications for the types of detection models used to explain the detection patterns, as outlined in Sec. IV.

TABLE VI. Comparisons between interaural conditions presented in terms of r^2. Correlations were based on responses for all four stimulus sets.

Subject	P(Y|W)	P(Y|T+N)	P(Y|N)
S1	0.60[a]	0.19[a]	0.93[a]
S2	0.02[a]	0.00	0.46[a]
S3	0.30[a]	0.00	0.02
S4	0.65[a]	0.15[a]	0.93[a]
S5	0.02[a]	0.00	0.33[a]
S6	0.13[a]	0.01	0.00

[a] $p < 0.05$.

IV. DISCUSSION

A. Summary of results

As discussed in Sec. I, there is an extensive literature indicating that monaural, diotic, and dichotic tone-in-noise detection can be partially, but not completely, predicted based on across-wave-form variations in energy. In order to better understand other cues that might also contribute to detection performance, this experiment investigated the roles of stimulus envelope and fine structure when energy differences among stimuli were eliminated. A simple multiple regression statistical model was unable to explain all of the predictable variance in the detection pattern residuals, yielding observed R^2 values between 0.32 and 0.75 for N_0S_0 stimuli and between 0.39 and 0.68 for N_0S_π stimuli. The predictable variance was estimated by r^2_{max},[4] which ranged from 0.77 to 0.99 for both N_0S_0 and N_0S_π stimuli and was always substantially higher than the observed R^2.

The envelope and fine-structure provide a complete description of the energy-normalized stimuli used in this study. The fact that envelope and fine structure were unable to predict the subjects' responses separately, or when combined

linearly, suggests that this description is somehow inadequate or that this method of decomposing the stimuli is not consistent with underlying physiological and psychophysical processes. The implications of these results for future efforts to model detection are discussed below. Successful models must consider alternative descriptions of the stimuli and/or must capture the impact of temporal interactions between the envelope and the fine structure of the stimuli.

B. Implications for computational models of detection

1. Comparisons between detection patterns estimated with base line and chimeric stimuli

With randomly generated tone-in-noise stimuli, many putative detection cues tend to co-vary. The approach in this experiment was to eliminate overall energy differences among the stimuli for both N and T+N wave forms under the N_0S_0 condition and for only N wave forms under the N_0S_π condition and thereby to force the subjects to base their detection judgments on envelope cues and/or fine-structure cues. The assumption was that these two sets of temporal cues were separable and independent. The fact that a substantial portion of the variance in the responses to the base line stimulus sets was not predicted by the multiple regression statistical model suggests that this assumption is wrong, at least in part. Two broad possibilities are suggested: (1) Although the fine structure and the envelope are *obvious* visual features of the narrow-band noise wave form, and there is considerable physiological evidence that both fine-structure and envelope information are encoded at least somewhat independently in the firing patterns of auditory neurons (e.g., Joris and Yin, 1992), these cues may not be used to determine the presence or absence of the tone in the authors' narrow-band detection task. Perhaps the subjects base their decisions on some other representation (e.g., spectral shape). (2) Short-time interactions between the wave form envelopes and the wave form fine structures are critical to understanding these data. That is, envelope and fine-structure cues may be used to detect the tone, but they are not independent. If so, for example, it is unwise for models to independently extract envelope and fine structure, unless some interaction between the two occurs before the decision variable is computed. Indeed, temporal interactions between envelope and fine structure occur in narrow-band Gaussian noise [Davenport and Root, 1958, pp. 159–160 (e.g., rapid changes in instantaneous frequency or phase often occur when the instantaneous amplitude is low)]. Moreover, the wave forms used in the present study may have underrepresented this problem because the stimulus generation algorithm tended to reject wave form pairs for which the chimeric recombination would temporally align high envelope and rapid frequency fluctuations.

Some previous studies attempting to explain detection patterns for narrow-band stimuli with computational models have omitted peripheral filtering and nonlinearities under the assumption that these do not contribute to the detection process (e.g., Isabelle, 1995; Davidson *et al.*, 2006). However, this type of processing may be critical to capture the interactions between envelope and fine structure that are suggested by the present data.

Several candidate models remain in contention for both diotic and dichotic signal detection, and each will be tested in detail in further studies of these experimental results. These models are worth briefly mentioning here. In general, these models either operate on a spectral representation of the stimulus or incorporate some sort of dynamic interaction of envelope and fine structure; that is, each computes the decision variable from the entire stimulus (rather than stripping the stimulus envelope or fine structure apart for separate analyses). An example of a diotic model that remains under consideration is the multiple-detector model (e.g., Ahumada and Lovell, 1971; Gilkey and Robinson, 1986), which uses monaural banks of filters that are weighted and combined linearly to produce a decision variable. With respect to binaural models, equalization-cancellation-style models with realistic peripheral processing stages (e.g., Breebaart *et al.*, 2001) should remain under consideration. Cross-correlation-style models (e.g., Colburn, 1977) with realistic peripheral processing should also remain under consideration, given that these models operate on the entire stimulus wave form rather than on envelope or fine structure alone.

2. N_0S_π noise-alone data

This study, like numerous previous studies (e.g., Evilsizer *et al.*, 2002; Gilkey *et al.*, 1985; Isabelle, 1995; Siegel and Colburn, 1989), found reliable detection patterns (with significant across-wave-form variation in the probability of a "yes" response) under the N_0S_π condition for noise-alone stimuli. In the present study, these reliable differences in responding were found even though across-wave-form energy variations were eliminated from the noise-alone stimuli. This finding is particularly significant for modeling efforts because several hypothesized models (cf., Isabelle, 1995; Isabelle and Colburn, 2004; Goupell and Hartmann, 2007) rely only on interaural differences to compute decision variables and thus would require some sort of internal noise mechanism to generate decision variables for the diotic noise-alone stimuli. If independent internal noise processes dominated over external noise at each ear, any left-right-symmetric binaural processing would not result in a stable detection pattern for noise-alone stimuli (assuming *additive* internal noise). If the noise were additive, the response on each trial would simply be based on interaural differences that resulted from the internal noise processes. Over large numbers of trials, such noise-generated interaural differences would produce "flat" detection patterns with no reliable differences in detection probabilities from noise to noise. A multiplicative internal noise source may be able to produce a stable pattern, but a generally applicable model of such processing is not available [see Colburn *et al.* (1997) for a description of how multiplicative internal noise could lead to internal interaural differences that are dependent on the external diotic stimulus in the context of the equalization-cancellation model of Durlach (1963)].

Another mechanism that would generate reliable detection patterns for noise-alone stimuli is a static frequency mis-

TABLE VII. Means and standard deviations for correlations (r^2) between corresponding wave forms in the E_1F_1 and E_2F_2 stimulus sets for both T+N and N stimuli. The levels of the tone were set to each subject's threshold (given as E_S/N_0 in dB) for the N_0S_0 condition. The mean and standard deviations are also given for correlations between 10 000 pairs of randomly generated T+N and correlations between 10 000 pairs of randomly generated N wave forms ("Random"). Statistics are also given for the envelopes and fine structures of each wave form (see text for details).

			Entire wave form		Envelopes only		Fine structure only	
Wave form	Subject	E_s/N_0	Mean r^2	SD r^2	Mean r^2	SD r^2	Mean r^2	SD r^2
T+N	S1	10	0.44	0.18	0.31	0.23	0.47	0.19
	S2	10	0.44	0.18	0.31	0.23	0.47	0.19
	S3	10	0.44	0.18	0.31	0.23	0.47	0.19
	S4	11	0.52	0.16	0.33	0.24	0.56	0.17
	S5	11	0.52	0.16	0.33	0.24	0.56	0.17
	S6	11.5	0.55	0.15	0.34	0.24	0.60	0.16
	Random	10	0.45	0.12	0.24	0.23	0.49	0.22
	Random	11	0.50	0.18	0.26	0.24	0.56	0.21
	Random	11.5	0.54	0.18	0.27	0.24	0.60	0.20
N	All	···	0.14	0.18	0.35	0.28	0.12	0.15
	Random	···	0.11	0.13	0.16	0.12	0.08	0.11

match (e.g., van der Heijden and Trahiotis, 1998), or a static internal interaural delay or internal interaural attenuation. These mechanisms would be stable over time and would generate specific detection patterns based on the processing asymmetry. The magnitudes and types of plausible processing asymmetries will be examined in future work.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. F31 077798 (S.A.D.), NIDCD Grant No. 001641 (L.H.C. and S.A.D.), NIH Grant No. DC00100 (H.S.C.). We acknowledge Dr. Marty Sliwinski for statistical advice and Susan Early for editorial assistance.

APPENDIX: STIMULUS SELECTION PROCEDURE AND RELATED ISSUES

Because it is impossible to modify the temporal structure of the stimulus without also impacting the spectrum (e.g., Davenport and Root, 1958, pp. 159–160), the process of assembling chimeric stimuli in some cases resulted in wave forms with spectral splatter and associated temporal distortions. These distortions could interfere with the task and cause unintended interaural differences (in the N_0S_π condition), so the stimulus wave forms were examined for excessive spectral splatter and eliminated based on specific criteria. Note that if absolutely no spectral splatter were allowed, the stimulus-creation algorithm would have eventually created four sets of stimuli with identical corresponding wave forms. The spectrum of each chimeric stimulus (after applying the $\cos^2$ ramps) was checked to ensure that the magnitude of each spectral component more than 50 Hz away from the target was at least 15 dB below the wave form's spectral peak, and that each spectral component more than 90 Hz away from the target was at least 25 dB below the wave form's spectral peak. When a wave form failed this test, it was eliminated, and the corresponding wave forms (N, T_0+N for N_0S_0, T_0+N for N_0S_π, and T_π+N for N_0S_π) in all four sets of stimuli were also eliminated for all six subjects.

Stimuli that were eliminated tended to have large frequency modulations in the fine structure that were temporally positioned at relatively high envelope values when recombined (see Amenta *et al.*, 1987). Such a combination naturally increased the bandwidth of the wave form. When stimuli were eliminated, two new base line wave forms were created using random noise, and corresponding chimeric stimuli were created. The stimuli were scaled, tones were added, and the resulting wave forms were tested. The algorithm ran for approximately 12 h on a Pentium M computer (1.86 GHz) and eliminated thousands of candidate stimulus wave forms before obtaining the set used in the present study (the exact number of eliminated wave forms was not recorded). This process resulted in stimuli that had little distortions or spectral splatter.

One initial concern with limiting the amount of spectral splatter was that corresponding E_1F_1 and E_2F_2 wave forms could be too highly correlated (e.g., the only stimuli that could swap envelopes without generating any splatter would be identical stimuli), such that the detection patterns resulting from the E_1F_1 and E_2F_2 stimulus sets would be the same. Table VII shows the mean r^2 values for correlations between corresponding wave forms in the E_1F_1 and E_2F_2 stimulus sets. For comparison, Table VII also shows the mean of 10 000 r^2 values computed between pairs of randomly generated tone-plus-noise (T+N) stimuli, as well as between pairs of random noise-alone (N) stimuli. These statistics are also presented for the envelopes (computed as the absolute value of the complex analytic wave form, which adds the Hilbert transform of the original wave form as the imaginary part to the original real wave form) and the fine structure (computed as the cosine of the angle of the complex analytic function) of each wave form. Several Mann–Whitney U tests were performed comparing the r^2 values from the random

TABLE VIII. Comparisons between detection patterns estimated with wave forms from the E_1F_1 and E_2F_2 stimulus sets. Presented in terms of r^2.

Interaural condition	Subject	P(Y|W), r^2	P(Y|T+N), r^2	P(Y|N), r^2
N_0S_0	S1	0.13[a]	0.01	0.05
	S2	0.14[b]	0.01	0.02
	S3	0.15[b]	0.22[a]	0.01
	S4	0.16[b]	0.00	0.03
	S5	0.05	0.03	0.10
	S6	0.18[b]	0.04	0.16[a]
N_0S_π	S1	0.26[b]	0.01	0.06
	S2	0.12[a]	0.00	0.00
	S3	0.24[b]	0.00	0.15
	S4	0.26[b]	0.03	0.07
	S5	0.22[b]	0.01	0.05
	S6	0.24[b]	0.03	0.02

[a]$p<0.05$.
[b]$p<0.01$.

stimuli to those of the reproducible stimuli. For N stimuli, the envelope correlation was significantly ($p<0.01$) larger for the reproducible stimuli with respect to the random stimuli, but the whole-wave-form and fine-structure-only correlations were not significantly different ($p>0.5$). For T+N stimuli, subjects S1–S5 showed significantly ($p<0.01$) larger correlations between the envelopes of the reproducible stimuli for corresponding wave forms in the E_1F_1 and E_2F_2 stimulus sets than those present for correlations between the envelopes of random stimuli ($p>0.05$). Whole-wave-form and fine-structure correlations were not significantly larger when computed between the E_1F_1 and E_2F_2 stimuli than when computed between randomly generated T+N stimuli.

The similarity of the resulting E_1F_1 and E_2F_2 detection patterns (i.e., the patterns of hit and false-alarm rates for a particular group of the reproducible wave forms) was also examined. If the detection patterns for the responses to E_1F_1 and E_2F_2 had been highly correlated, results from the regression analysis (Sec. II D) would be questionable because of the perceptual similarity of the two groups (and likely similarity of the E_2F_1 and E_1F_2 wave forms). Table VIII shows correlations in terms of r^2 [the squared Pearson correlation coefficient of the z-scores (but not residuals) computed from the two detection patterns] between the E_1F_1 and E_2F_2 detection patterns for the subjects in this study. If both T+N and N responses are considered together [i.e., the detection pattern for the probability of saying "yes, the tone is present" across all reproducible wave forms, P(Y|W)], detection patterns from the E_1F_1 and E_2F_2 stimulus sets were significantly, albeit weakly, correlated (as one would expect because on average $P(Y|T+N)>P(Y|N)$, introducing correlation between any two sets of stimuli for which $d'>0$). However, for N_0S_0 stimuli the detection patterns for the E_1F_1 and E_2F_2 stimulus sets that were based on only T+N trials, P(Y|T+N), were not significantly correlated for any subjects except S2, and the detection patterns based only on N trials, P(Y|N), were not significantly correlated except for S6. Thus, the similarity of the P(Y|T+N) and P(Y|N) detection patterns for the E_1F_1 and E_2F_2 stimulus sets was not of concern. Moreover, the correlations that did exist were removed by the statistical blocking procedure described in Sec. II D).

Blauert (1981), Ghitza (2001), and Zeng *et al.* (2004) pointed out that an envelope may be recovered when relatively broadband stimuli are filtered in the auditory periphery with a filter narrower than the stimulus bandwidth. This was not likely to occur given the approximately 75-Hz critical bandwidth at 500 Hz and the fact that a 50-Hz noise bandwidth was used. Nevertheless, stimuli were diagnostically tested for possible envelope recovery by filtering all stimuli with a 50-Hz bandwidth, fourth-order gammatone filter at center frequencies from 400–600 Hz in 1-Hz steps. Envelopes were then recovered from the stimuli by half-wave rectification and filtering with a first-order low-pass filter with a 50-Hz cutoff frequency. First, envelopes from the filtered chimeric stimulus sets (E_1F_2 and E_2F_1) were compared to the envelopes from the filtered original stimulus sets (E_1F_1 and E_2F_2, respectively). The correlation value did not fall below 0.977 in any situation (i.e., at any filter center frequency or for any wave form). Then, the correlations between the envelopes extracted in the E_1F_1 and E_2F_2 stimulus sets were eliminated from the correlations between the envelopes extracted in the E_1F_2 and E_2F_2 stimulus sets and also from the correlations between the envelopes extracted from the E_2F_1 and E_1F_1 stimulus sets, all after filtering. This comparison examined whether the envelope of the base line stimulus sets (e.g., E_1) was recovered from the fine structures of the chimeric stimulus sets (e.g., E_2F_1), thereby potentially increasing the correlation of the envelopes of the base line and chimeric stimulus sets with respect to the two base line stimulus sets. The correlations never differed by more than 0.05, indicating that recovery of envelope information from stimulus fine structure by peripheral filtering was unlikely at these stimulus and filter bandwidths.

[1]Each wave form set had 100 wave forms. There were 25 noise-alone (N) wave forms, which were presented diotically for the N trials of both the N_0S_0 and the N_0S_π conditions. There were 25 tone-plus-noise (T+N) wave forms, with the tone set to the N_0S_0 threshold SNR for that listener, which were presented diotically for the N_0S_0 T+N trials. There were 50 T+N wave forms, which were presented dichotically (i.e. 25 left/right stimulus pairs) for the N_0S_π condition; these were created by adding and subtracting tones, with the tone level set to the N_0S_π threshold SNR for that listener.

[2]To maintain compatibility with previous work (and with much of the tone-in-noise masking literature to which these findings may generalize), the threshold level of the signal is presented in terms of $10\log_{10}(E_S/N_0)$, the ratio of the energy in the signal to the spectrum level of the noise in decibels. However, because of the unusual nature of these stimuli, some additional explanation of these values is needed. Although the masker, if extended to a steady state wave form by periodic extension, could be viewed as a collection of five, 50-dB tones, the 0.1-s duration wave form that is presented to the listeners is a narrow-band wave form with a continuous spectrum. Specifically, the Fourier transform has a bandwidth, BW, of approximately 50-Hz, so that the overall noise power P_N of 57 dB SPL results in an approximate power spectrum level of 40 dB SPL [57 dB SPL$-10\log_{10}$(BW)]. [Note that the exact bandwidth is not critical: Assuming a 40-Hz BW, i.e., 520–480 Hz, would decrease the computed values of $10\log_{10}(E_S/N_0)$ by about 1 dB]. The energy in the signal, E_S, is determined by $10\log_{10}(E_S)=10\log_{10}(P_ST)$, where P_S is the signal power at threshold and T is the duration of the signal in seconds). So, one can compute the signal to noise power ratio from $10\log_{10}(E_S/N_0)$ as $10\log_{10}(P_S/P_N)=10\log_{10}(E_S/N_0)-10\log_{10}(T\cdot BW)=10\log_{10}(E_S/N_0)-7$ dB. It is also possible to compute P_S (in dB SPL) as $10\log_{10}(P_S)$

$=10\log_{10}(E_S/N_0)+10\log_{10}(P_N/(T\cdot BW))=10\log_{10}(E_S/N_0)+50$ dB SPL. However, note that the T+N wave forms for the N_0S_0 condition are rescaled after the tone is added, so this value is only approximate.

[3]Probabilities of 0 and 1 (the z-scores of which are unbounded) were replaced with 1/100 and 99/100 (z-scores of -2.33 and +2.33), respectively; this occurred for only 54 of the 2400 probabilities in the stimulus set [50 $P(Y|W)\times 2$ interaural conditions $\times 4$ stimulus conditions $\times 6$ subjects].

[4]The relation between r_{12}, the split-half correlation, and r_{max}, the correlation between conditions a and b that are assumed to be identical conditions except for random variability (e.g., for comparisons across stimulus sets, subjects, or interaural conditions), is derived here. First, consider two random variables, x_a and x_b, that share a common predictable component, with variance σ_p^2, and two additive independent and unpredictable components, with variances σ_{Ea}^2 and σ_{Eb}^2. The correlation between these two random variables is described by $r_{max}=(\sigma_p^2/\sqrt{\sigma_p^2+\sigma_{Ea}^2}\sqrt{\sigma_p^2+\sigma_{Eb}^2})$ (Robinson and Jeffress, 1963). If it is assumed that $\sigma_p^2=1$ and that $\sigma_{Ea}^2=\sigma_{Eb}^2=\sigma_E^2$, then $r_{max}=[1/(1+\sigma_E^2)]$. Rearranging this equation provides an expression for the unpredictable variance $\sigma_E^2=[(1-r_{max})/r_{max}]$. This variance can also be approximated using the split-half correlation, r_{12}, as $\sigma_E^2=[(1-r_{12})/2r_{12}]$, where the factor of 2 in the denominator was introduced to account for the fact that in the case presented here the predictable variance for comparison of two 100-trial data sets was half that estimated from the split-half correlation (i.e., r_{12} was computed from a single set of 100 trials divided into two sets of 50). Thus, combining the expressions for r_{max} and σ_E^2 provides an expression for r_{max} in terms of r_{12}, which was estimated from the data. Finally, the proportion of predictable variance is provided by squaring r_{max}, $r_{max}^2=[1/(1+\sigma_E^2)]^2=(1/\{1+[(1-r_{12})/2r_{12}]\})^2$.

[5]Before employing these techniques, several tests were applied to the detection patterns (z-scores) to determine whether the data were consistent with the assumptions of the analysis procedure. First, the detection patterns were checked for normality using the Lilliefors hypothesis test of composite normality (Sheskin, 2000), keeping the individual-test alpha level at 0.05. No family-wise error-rate correction was implemented in order to maintain a conservative test criterion. Only two of the 144 detection patterns [$P(Y|W)$, $P(Y|T+N)$, and $P(Y|N)$ for 4 stimulus sets $\times 2$ interaural conditions $\times 6$ subjects] proved to be non-normal. Second, for all regression analyses, residuals were examined using the same test. Of the 324 regressions performed [3 predictors (E_1F_1, E_2F_2, and combined) $\times 3$ detection-pattern components $\times 6$ subjects $\times 3$ predictor models (envelope, fine structure, or both) $\times 2$ interaural conditions], only 10 showed significantly ($p<0.05$) non-normal residuals. Finally, examination of residual plots failed to find any serious issues of heteroscedasticity (unequal error variances). Correlations between predictor variables in the same analysis were computed to check for multicolinearity (high correlation of predictor variables). Typical values for the r^2 between predictor variables ranged from 0 to 0.1 (and were insignificant) and in no case exceeded 0.31, indicating that the data did not exhibit a large degree of multicolinearity. Overall, the results of these tests suggest that these assumptions were adequately satisfied for the tests to be meaningful.

[6]The test of significant differences between correlated but non-overlapping correlations (Raghunathan *et al.*, 1996) was conducted for each combination of envelope and fine structure (2), for each subject (6), and for $P(Y|W)$, $P(Y|T+N)$, and $P(Y|N)$, for a total of 36 tests. The question was whether or not the null hypothesis could be rejected, where *not* rejecting the null hypothesis was the desired outcome. Therefore, to produce a more conservative rejection criterion and to reduce the chance of a type-II error, a family-wise error alpha level was not computed, and the individual alpha level for each test was maintained at 0.05. None of the 36 tests under the N_0S_0 condition nor the 36 tests under the N_0S_π yielded significant differences ($p<0.05$) between predictions for E_1F_1 and E_2F_2 for either envelope or fine structure.

Ahumada, A., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Amenta, C. A., Trahiotis, C., Bernstein, L. R., and Nuetzel, J. M. (**1987**). "Some physical and psychological effects produced by selective delays of the envelope of narrow bands of noise," Hear. Res. **29**, 147–161.

Bernstein, L. R., and Trahiotis, C. (**1985**). "Lateralization of low-frequency, complex waveforms: The use of envelope-based temporal disparities," J. Acoust. Soc. Am. **77**, 1868–1880.

Bernstein, L. R., and Trahiotis, C. (**1996**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3784.

Bernstein, L. R., and Trahiotis, C. (**2002**). "Enhancing sensitivity to interaural delays at high frequencies by using 'transposed stimuli'," J. Acoust. Soc. Am. **112**, 1026–1036.

Blauert, J. (**1981**). "Lateralization of jittered tones," J. Acoust. Soc. Am. **70**, 694–698.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**1999**). "The contribution of static and dynamically varying ITDs and IIDs to binaural detection," J. Acoust. Soc. Am. **106**, 979–992.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

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.

Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S., Isabelle, S. K., and Tollin, D. J. (**1997**). "Modeling binaural detection performance for individual masker waveforms," *Binaural and Spatial Hearing* (Erlbaum, NJ).

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.

Davenport, W. B., and Root, W. L. (**1958**). *An Introduction to the Theory of Random Signals and Noise* (McGraw-Hill, New York), p. 393.

Davidson, S. A. (**2007**). "Detection of tones in reproducible noise: Psychophysical and computational studies of stimulus features and processing mechanisms," Ph.D. dissertation, Syracuse University (www.bme.rochester.edu/carney.html).

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (**2006**). "Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results," J. Acoust. Soc. Am. **119**, 2258–2275.

Durlach, N. I. (**1963**). "Equalization and cancellation theory of binaural masking-level differences," J. Acoust. Soc. Am. **35**, 1206–1218.

Eddins, D. A., and Barber, L. E. (**1998**). "The influence of stimulus envelope and fine structure on the binaural masking level difference," J. Acoust. Soc. Am. **103**, 2578–2589.

Edwards, A. L. (**1979**). *Multiple Regression and the Analysis of Covariance* (Freeman, New York).

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Gilkey, R. H., and Robinson, D. E. (**1986**). "Models of auditory masking: A molecular psychophysical approach," J. Acoust. Soc. Am. **79**, 1499–1510.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Goupell, M. J., and Hartmann, W. M. (**2007**). "Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models," J. Acoust. Soc. Am. **122**, 1029–1045.

Hafter, E. R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116–1122.

Hall, J. W., Haggard, M. P., and Fernandes, M. A. (**1984**). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.

Isabelle, S. K. (**1995**). "Binaural detection performance using reproducible stimuli," Ph.D. dissertation, Boston University, Boston, MA.

Isabelle, S. K., and Colburn, H. S. (**2004**). "Binaural detection of tones masked by reproducible noise: Experiment and models," Report No. BU-HRC 04-01, Boston University, Boston, MA.

Joris, P. X. (**2003**). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," J. Neurosci. **23**, 6345–6350.

Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated sounds in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.

Kay, R. H. (**1982**). "Hearing of modulation in sounds," Physiol. Rev. **62**,

894–975.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).
Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Puschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acust. **83**, 659–669.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press., New York).
McFadden, D., Jeffress, L. A., and Ermey, H. L. (**1971**). "Differences of interaural phase and level in detection and lateralization: 250 Hz," J. Acoust. Soc. Am. **50**, 1484–1493.
Moore, B. J. C. (**1975**). "Mechanisms of masking," J. Acoust. Soc. Am. **57**, 391–399.
Neff, D. L., and Callaghan, B. P. (**1988**). "Effective properties of multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.
Osman, E. (**1973**). "A correlation model of binaural masking level differences," J. Acoust. Soc. Am. **50**, 1494–1511.
Raghunathan, T. E., Rosenthall, R., and Rubin, D. B. (**1996**). "Comparing correlated but nonoverlapping correlations," Psychol. Methods **1**, 178–183.
Richards, V. M. (**1992**). "The detectability of a tone added to narrow bands of equal energy noise," J. Acoust. Soc. Am. **91**, 3424–3425.
Robinson, D. E., and Jeffress, L. A. (**1963**). "Effect of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**, 1947–1952.
Sheskin, D. J. (**2000**). *Parametric and Nonparametic Statistical Procedures* (Chapman and Hall, New York).
Siegel, R. A., and Colburn, H. S. (**1989**). "Binaural processing of noisy stimuli: Internal/external noise ratios under diotic and dichotic stimulus conditions," J. Acoust. Soc. Am. **86**, 2122–2128.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
van de Par, S., and Kohlrausch, A. (**1997**). "A new approach to comparing binaural masking level differences at low and high frequencies," J. Acoust. Soc. Am. **101**, 1671–1680.
van de Par, S., and Kohlrausch, A. (**1998**). "Diotic and dichotic detection using multiplied-noise maskers," J. Acoust. Soc. Am. **103**, 2100–2110.
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.
Zeng, F., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

An evaluation of models for diotic and dichotic detection in reproducible noises

Sean A. Davidson
Department of Biomedical and Chemical Engineering, Institute for Sensory Research, Syracuse University, 621 Skytop Road, Syracuse, New York 13244

Robert H. Gilkey
Department of Psychology, Wright State University, Dayton, Ohio 45435 and Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433

H. Steven Colburn
Department of Biomedical Engineering, Boston University Hearing Research Center, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Department of Biomedical and Chemical Engineering, Department of Electrical Engineering and Computer Science, and The Institute for Sensory Research, Syracuse University, 621 Skytop Road, Syracuse, New York 13244

(Received 22 August 2008; revised 29 May 2009; accepted 27 July 2009)

Several psychophysical models for masked detection were evaluated using reproducible noises. The data were hit and false-alarm rates from three psychophysical studies of detection of 500-Hz tones in reproducible noise under diotic (N_0S_0) and dichotic (N_0S_π) conditions with four stimulus bandwidths (50, 100, 115, and 2900 Hz). Diotic data were best predicted by an energy-based multiple-detector model that linearly combined stimulus energies at the outputs of several critical-band filters. The tone-plus-noise trials in the dichotic data were best predicted by models that linearly combined either the average values or the standard deviations of interaural time and level differences; however, these models offered no predictions for noise-alone responses. The decision variables of more complicated temporal models, including the models of Dau *et al.* [(1996a). J. Acoust. Soc. Am. **99**, 3615–3622] and Breebaart *et al.* [(2001a). J. Acoust. Soc. Am. **110**, 1074–1088], were weakly correlated with subjects' responses. Comparisons of the dependencies of each model on envelope and fine-structure cues to those in the data suggested that dependence upon both envelope and fine structure, as well as an interaction between them, is required to predict the detection results. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3206583]

PACS number(s): 43.66.Dc, 43.66.Ba, 43.66.Pn [RYL] Pages: 1906–1925

I. INTRODUCTION

The traditional goal of psychophysical experiments examining masked detection has been to characterize threshold signal-to-noise ratios (SNRs) as functions of physical parameters of the stimuli (e.g., signal frequency, noise bandwidth, and interaural phase difference of the signal). These threshold SNRs have been estimated using masker waveforms drawn on each trial without replacement from an effectively infinite set, such that no sample of masking noise is ever presented more than once. More recently, a number of studies have collected data using reproducible maskers (e.g., Pfafflin and Matthews, 1966; Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey *et al.*, 1985; Siegel and Colburn, 1989; Isabelle and Colburn, 1991; Isabelle, 1995; Isabelle and Colburn, 2004; Evilsizer *et al.*, 2002; Davidson *et al.*, 2006), allowing each sample of masking noise to be presented numerous times. These studies characterize detection responses for each individual stimulus waveform in a set of masking noise samples, rather than describing a single threshold estimated using maskers from an infinite set of noise waveforms. Such data present a more rigorous test for models of masked detection because, in addition to predicting average threshold, the models must predict detection statistics for individual waveforms. As shown here and in other works, models that accurately predict average thresholds may fail to predict responses to individual waveforms (e.g., Isabelle, 1995; Isabelle and Colburn, 2004).

The models tested in this study were selected because they have successfully predicted reproducible noise data in the past (e.g., Fletcher, 1940; Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986), because they have been used with some success to predict thresholds for a broad spectrum of psychophysical detection tasks (e.g., Dau *et al.*, 1996a, 1996b; Breebaart *et al.*, 2001a, 2001b, 2001c), because they are straightforward adaptations of ob-

[a)]Author to whom correspondence should be addressed. Also at Department of Neurobiology and Anatomy, University of Rochester, Box 603, 601 Elmwood Ave., Rochester, NY 14642. Electronic mail: laurel.carney@rochester.edu

served physiological phenomena (e.g., McAlpine *et al.*, 2001; Marquardt and McAlpine, 2001), or because they use a processing strategy that involves an interaction between stimulus envelope and fine structure (e.g., Goupell and Hartmann, 2007). The interaction of envelope and fine structure was of particular interest because such an interaction has been suggested by recent empirical studies of detection of low-frequency tones in reproducible maskers (Davidson, 2007; Davidson *et al.*, 2009).

II. METHODS

A. Target data

Data sets from three psychophysical studies (Isabelle, 1995; Evilsizer *et al.*, 2002; Isabelle and Colburn, 2004; Davidson, 2007; Davidson *et al.*, 2009) that shared similar experimental methods are modeled in this work. In these studies, an approximate threshold was estimated, drawing from an infinite set of masker waveforms without replacement, first in a two-interval up/down tracking experiment and subsequently verified in a single-interval experiment with fixed SNR. Then a fixed-SNR (i.e., at the estimated threshold), single-interval experiment was performed with a small (25–100) closed set of reproducible maskers. On each trial, the masker was randomly drawn, with replacement, from the closed set, with the constraint that all tone-plus-noise (T+N) and noise-alone (N) waveforms were presented an equal number of times during the course of the entire experiment (50–100 presentations, depending on the study). Upon completion of the experiment, the hit rate, or proportion of "yes" responses when the tone was present [$P(Y|T+N)$], and the false-alarm rate, or proportion of yes response when the tone was not present [$P(Y|N)$], were calculated separately for each individual masker waveform in the set. The resulting set of hit and false-alarm rates is termed the detection pattern. Note that although these hit and false-alarm rates could be used to calculate some performance metric [e.g., P(C), d', etc.] on a per noise waveform basis, that was not the focus here. Instead, N and T+N trials were considered separately, and the hit and false-alarm rates were used to estimate the tendency of the subject to respond "tone present" (presumably, based on how much a particular waveform "sounded like" it contained the tone).

Hit and false-alarm rates from Isabelle (1995) (*Study 1*), Evilsizer *et al.* (2002) (*Study 2*), and Davidson *et al.* (2009) (see also Davidson, 2007) (*Study 3*) served as the data for the modeling presented in this study. These data were selected because collectively, they established a set of detection patterns estimated under diotic (N_0S_0) and dichotic (N_0S_π) interaural configurations, with several noise bandwidths [50 (Study 3), 100 (Study 2), 115 (Study 1), and 2900 Hz (Study 2)], at a single tone frequency of 500 Hz. Study 1 examined the N_0S_π configuration only, whereas the other studies examined N_0S_0 and N_0S_π configurations and used the same noise masker samples under both conditions.

In contrast to Studies 1 and 2, in which the sets of maskers were randomly generated, Study 3 examined four stimulus sets under each interaural configuration. These stimulus sets were denoted E_1F_1, E_2F_2, E_1F_2, and E_2F_1, with E denoting envelope and F denoting fine structure. Corresponding stimuli (N or T+N) within the E_1F_1 and E_1F_2 stimulus sets and within the E_2F_1 and E_2F_2 stimulus sets shared the same temporal envelopes. Similarly, corresponding stimuli within the E_1F_1 and E_2F_1 stimulus sets and within the E_1F_2 and E_2F_2 stimulus sets shared the same fine structures (i.e., had the same zero crossings). The energies of T+N and N waveforms were equalized for all N_0S_0 stimuli in Study 3, thus eliminating detection cues related to overall energy. Davidson *et al.* (2009) and Davidson (2007) provided details regarding stimulus construction for Study 3.

B. General modeling strategy

The models were implemented without internal noise and without a decision stage. The output of each model (i.e., the decision variable) was calculated and compared to the responses of each subject on a waveform-by-waveform basis. To do this, the value of $P(Y|T+N)$ or $P(Y|N)$ obtained for each subject in response to each waveform was converted to a z-score using the inverse cumulative normal distribution function[1] as in Evilsizer *et al.* (2002). This conversion is equivalent to corrupting the model's decision variable (DV) with normally-distributed, additive internal noise. That is, it was assumed that the subject's DV was the sum of external and internal noise components, $DV=DV_{ext}+DV_{int}$. The external component, DV_{ext}, was computed in response to the external stimulus and was assumed to be fixed across trials on which the same stimulus waveform was presented, but to vary across stimulus waveforms. The internal component, DV_{int}, was assumed to be randomly drawn from a normal distribution with mean equal to zero and constant variance, independent of the trial or waveform presented. Under these assumptions, the z-score provides an estimate proportional to the distance from the subject's criterion to DV_{ext} for a particular waveform and subject. (Criterion variation, if present, is one form of internal noise and is not separately considered here.) Thus, the computed (noise-free) DV of a correct model should be linearly related to these z-scores (i.e., both the subject z-scores and the model DV should be linearly related to DV_{ext}). The proportion of variance accounted for by each model was simply computed as the square of the Pearson product-moment correlation (r^2) between the model DV and the subject z-score.[2] Because it was assumed *a priori* that different subjects might employ different detection strategies (indeed, this appears to have been the case, at least in Study 3), each model was compared to each subject's data individually (the analyses did not consider data that were averaged across subjects or attempt to predict the across-subject variance). Previous studies that modeled data that were averaged across subjects have been able to explain more of the variance in the data (e.g. Isabelle, 1995; Isabelle and Colburn, 2004; Davidson *et al.*, 2006); however, in those studies there were energy differences across waveforms that were presumably a common source of variation in the detection patterns that could be enhanced by averaging.

Implementing the models and representing the data in this way has some important implications. First, this approach allowed an evaluation of models without developing

a sophisticated internal noise model; however, many of the dichotic models considered require internal noise, or some other modification, to produce non-zero outputs on N trials. Therefore, those models were not applied on N trials. Second, in some cases it was advantageous to combine the outputs of two models in order to predict the subjects' responses. Because a linear relationship between the decision variables of the models and the *z*-scores of the subjects was expected, linear multiple regression techniques could be applied.

Although the subjects' thresholds were estimated using traditional psychophysical techniques before the experiment began, estimated thresholds were only used to set the SNR during the experiment. The analyses reported here *do not* attempt to predict thresholds (indeed, N and T+N trials were analyzed separately).[3] The focus in this study was on the responses of subjects to individual reproducible stimulus waveforms (i.e., how likely the subjects were to say the target was present when that stimulus waveform was presented). The SNR for the models and the SNR for the subject were identical since both responded to identical stimulus waveforms.

When evaluating the success of these models it is important to establish an upper limit of expected performance for any given prediction.[4] Isabelle (1995) (see Isabelle and Colburn, 2004) described that the reasonable upper limit for predicting their N_0S_π data (Study 1) was an r^2 of about 0.88. Evilsizer *et al.* (2002) (Study 2) reported first-half, last-half correlations that yield predictable variances (V_P) (Ahumada and Lovell, 1971) from 0.80 to 0.97. Predictable variances for Study 3 (Davidson, 2007; Davidson *et al.*, 2009) ranged from 0.85 to 0.99 for data averaged across the baseline stimulus sets in the N_0S_0 condition and from 0.92 to 0.99 for the N_0S_π data. Model results are presented here in terms of r^2; qualitatively similar results presented in terms of an estimate of the proportion of *predictable* variance explained, computed as the ratio of r^2 over V_P, are available in Davidson (2007).

III. MODEL DESCRIPTIONS, RESULTS, AND DISCUSSION

A brief description of each diotic and dichotic model is provided below, along with the results for those models. Detailed descriptions of model implementations are provided in Appendixes A through H. Table I provides a list of the models studied here and includes a brief description of their decision variables. In the following discussion, the effect of stimulus energy on each model's predictions is considered for both the diotic and dichotic stimulus conditions (i.e., to what extent is the performance of a model dependent on its correlation with energy). Finally, the extent to which each model relies on stimulus envelope or fine structure is compared to the empirical dependency observed for human subjects in Study 3.

A. Diotic models

There are six sets of models considered here for the diotic data. The first two are related: single critical-band (CB) models and multiple-detector (MD) models that linearly combine outputs of multiple critical bands. To date, detection patterns estimated under diotic conditions have been best predicted by a MD model (Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986; Davidson *et al.*, 2006). The MD model accounted for up to 90% of the variance in one subject's responses in Ahumada and Lovell (1971) and up to 72% of the variance in one subject's responses in Gilkey and Robinson (1986). Davidson *et al.* (2006) used the MD model to predict the data of Study 2 and found that the model accounted for 78%–90% of the variance in the average subject's responses, depending on bandwidth and interaural configuration [monaural (N_mS_m) or diotic]. The MD model's DV is the weighted sum of energies at the outputs of several auditory filters surrounding the tone frequency. Thus, the MD model is an extension of Fletcher's (1940) proposal that detection responses are determined by the energy at the output of a single CB centered at the tone frequency. Davidson *et al.* (2006) showed that the CB model predicted 64%–82% of the variance in their average subject's responses.

Two simple models that depend on temporal cues were also considered here: a modified version of the Richards (1992) envelope-slope (ES) model (Zhang, 2004) and the phase-opponency (PO) model (Carney *et al.*, 2002). Davidson *et al.* (2006) showed that the ES and PO models predicted about 60% of the variance in narrowband and wideband N_0S_0 and N_mS_m detection patterns. Neither model has previously been tested using detection patterns estimated from stimuli where energy was equalized across stimulus waveforms (as was the case in Study 3). Presumably, the likelihood that the subjects will use temporal cues, and thereby the success of these models, will increase when energy cues are not available.

Finally, two relatively more complex models were evaluated (Dau *et al.*, 1996a; Breebaart *et al.*, 2001a). These models combine temporal and energy information and also include a basic representation of both peripheral filtering and adaptation. Each of these models creates an internal-representation template through an iterative method, and decision variables for the detection task are derived based on comparisons to this template.

1. CB model

The DV for the CB model (Fig. 1) is the rms output of a fourth-order gamma-tone filter centered at 500 Hz. The equivalent rectangular bandwidth (ERB) of the filter was set at 75 Hz (Glasberg and Moore, 1990). The CB model was the simplest model tested in this study and, in general, was able to predict a significant and substantial proportion of the variance in the detection patterns for all subjects in Study 2 [Fig. 2(A)]. Recall that overall energies were equalized for all diotic stimuli in Study 3. The CB model made relatively poor predictions of the detection patterns in the equal-energy cases,[5] as expected. In Study 3, where energy cues were not available, the predictions of the CB model were significantly correlated to the detection patterns for only 12 out of 24

TABLE I. List of models tested in this study.

Model	Periphery	Decision variable
	N_0S_0	
CB[a]	GT(4)[b]	Energy
MD[c]	GT(4), MF[d]	Linear combination of energy across channels
MDS	GT(4), MF	Linear combination of energy across channels
ES[e]	GT(4), Extract Env[f]	Average slope of the envelope
DA[g]	GT(4), Adapt. Loops[h]	Similarity to peripherally transformed "noisy" tone template
BR[i]	GT(3), Adapt. Loops, MF	Difference from peripherally transformed noise-alone template
PO[j]	GT(4), AN model,[k] MF	Monaural cross-frequency coincidence detection
	N_0S_π	
EN[l]	None	Energy
sT[l]	None	Standard deviation (Std) of ITDs
Sl[l]	None	Std of ILDs
Wst[l]	None	Linear combination of Stds of ITD and ILD
Wav[m]	None	Linear combination of averages of ITD and ILD
Xst[m]	None	Std of linear combination of ITD and ILD
Xav[m]	None	Average of linear combination of ITD and ILD
Lp[l,n]	None	Std of linear combination of ITD and ILD
FCc[o]	GT(4), AN model	Linear combination of time-delayed binaural cross correlations and cancellations
FCn[o]	GT(4), AN model	Linear combination of time-delayed normalized binaural cross correlations and cancellations
BR[i]	GT(3), Adapt. Loops, MF	Difference from peripherally transformed noise-alone template

[a]Fletcher (1940).
[b]GT(N) indicates *N*th-order gammatone filter.
[c]Ahumada and Lovell (1971); Ahumada *et al.* (1975); Gilkey and Robinson (1986).
[d]MF indicates multiple frequency channels.
[e]Richards (1992); Zhang (2004).
[f]Envelope extraction performed using Hilbert transform.
[g]Dau *et al.* (1996a).
[h]Adaptation loops from Dau *et al.* (1996a).
[i]Breebaart *et al.* (2001a).
[j]Carney *et al.* (2002).
[k]Auditory-nerve model of Heinz *et al.* (2001).
[l]Isabelle (1995).
[m]Goupell (2005); Goupell and Hartmann (2007).
[n]Hafter (1971).
[o]Marquardt and McAlpine (2001).

cases for P(Y|T+N) and for only 18 out of 24 cases for P(Y|N). This finding is in agreement with the results of Richards (1992).

2. MD models

As stated above, the DV of the MD model (Gilkey and Robinson, 1986) was a weighted sum of energies at the outputs of several auditory filters surrounding the tone frequency (Fig. 1). The MD model (as described in the literature) uses a fit to the subjects' data, rather than a decision-theoretic weighting strategy. Here, we considered both this classic MD model and a multiple-detector model with suboptimal weights (MDS), which were computed using a decision-theoretic weighting scheme. This scheme is suboptimal in the sense that the chosen weights would maximize d' for the model, not the fit (r^2) between the detection patterns for the model and these specific data. Note that these models were only applied to Study 2, in which the stimulus bandwidth exceeded one critical band. Appendix A includes details of the MD and MDS implementations.

Plots showing the weights resulting from the fit to the data (MD model) and from the suboptimal computation (MDS model) provide insight into the differences between these models (Fig. 3). The negative weights for the MD model found above and below the target frequency were consistent with weighting patterns in previous MD model results (Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986), but only positive weights were possible in the suboptimal weighting scheme because these weights were derived from rms metrics. Note that the weights that fit to the responses of S3 for MD with the 100-Hz bandwidth were close to those fitted to the MDS model. Figure 2 shows that the proportions of variance accounted for by the MD and MDS models were also similar for that subject. The MDS model made poorer predictions for subjects that tended to have more negative MD weights, as expected. In fact, the MDS model predictions were even more poorly correlated to the data of S2 and S4 than the CB model predictions. For the MD model, significant and substantial predictions were made for all subjects' detection pat-

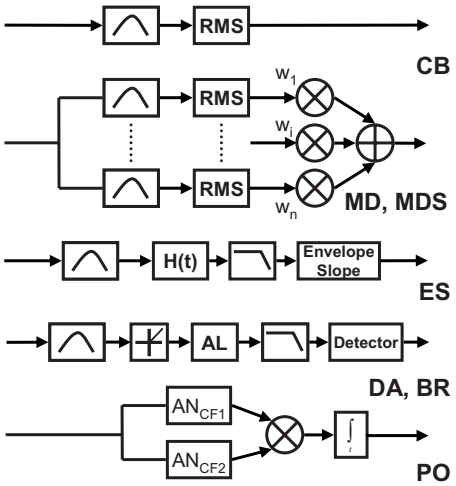


FIG. 1. Block diagrams of the models used to predict N_0S_0 detection patterns. The models listed from top to bottom are CB, critical band (Fletcher, 1940); MD, multiple detector with fit weights (Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986); MDS, multiple detector with suboptimal (independent channel) weights; ES, envelope slope (Richards, 1992; Zhang, 2004) DA, Dau model (Dau *et al.*, 1996a); BR, Breebaart model (Breebaart *et al.*, 2001a); PO, phase opponency (Carney *et al.*, 2002). $H(t)$ denotes the Hilbert transform used to recover the absolute value of the complex-analytic signal. AL denotes the adaptation loops as described in Dau *et al.* (1996a). AN denotes the auditory-nerve model of Heinz *et al.* (2001).

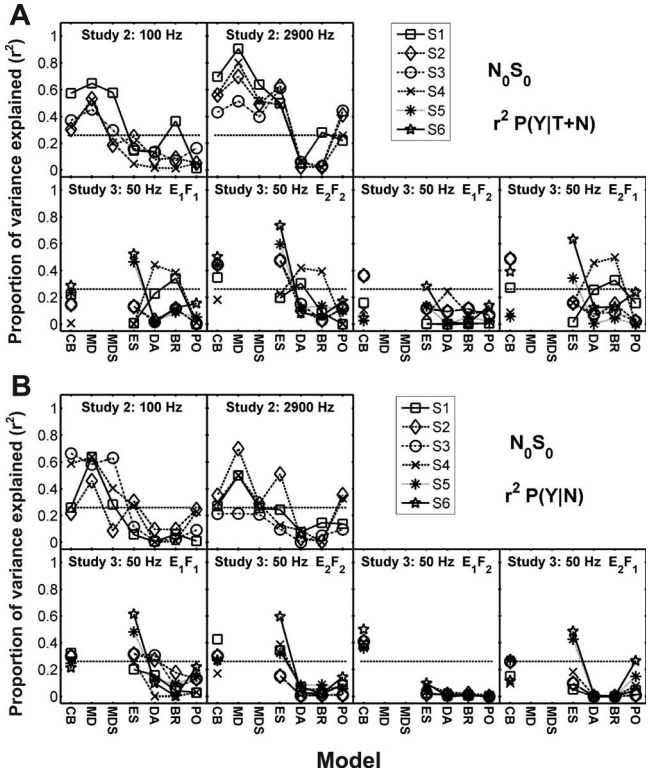


FIG. 2. Proportion of variance explained (r^2) in the z-scores of (A) $P(Y|T+N)$ and (B) $P(Y|N)$ by the N_0S_0 model predictions. Model abbreviations from left to right: CB, critical band; MD, multiple detector; MDS, multiple detector with suboptimal weights; ES envelope slope; DA, Dau; BR, Breebaart; and PO, phase opponency. Different subjects are indicated with different symbols connected with lines to facilitate intersubject and cross-model comparisons. Note that subject identification numbers do not correspond to the same subjects across studies. The critical r^2 value for a significant prediction would be 0.16 for comparison of one model to the detection pattern for one subject. A Bonferroni correction was used to compensate for the comparison of each subject's data to seven models in Study 2 resulting in a criterion of significance for r^2 equal to 0.28 and for five models in Study 3 resulting in a criterion equal to 0.26 (horizontal-dashed lines).

terns for $P(Y|T+N)$ and for 7 out of 8 cases for $P(Y|N)$, whereas while the MDS model prediction reached significance in 6 out of 8 cases for $P(Y|T+N)$ and in 5 out of 8 cases for $P(Y|N)$. The amount of variance predicted by the MDS model was lower than that predicted by the MD model in all but one case. It should be noted that the MD model includes a fit to the subject data, which partly explains the success of this model (see discussion of this issue in Davidson *et al.*, 2006) relative to the models that have fixed parameter values, except for the mean and slope of the regression line. However, Davidson *et al.* (2006) showed that the weighting strategy of the MD model produced r^2 values significantly greater than would be expected by simply adding free parameters to the CB model.[6] Overall, the MD model accounted for more of the variance in the subjects' detection patterns than any other model tested in this study. Davidson *et al.* (2006) found relatively little variation across subjects in the diotic results (e.g., as compared to the dichotic results, see below), which contributes to the success in fitting these data sets with a single model.

3. ES model

A modified version of Richards' (1992) ES model (Zhang, 2004) was evaluated. The ES model estimates the rate and magnitude of fluctuation in the stimulus envelope as a decision variable. Although, as will be shown later, envelope fluctuation co-varies with stimulus energy when energy variations are present, the ES model is not strictly dependent on energy, and this model can make meaningful predictions even when energy is normalized, as it was in Study 3. Implementation details for the ES model are described in Appendix B.

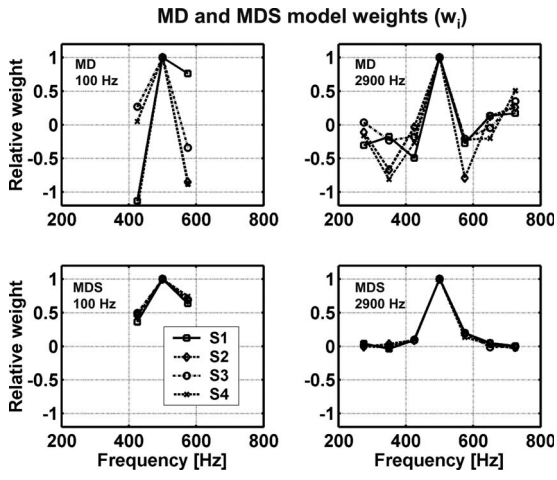


FIG. 3. Weights computed for MD and MDS models for the 100- and 2900-Hz data in Study 2. Weights are shown for the four subjects. Note that weights in each condition were normalized to the maximum weight (occurring at 500 Hz). These weights correspond to the w_i for the MD and MDS models in Fig. 1 and Eq. (A2).

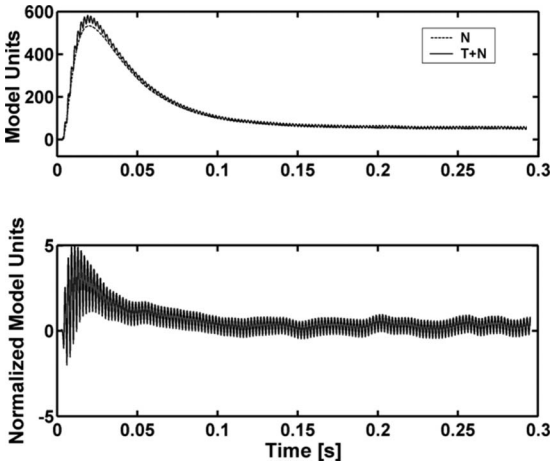


FIG. 4. Templates used in the Dau model decision device. The T+N and N templates were computed as the mean of 500 internal representations of T+N and N stimuli, respectively. The templates (plotted in arbitrary model units) are the low-pass filtered outputs of the adaptation loops in the Dau model (see Appendix C). The lower panel shows the normalized difference between the two templates in the upper panel.

In general, the ES model predictions were no better than, and often poorer than, the predictions made with the CB, MD, and MDS models, with only 13 of the 32 predictions reaching significance for P(Y|T+N) [Fig. 2(A)] and only 12 of the 32 predictions reaching significance for P(Y|N) [Fig. 2(B)]. Model predictions were highly variable across subjects in all studies, except for P(Y|T+N) for the wideband results of Study 2. Despite this model's simplicity, it was able to account for more variance in some of the subject's detection patterns than the more complicated temporal models (i.e., DA, BR, and PO; see below), particularly in the equal-energy cases of Study 3 for P(Y|N) [Fig. 2(B)].[7]

4. Dau model

The Dau model (Dau *et al.*, 1996a) has been used to predict thresholds in a number of monaural and diotic psychophysical tasks, including detection of tones in random and frozen noise as a function of the temporal position, duration, and frequency of the tone, as well as forward and backward masking tasks (Dau *et al.*, 1996b). The model includes bandpass filters to represent peripheral filtering, plus rectification, a simplified model of adaptation, and a low-pass filter to extract the envelope (Fig. 1). This model's DV is computed by comparing an "internal representation" of the response for each stimulus to a template. Details of the implementation of the Dau model are included in Appendix C.

Like the ES model, the Dau model relies primarily on the temporal envelope of the stimulus waveform, but the Dau model allows some fine structure to pass onto the decision device (Fig. 4) because the low-pass filter used for envelope extraction is only first-order. This model uses a distinct template-matching strategy, where a previously computed N template is subtracted from the waveform on each trial, and the result is compared to a normalized version of the difference between previously computed T+N and N templates. Figure 4 shows representative T+N and N templates (top panel) and the normalized difference template (bottom panel) for the 100-Hz bandwidth stimuli of Study 2. Each trace in the top panel shows the output of the model's adaptation loops (see Appendix C) averaged over 500 stimulus waveforms. It is clear that the averaging process brings out some fine-structure information related to the tone frequency in the T+N template. This information is effectively increased by the normalization with respect to the difference between the two templates (Fig. 4, bottom panel). The difference between templates is largest at the onset of the noise waveform because of the lack of compression in the adaptation loops for stimuli with fast changes in sound pressure level (whereas the latter portion of the difference is compressed). The covariation in time of the fine structure present in the stimulus waveform and in the internal template (requiring the detector to have knowledge of the phase of the target tone) also contributes to the decision variable.

In general, the predictions of the Dau model were not significantly correlated to the subjects' detection patterns. Only 5 of the 32 P(Y|T+N) predictions [Fig. 2(A)] and only 2 of the 32 P(Y|N) predictions [Fig. 2(B)] were significant. Note also that these results were obtained despite the fact that the Dau decision variables are at least partially correlated to overall energy, as discussed further below.

5. Breebaart model

The peripheral processing in the Breebaart model (Breebaart *et al.*, 2001a) is similar to that of the Dau model. Differences between the predictions of the Breebaart and Dau models result from differences in the decision devices, including the template mechanisms. For example, in the Breebaart model, the N template is subtracted from the internal representation of each stimulus waveform as a measure of the "distance" from the N stimulus, which differs from the normalized difference strategy of the Dau model. Other features in the Breebaart model include temporal weighting using a double-sided exponential window and spectral weighting across multiple frequency channels. Details of the implementation are presented in Appendix D.

Representative templates of the Breebaart model are shown in Fig. 5 for the 100-Hz bandwidth stimuli of Study 2 for the three frequency channels used. The frequency weighting [see Appendix D, Eq. (D1)] is shown in Fig. 5. In this illustration, the time-varying weights are summed over time and normalized to facilitate comparison to the weights for the MD and MDS models. The weights for both narrowband and wideband results are similar to those in Fig. 3 for the MDS model. For our implementation of the Breebaart model, only 7 predictions of the 32 made for P(Y|T+N) were significant [Fig. 2(A)], and none of the 32 made for P(Y|N) were significant [Fig. 2(B)].

6. PO model

The PO (Fig. 1) model (Carney *et al.*, 2002) is a detection model that is based primarily on temporal cues in the stimulus fine structure. These cues are extracted using cross-frequency coincidence detection. This model successfully predicts that the detection threshold should be robust even in

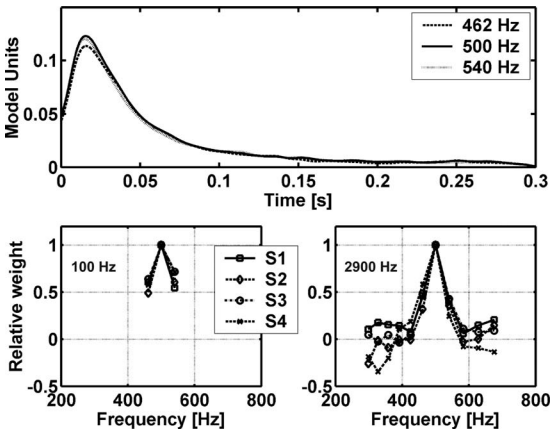


FIG. 5. Internal noise-alone templates (top panel) for the different frequency channels in the narrowband condition of Study 2 and frequency weighting (bottom two panels) imposed by the model of Breebaart *et al.* (2001a) on the internal representations of the stimulus waveforms. The internal representations (plotted in arbitrary model units) are the responses of Breebaart's peripheral model, followed by half-wave rectification and low-pass filtering to represent transduction by the inner hair cells, and adaptation loops (see Appendix D). The frequency weights were summed over time and normalized to the peak value for the 100- and 2900-Hz conditions of Study 2 for the left and right panels, respectively.

a roving-level paradigm (Carney *et al.*, 2002); however, it has not been previously tested with detection patterns estimated from stimuli with energy equalized across stimulus waveforms (as was the case for Study 3). Details of the implementation of this model are presented in Appendix E.

The PO model's ability to predict detection patterns was comparable to that of the DA model, with only 3 of the 32 $P(Y|T+N)$ predictions reaching significance [Fig. 2(A)] and 3 of the 32 $P(Y|N)$ predictions reaching significance [Fig. 2(B)]. This model performed no better for experiments in which energy was equalized (i.e., when fine structure might be expected to play more of a role) than for experiments with energy cues present.

B. Dichotic models

When out-of-phase tones are added to identical noise waveforms, as is done in the N_0S_π condition, the resulting tone-plus-noise stimulus waveforms have instantaneous interaural time differences (ITDs) and interaural level differences (ILDs) that vary over time due to interactions between the tones and the noise masker. Models for dichotic detection allow tests of hypotheses about the relations between these cues and the detection results and of hypotheses about the binaural mechanisms used to process these cues, such as cross correlation and equalization-cancellation. It should be noted that the ITDs and ILDs that are present in the tone-plus-noise stimuli are dynamic cues that vary throughout the time course of the stimuli, rather than the static interaural differences that are used in lateralization experiments. Some models for dichotic detection are based on the time averages of these interaural cues over the course of a stimulus waveform; others use the instantaneous, time-varying cues.

In general, dichotic models have been less successful at predicting the N_0S_π detection patterns than diotic models have been at predicting the N_0S_0 detection patterns. Isabelle (1995; i.e., Study 1) and Colburn *et al.* (1997) analyzed several decision variables for N_0S_π detection patterns. Colburn *et al.* (1997) considered the equalization-cancellation (EC) model and normalized cross correlation (NCC) and unnormalized cross correlation (UCC) models. They found that the EC DV was too strongly influenced by monaural stimulus energy, which did not correlate with subject performance, as compared to predictions based on non-linear combinations of interaural difference cues present in the acoustic stimuli, which showed significantly better correlation with performance in N_0S_π experiments (however, an internal noise model would be required for the interaural difference models to make predictions on N trials). They also found that the UCC model was too dependent on masker waveform, rather than on the addition of the tone to the masker waveform, for tone-plus-noise stimuli. Decision variables for the UCC model were almost identical regardless of signal presence (that is, hit and false-alarm rates were more similar for the model predictions than in the data). Finally, Colburn *et al.* (1997) found that the NCC model is equivalent to the EC model when using multiplicative time and amplitude jitter, such that the DV was again heavily dependent on the energy in each waveform. Isabelle (1995) showed that the variation in the NCC DV based on the addition of the tone was too weak compared to the dependence of the NCC DV on masker energy to predict his data. Isabelle (1995) (see Isabelle and Colburn, 2004) was able to explain at most about 50% of the variance in his N_0S_π data or in the Isabelle and Colburn (1991) data using stimulus energy (as a substitute for the EC and NCC models), standard deviations of ITDs and ILDs, and decision variables computed using various combinations of ITDs and ILDs. The highest model correlations to subject data in the Isabelle (1995) study were based on Webster's (1951) time-deviation model, which included a subject-dependent (i.e., fitted) parameter related to the threshold of time-deviation detection. However, the correlation of this model to the data was not significantly better than those of the simpler model based on the standard deviation of ITD; the latter model was included in the results presented here.

In the current study, several decision variables (standard deviations of ITD, ILD, and combinations thereof) related to those used in Isabelle (1995) were re-examined using the data from Studies 2 and 3. In addition, the related decision variables from Goupell and Hartmann (2007) were also examined. The Goupell and Hartmann (2007) decision variables extended the Isabelle (1995) decision variables and included two distinct classes that make use of both ITD and ILD: "Independent-center" models, in which integration over time occurs separately for the decision variables based on ITD and ILD, and "auditory-image" models, in which ITDs and ILDs interact as a function of time, before integration across time. The results from Study 3 (Davidson, 2007; Davidson *et al.*, 2009) suggest that the Isabelle (1995) decision variables could not predict the detection patterns because they do not allow envelope (ILDs) and fine structure (ITDs) to interact temporally. Thus, it was of interest to determine the effectiveness of the Goupell and Hartmann (2007) auditory-image decision variables that allow for this

interaction. The lateral position model (Hafter, 1971; Isabelle, 1995; Isabelle and Colburn, 2004) was also evaluated because it includes an explicit interaction between envelope and fine-structure cues in the form of a trading ratio.

A variant of the Marquardt and McAlpine (2001) model for masked detection was also tested; this model has been shown to successfully predict masked-detection thresholds using only four binaural delay channels [henceforth referred to as the four-channel (FC) model]. This model was inspired by the findings of McAlpine *et al.* (2001), who reported that interaural phase tuning of delay-sensitive neurons in the guinea pig inferior colliculus was centered around 45°, regardless of the neurons' best frequencies. The binaural counterpart of the Breebaart model was also tested; this model makes use of temporal fine structure in the binaural processor.

1. Independent-center and auditory-image models

As described above, Isabelle (1995) used several decision variables that were based either on fine structure (i.e., ITDs) or on envelope (i.e., ILDs). These included the standard deviations of ITD and ILD and a weighted linear combination of the variances of ITD and ILD. Goupell and Hartmann (2007) referred to these as independent-center models because the variances of ITD and ILD were computed *before* their weighted combination was computed. Goupell and Hartmann (2007) introduced what they referred to as auditory-image decision variables, in which ITD and ILD were combined before computing the variance or averaging over time. The auditory-image decision variables used here included the standard deviation of a temporal combination of ITD and ILD as well as the average absolute value of the temporal combination of ITD and ILD. Finally, Isabelle's (1995) implementation of Hafter's (1971) lateral position model was evaluated in the present study; this model was also one of the auditory-image models considered by Goupell and Hartmann (2007).

The Isabelle (1995) and Goupell and Hartmann (2007) decision variables are based on interaural differences calculated directly from the stimulus waveforms.[8] Because internal noise was not used, the decision variables described in this section would have been identically zero for noise-alone stimuli; therefore, predictions were not computed for P(Y|N). Details of the calculations of these decision variables are provided in Appendix F.

In order to provide a comparison and to confirm previous work, a simple energy-based model was also used to predict the detection patterns for the dichotic condition. The energy of the stimulus waveforms delivered to the two ears differed very slightly due to the addition of out-of-phase tones to the two waveforms, but on average this difference was very small for stimuli with tones added at N_0S_π threshold levels. Therefore, the energy (EN) based model used here was simply based on the rms energy of the right stimulus waveform. The EN model performed poorly, with none of the 37 predictions reaching significance (Fig. 6), consistent with Isabelle (1995) and indicating that the cue used by subjects to perform the detection task was not simply correlated to energy. Standard deviations of ITDs and ILDs (sT and sI)

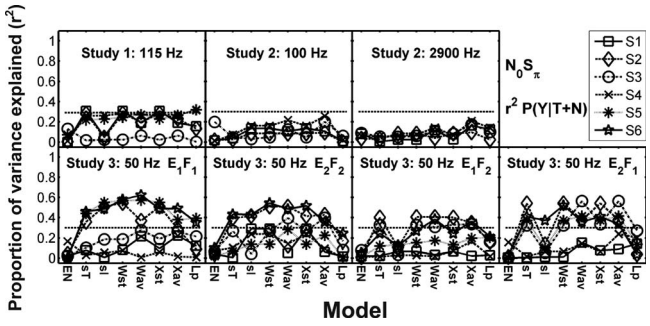


FIG. 6. Proportion of variance explained (r^2) by the N_0S_π model predictions for selected Isabelle (1995) and Goupell (2005) decision variables for z-scores of P(Y|T+N). EN is the rms energy of the right stimulus waveform, sT is the standard deviation of ITDs, sI is the standard deviation of ILDs, Wst is a linear combination of the standard deviations of ITDs and ILDs, Wav is a linear combination of the average absolute values of ITDs and ILDs, Xst is the standard deviation of a linear combination of ITDs and ILDs, Xav is the average value of a linear combination of ITDs and ILDs, and Lp is a lateral position model relating ITDs and ILDs with a trading ratio. Results for Studies 1 and 2 are shown on the first row, and Study 3 is shown on the second row. Each column represents a different bandwidth condition within Study 2 or a different stimulus set in Study 3. Different subjects are indicated with different symbols connected with lines to facilitate intersubject and cross-model comparisons. Note that subject identification numbers do not correspond to the same subjects across studies. The critical r^2 value for reaching a significant prediction ($p<0.05$), including a Bonferroni correction for comparison of each data set to 11 models (across Figs. 6 and 9(A)), is 0.29 for Study 1 and 0.30 for Studies 2 and 3, as denoted by the horizontal dashed lines. Symbols are connected to facilitate comparisons across models.

performed somewhat better, with 12 and 6 of the 37 predictions reaching significance, respectively (Fig. 6). Linear combinations of the standard deviations or of the average absolute values of ITD and ILD performed slightly better than the other models, suggesting that both envelope and fine-structure cues contribute to the detection process. Of the 37 predictions, 12 that used a weighted combination of the standard deviations of ITD and ILD as a DV (Wst) reached significance, and 13 that used an average of the absolute value of the weighted combination of ITD and ILD as a DV (Wav) reached significance. Note that the models that depended on weighted combinations of cues involved a fit to the subjects' data. Predictions for models that computed separate decision variables before combining across ITD and ILD processors (Wst and Wav) accounted for about the same amount of variance in P(Y|T+N) as those that first combined ITDs and ILDs as a function of time before computing decision variables. These variables were the standard deviation of the temporal combination of ITD and ILD (Xst), the average of the absolute value of the temporal combination of ITD and ILD (Xav), and the estimated lateral position (Lp) (see Appendix F for details). The Xst, Xav, and Lp decision variables made 12, 12, and 4 significant predictions, respectively, for the 37 comparisons performed (Fig. 6). In summary, the independent-center (Wst and Wav) and "auditory-image" (Xst and Xav) decision variables had about the same predictive power.

The weights placed on ITD or ILD decision variables were also examined for possible trends across subjects and for relations to the threshold SNR. Figure 7 shows weights organized by model and subject for all three studies (see

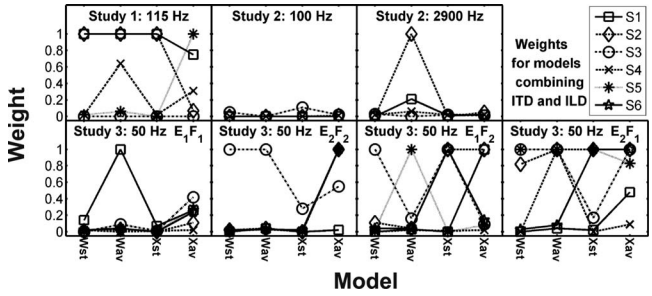


FIG. 7. Model weights for decision variables based on both ITDs and ILDs. A weight approaching 1 indicates reliance on ITD and a weight approaching 0 indicates reliance on ILD. Note that subject identification numbers do not correspond to the same subjects across studies, but corresponding symbols are used in Figs. 6 and 7. Different subjects are indicated with different symbols connected with lines to facilitate intersubject and cross-model comparisons.

Appendix F for details of weight calculations). Weights were bounded by 0 and 1, with 1 indicating total reliance on ITD and 0 indicating total reliance on ILD. Figure 7 shows that the results of these models were largely due to their ability to exclusively select the DV that was better correlated with the individual subject's responses from either sT or sI. Certain subjects, however, used a true weighted combination of the two decision variables (e.g., S4 in Study 1), and in almost all cases, these subjects had relatively low thresholds. Subjects with higher thresholds were fitted more reliably with weights of either 0 or 1, indicating that they relied solely on ITDs or ILDs.

2. FC model

The general structure of the FC model (Marquardt and McAlpine, 2001) is shown in Fig. 8 (upper panel). The right and left stimulus waveforms were processed using the Heinz *et al.* (2001) auditory-nerve model. The output of each filter was passed to a delay line with a phase shift of 45° on each side, which corresponds to a delay of 250 μs for a 500-Hz stimulus. The delayed stimulus from the ipsilateral side and delayed stimulus from the contralateral side converged onto

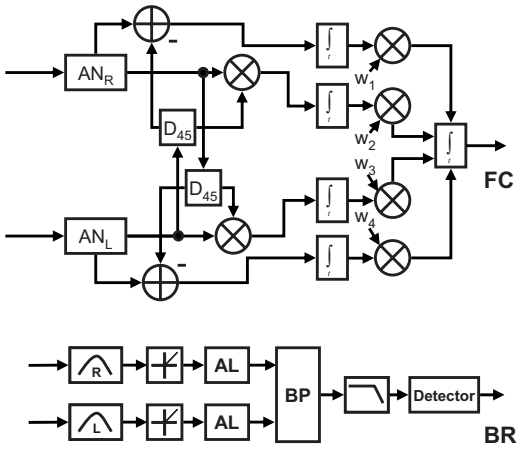


FIG. 8. Block diagrams of the FC and binaural Breebaart (BR) models used to predict N_0S_π detection patterns. AN denotes the auditory-nerve model of Heinz *et al.* (2001). *D* denotes a delay block computed based on the center frequency of each model auditory-nerve fiber. AL denotes the adaptation loops as described in Dau *et al.* (1996a). BP denotes a binaural processor.

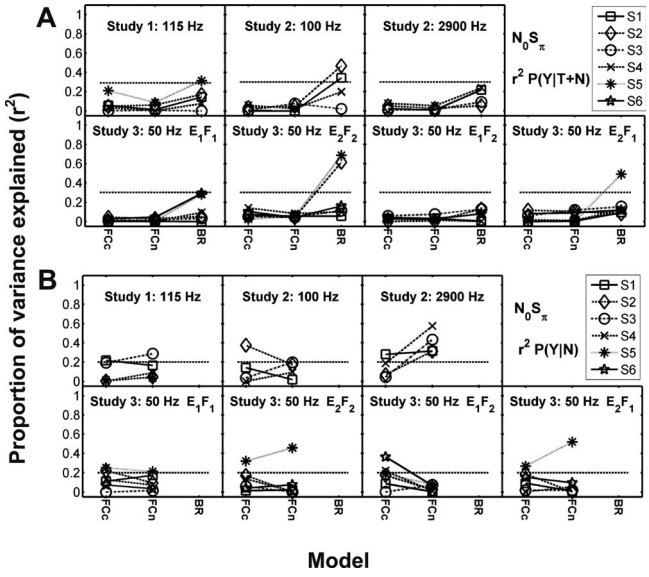


FIG. 9. Proportion of variance explained (r^2) by N_0S_π model predictions for *z*-scores of (A) P(Y|T+N) and (B) P(Y|N). Model abbreviations are FCc, FC model using an UCC for simulated peaker channels (i.e., channels with ITD curves that are characterized by a central peak); FCn, FC model using a NCC for peaker channels; and BR, the binaural Breebaart model. Different subjects are indicated with different symbols connected across models to facilitate intersubject comparisons. Note that subjects sharing the same number do not correspond across studies, but corresponding symbols are used in Figs. 6 and 7. Symbols are connected to facilitate comparisons across models. The critical r^2 value for reaching a significant prediction ($p<0.05$), including a Bonferroni correction for comparison of each data set to 11 models for P(Y|T+N) (across Figs. 6 and 9(A)), is 0.29 for Study 1 and 0.30 for Studies 2 and 3, and including a correction for comparison to two models for P(Y|N) (Fig. 9(B)) the critical r^2 value is 0.20 for all three studies, as indicated by the horizontal dashed lines.

binaural coincidence detectors; both the UCC and the binaural cancellation (difference) were computed for each channel in order to approximate neurons that are excited by stimuli to both ears (EE) and neurons that are excited by stimulation of one ear and inhibited by the other (EI), respectively. The resulting FCs thus corresponded to cells tuned to ±45° and ±135°, spanning the entire range of possible interaural phase differences at 500 Hz in relative increments of 90°. Note that this model is a special case of the standard cross correlation model; it differs from the general model because it is restricted to a particular subset of channel correlations. The outputs of the four binaural channels were suboptimally weighted and summed using the same strategy as was used for the MDS model (see Appendix G for details).

Two versions of this model were implemented; the common structure of both versions is shown in the upper part of Fig. 8 (FC). The first (FCc) used a cross correlation (product) of the inputs for the channels tuned to ±45° (weighted by w_2 and w_3 in Fig. 8), while the second (FCn) used a NCC (Colburn *et al.*, 1997) for these channels. Neither version of the model made consistently significant predictions of the detection patterns (Fig. 9), with predictions reaching significance for none of the 74 comparisons (including both FCc and FCn) for P(Y|T+N) [Fig. 9(A)], 9 of the 37 P(Y|N) comparisons were significant for FCc, and 8 of the 37 P(Y|N) comparisons were significant for FCn [Fig. 9(B)].

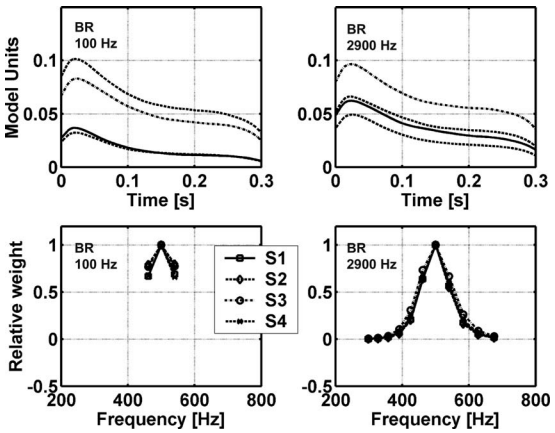


FIG. 10. Representative temporal weights (top) and spectral weights (bottom) for the binaural Breebaart model for two different masker bandwidths (100 and 2900 Hz). Note that the onset of the stimulus is weighted more heavily than the steady-state portion of the stimulus. The temporal weights (plotted in arbitrary model units) are the internal representations for the binaural model and thus include monaural peripheral processing (tuning, rectification, low-pass filtering, and adaptation loops) followed by binaural excitatory-inhibitory (EI) processing, a double-exponential filter, and logarithmic compression (see Appendix H).

3. Binaural Breebaart model

The binaural version of the Breebaart model is an extension of the monaural model described above and is based on an interaural subtraction (EI) algorithm. A simplified block diagram is shown in Fig. 8 (BR). The output of the adaptation loops from the ipsilateral and contralateral sides passes to a binaural processor that simulates an excitatory-inhibitory interaction. As originally designed (Breebaart *et al.*, 2001a), the model included a series of attenuation taps and delays with different values and selected the single delay and attenuation channel that showed the greatest change in output between T+N and N stimuli. However, because the zero delay and zero attenuation channel always has the greatest change in output for N_0S_π stimuli, the model was reduced to this single channel. Thus, this model has a structure that is generally similar to the EC model, but it differs in the details of the decision variable. Details of the implementation of this model are provided in Appendix H.

As with the Isabelle (1995) and Goupell and Hartmann (2007) decision variables, the binaural version of the Breebaart model produced decision variables that were identically 0 for N stimuli because of the subtraction mechanism [see Eq. (H1)]; thus, predictions for P(Y|N) were not made. Figure 10 shows representative temporal and spectral weights for the binaural version of the model computed for the two bandwidths of Study 2. Note that the onset was weighted more heavily than the steady-state portion of the stimulus because of the action of the adaptation loops. This model produced significant predictions for only 6 of the 37 comparisons to P(Y|T+N) (Fig. 9), performing more poorly than Wst or Wav despite its more complex and arguably more physiologically realistic structure. Note that although a few individual r^2 values in Fig. 9 are high for the BR model, its predictions are not consistently significant across the stimulus sets used in Study 3, as was true for the Wst and Wav predictions (see Fig. 6).

C. Comparisons of diotic models

Although none of the models were able to predict a significant proportion of the variance in subjects' detection patterns in every case, it was of interest to determine how similar or different each model's DV was to those of the other models. Because the models operated at different SNRs for each subject, the models' decision variables varied slightly across subjects. To simplify comparisons between diotic models, and because the threshold SNRs were within 3 dB under diotic condition for all subjects within each study, the SNR of the subject closest to the median threshold for each study (see Table II) was selected. Correlations between model decision variables in response to the reproducible stimuli are presented in terms of r^2 in Tables II and III for P(Y|T+N) and P(Y|N), respectively, for the each of the models in Fig. 1. Blank values indicate study conditions for which a given model did not apply.

Tables II and III show that the CB model was significantly correlated to each of the diotic models tested here. The most highly correlated models were the CB and MDS models, with r^2 values from 0.95 to 0.98. Both of these models were also significantly correlated to the MD model, which is not surprising given that all three models use energy at the output of one or more critical bands as decision variables. The CB and ES models were also significantly, albeit weakly, correlated for the stimuli of Study 2 and for the non-chimeric stimulus sets in Study 3. The CB, MD, and MDS models were significantly correlated to the DA and the BR models for most noise-alone [P(Y|N)] results for experiments that had differences in energy across stimulus waveforms (i.e., Study 2). These correlations were expected, given the Dau and BR models' envelope dependences. Finally, the Dau and BR models were significantly correlated for every case tested, as were the PO and ES models. It is also interesting to note that nearly all of the models were significantly correlated for the wideband stimuli of Study 2.

The contribution of stimulus energy to each of the model decision variables was tested with a multiple regression approach: two models were used together to predict the subjects' data, and the CB model was one of the two predictor models. An incremental F test (Edwards, 1979) was used to determine if the addition of the second model significantly increased the proportion of predicted variance. This procedure was equivalent to testing the significance of the partial correlation coefficient or the significance of the slope of a predictor variable in a multiple regression analysis. Results are briefly summarized in the text below in terms of the increase in R^2 (the proportion of variance explained, with the upper-case R indicating a result of a multiple regression) achieved by adding the second model to the CB model for both P(Y|T+N) and P(Y|N).

Of all the diotic models tested in combination with the CB model [192 tests were run in total for P(Y|T+N) and 192 for P(Y|N); six models were tested for Study 2 with four subjects and two bandwidths and for Study 3 with six subjects and four stimulus sets], only regression analyses that included the MD, ES, or PO models as second predictors yielded significantly better predictions than the CB model

TABLE II. Correlations between model decision variables in response to reproducible T+N stimuli under the N_0S_0 condition are shown in terms of r^2. Model abbreviations are the same as in Fig. 1. Note that the MD and MDS models were not applied to the 50-Hz bandwidth of Study 3. Each model was run at an E_S/N_0 of 11.8 dB for the 100-Hz bandwidth condition of Study 2, 10.8 dB for the 2900-Hz bandwidth condition of Study 2, and 10 dB for all stimulus sets in Study 3.

	Study 2		Study 3			
Model comparison	100 Hz	2900 Hz	E_1F_1	E_2F_2	E_2F_1	E_1F_2
CB-MD	0.78[a]	0.70[a]				
CB-MDS	0.95[a]	0.98[a]				
CB-ES	0.33[a]	0.56[a]	0.22[a]	0.51[a]	0.07	0.12
CB-DA	0.00	0.09	0.04	0.00	0.17[a]	0.01
CB-BR	0.10	0.28[a]	0.01	0.00	0.11	0.09
CB-PO	0.01	0.20[a]	0.23[a]	0.30[a]	0.14	0.10
MD-MDS	0.79[a]	0.65[a]				
MD-ES	0.34[a]	0.55[a]	0.22[a]	0.51[a]	0.07	0.12
MD-DA	0.00	0.07	0.04	0.00	0.17[a]	0.01
MD-BR	0.07	0.30[a]	0.01	0.00	0.11	0.09
MD-PO	0.00	0.21[a]	0.23[a]	0.30[a]	0.14	0.10
MDS-ES	0.21[a]	0.50[a]	0.22[a]	0.51[a]	0.07	0.12
MDS-DA	0.00	0.08	0.04	0.00	0.17[a]	0.01
MDS-BR	0.09	0.25[a]	0.01	0.00	0.11	0.09
MDS-PO	0.00	0.19[a]	0.23[a]	0.30[a]	0.14	0.10
ES-DA	0.00	0.05	0.11	0.00	0.16[a]	0.00
ES-BR	0.01	0.33[a]	0.03	0.00	0.03	0.03
ES-PO	0.34[a]	0.51[a]	0.49[a]	0.41[a]	0.46[a]	0.55[a]
DA-BR	0.31[a]	0.38[a]	0.86[a]	0.81[a]	0.74[a]	0.89[a]
DA-PO	0.00	0.13	0.26[a]	0.24[a]	0.36[a]	0.22[a]
BR-PO	0.05	0.20[a]	0.18[a]	0.26[a]	0.22[a]	0.09

[a] $p < 0.05$.

TABLE III. Correlations between model decision variables in response to reproducible N stimuli under the N_0S_0 condition are shown in terms of r^2. Model abbreviations are the same as in Fig. 1. Note that the MD and MDS models were not applied to the 50-Hz bandwidth of Study 3. SNRs are the same as for Table II.

	Study 2		Study 3			
Model comparison	100 Hz	2900 Hz	E_1F_1	E_2F_2	E_2F_1	E_1F_2
CB-MD	0.57[a]	0.56[a]				
CB-MDS	0.91[a]	0.97[a]				
CB-ES	0.20[a]	0.44[a]	0.30[a]	0.23[a]	0.06	0.01
CB-DA	0.40[a]	0.34[a]	0.00	0.02	0.05	0.05
CB-BR	0.41[a]	0.08	0.01	0.03	0.10	0.13
CB-PO	0.11	0.37[a]	0.03	0.03	0.00	0.00
MD-MDS	0.60[a]	0.50[a]				
MD-ES	0.22[a]	0.35[a]	0.30[a]	0.23[a]	0.06	0.01
MD-DA	0.35[a]	0.35[a]	0.00	0.02	0.05	0.05
MD-BR	0.20[a]	0.20[a]	0.01	0.03	0.10	0.13
MD-PO	0.05	0.33[a]	0.03	0.03	0.00	0.00
MDS-ES	0.05	0.37[a]	0.30[a]	0.23[a]	0.06	0.01
MDS-DA	0.33[a]	0.27[a]	0.00	0.02	0.05	0.05
MDS-BR	0.41[a]	0.05	0.01	0.03	0.10	0.13
MDS-PO	0.01	0.31[a]	0.03	0.03	0.00	0.00
ES-DA	0.20[a]	0.22[a]	0.01	0.00	0.00	0.03
ES-BR	0.03	0.32[a]	0.04	0.03	0.01	0.02
ES-PO	0.78[a]	0.58[a]	0.58[a]	0.29[a]	0.56[a]	0.39[a]
DA-BR	0.35[a]	0.47[a]	0.81[a]	0.78[a]	0.69[a]	0.88[a]
DA-PO	0.12	0.35[a]	0.23[a]	0.01	0.05	0.02
BR-PO	0.02	0.34[a]	0.34[a]	0.02	0.11	0.06

[a] $p < 0.05$.

TABLE IV. Correlations between model decision variables in response to reproducible T+N stimuli under the N_0S_π condition are shown in terms of r^2. Here, models were tested at the *lowest* subject threshold from each study; thus each model was run at an E_S/N_0 of 1 dB for Study 1, −6.2 dB for the 100-Hz bandwidth condition of Study 2, −2.2 dB for the 2900-Hz bandwidth condition of Study 2, and −17 dB for all stimulus sets in Study 3. Model abbreviations are the same as in Fig. 8 and Sec. III B 1.

	Study 1	Study 2		Study 3			
Model comparison	115 Hz	100 Hz	2900 Hz	E_1F_1	E_2F_2	E_2F_1	E_1F_2
EN-Wav	0.30[a]	0.43[a]	0.30[a]	0.05	0.15	0.08	0.01
EN-Xav	0.30[a]	0.43[a]	0.32[a]	0.01	0.12	0.34[a]	0.01
EN-FCn	0.23[a]	0.12	0.08	0.00	0.04	0.04	0.00
EN-BR	0.06	0.20[a]	0.10	0.04	0.01	0.01	0.01
Wav-Xav	1.00[a]	1.00[a]	0.78[a]	0.93[a]	1.00[a]	0.83[a]	1.00[a]
Wav-FCn	0.07	0.01	0.03	0.01	0.14	0.03	0.15
Wav-BR	0.18[a]	0.42[a]	0.29[a]	0.14	0.52[a]	0.56[a]	0.39[a]
Xav-FCn	0.08	0.01	0.02	0.03	0.13	0.01	0.15
Xav-BR	0.16[a]	0.42[a]	0.18[a]	0.11	0.52[a]	0.35[a]	0.39[a]
FCn-BR	0.05	0.00	0.01	0.05	0.07	0.11	0.00

[a]$p < 0.05$.

alone. That is, the variance explained by the other models (MDS, DA, and BR) "overlaps" with the variance already explained by the CB model. Significant increases in R^2 values by the addition of the MD model as a predictor were in the range of 0.10–0.33 for T+N stimuli and 0.10–0.36 for N stimuli, depending on the subject. Significant increases resulting from the addition of the ES model were in the range of 0.10–0.32 for T+N stimuli and of 0.10–0.52 for N stimuli. Significant increases resulting from adding the PO model were in the range of 0.08–0.46 for T+N stimuli and 0.16–0.21 for N stimuli.

These quantitative comparisons of the predictions of the diotic models illustrate the strong similarities between the CB, MD, MDS, DA, and BR models. The results suggest that a superior model might be constructed by combining the across-frequency structure of the MD model and the mechanisms in the ES and PO models. More detailed discussion of the diotic models is included below in Sec. IV.

D. Comparisons of dichotic models

Comparisons between EN, Wav, Xav, FCn, and BR models are shown in Tables IV and V for $P(Y|T+N)$ and for different levels. Recall that it appeared that subjects with substantially different thresholds were likely to be using distinct detection strategies. Therefore, Table IV presents comparisons when the SNR was equal to the threshold of the best subjects in each study, and Table V presents comparisons when the SNR was equal to the threshold of the poorest subjects in each study.

In Table IV, correlations between Wav and Xav, and Wav and BR are examined (Wst and Xst are not included because these models are highly correlated to Wav and Xav, respectively.) Stimulus energy (EN) was significantly, albeit modestly, correlated to Wav, Xav, and the BR models for some stimulus sets. The results of Table V are similar; however, as was expected for the responses of subjects with

TABLE V. Correlations between model decision variables in response to reproducible T+N stimuli under the N_0S_π condition are shown in terms of r^2. Here, models were tested at the *highest* subject threshold from each study; thus each model was run at an E_S/N_0 of 18 dB for Study 1, 3.8 dB for the 100-Hz bandwidth condition of Study 2, 4.8 dB for the 2900-Hz bandwidth condition of Study 2, and 0 dB for all stimulus sets in Study 3. Model abbreviations are the same as in Fig. 8 and Sec. III B 1.

	Study 1	Study 2		Study 3			
Model comparison	115 Hz	100 Hz	2900 Hz	E_1F_1	E_2F_2	E_2F_1	E_1F_2
EN-Wav	0.48[a]	0.35[a]	0.45[a]	0.03	0.22[a]	0.12	0.10
EN-Xav	0.25[a]	0.43[a]	0.39[a]	0.00	0.56[a]	0.21[a]	0.06
EN-FCn	0.31[a]	0.01	0.12	0.00	0.03	0.02	0.00
EN-BR	0.19[a]	0.34[a]	0.10	0.03	0.04	0.01	0.01
Wav-Xav	0.87[a]	0.75[a]	0.84[a]	0.95[a]	0.00	0.67[a]	0.63[a]
Wav-FCn	0.59[a]	0.05	0.06	0.09	0.02	0.34[a]	0.08
Wav-BR	0.37[a]	0.46[a]	0.37[a]	0.17[a]	0.34[a]	0.17[a]	0.14
Xav-FCn	0.39[a]	0.04	0.09	0.10	0.09	0.14	0.00
Xav-BR	0.25[a]	0.49[a]	0.24[a]	0.17[a]	0.02	0.44[a]	0.46[a]
FCn-BR	0.47[a]	0.07	0.10	0.11	0.02	0.19[a]	0.04

[a]$p < 0.05$.

TABLE VI. The contribution of cues based on envelope and fine structure to N_0S_0 model decision variables. Results for human subjects from Study 3 are presented along with results from each computational model. The entries R_E^2 and R_F^2 denote the proportion of variance in the baseline detection patterns explained by chimeric detection patterns measured using stimuli sharing the same envelopes or fine structures as the baseline stimuli, respectively. If the addition of envelope as a predictor to fine structure as a predictor significantly increased the proportion of variance explained (i.e., R_{EF}^2 was significantly higher than R_F^2), then R_E^2 is underlined. If the addition of fine structure as a predictor to envelope as a predictor significantly increased the proportion of variance explained (i.e., R_{EF}^2 was significantly higher than R_E^2), then R_F^2 is underlined. Note that the R_{EF}^2 values are the proportion of variance explained by a multiple regression including both envelope and fine structure as predictors; all R_{EF}^2 values were statistically significant ($p<0.05$). Model abbreviations are the same as in Fig. 1.

			P(Y|T+N)			P(Y|N)		
N_0S_0		E_S/N_0	R_E^2	R_F^2	R_{EF}^2	R_E^2	R_F^2	R_{EF}^2
Subjects	S1	10	0.17	0.18	0.41	0.17	0.59	0.69
	S2	10	0.26	0.29	0.37	0.18	0.48	0.54
	S3	10	0.05	0.28	0.32	0.00	0.71	0.72
	S4	11	0.69	0.11	0.73	0.45	0.30	0.68
	S5	11	0.35	0.07	0.38	0.20	0.39	0.49
	S6	11.5	0.66	0.25	0.66	0.35	0.36	0.58
Models	DA	10	0.97	0.16	0.98	0.91	0.30	0.94
	BR	10	0.88	0.12	0.93	0.71	0.20	0.90
	ES	10	0.94	0.03	0.95	0.88	0.02	0.92
	PO	10	0.82	0.21	0.85	0.40	0.43	0.71

higher thresholds, the correlations of many of the model decision variables to energy were stronger at the higher SNR. Note also that Wav and Xav were slightly less correlated at the higher SNRs. [Also, note that the perfect correlations between Wav and Xav in some cases (Table IV) are due to the fact that Xav with weight $a=0$ or 1 reduces to Wav with weight $a=0$ or 1.]

Colburn *et al.* (1997) and Isabelle (1995) discounted models based on UCC, NCC, and EC mechanisms because of their dependence on stimulus energy. The comparisons in Tables IV and V showed only moderate correlations between energy and either the binaural BR or FC model, both of which include correlation and/or cancellation mechanisms. (All stimulus sets of Study 3 had nearly equal energies; thus, correlations of the BR and FC models to energy would be expected to be near zero.) Because energy was not correlated to the subjects' detection patterns, the failure of these models can be partially explained by their moderate correlations to stimulus energy.

E. Interactions between envelope and fine-structure cues

Some of the models tested in this work included interactions between cues derived from envelope and fine structure before computing decision variables. It was of interest to determine if the nature of these interactions was appropriate as compared to the interactions observed in the empirical data collected in Study 3 (Davidson, 2007; Davidson *et al.*, 2009). This was examined by comparing the extent to which each of the models relied on envelope or fine structure and the extent to which each of the subjects relied on envelope or fine structure. The stimuli and analysis techniques used in the present study were the same as those used in Study 3.

The analysis procedure used by Davidson (2007) and Davidson *et al.* (2009) is briefly described here: Stimuli from four stimulus sets (E_1F_1, E_2F_2, E_1F_2, and E_2F_1) shared either envelopes (E) or fine structures (F); subscripts shared between stimulus sets indicate that the particular waveform component was shared between sets. Model detection patterns were computed, and subjects' detection patterns were measured for each of the four stimulus sets. A multiple linear regression was performed that used model detection patterns from the "chimeric" stimulus sets (E_1F_2 and E_2F_1) to predict the baseline model detection patterns (E_1F_1 and E_2F_2). To simplify the analysis, the detection patterns that shared the predictor, either envelope or fine structure, were combined (i.e., concatenated). If the model (or subject) relied exclusively on envelope, the detection patterns for the baseline stimulus sets and those for the stimulus sets sharing the same envelopes should have been identical. If the model (or subject) relied exclusively on fine structure, the detection patterns from the baseline stimulus sets and the stimulus sets sharing the same fine structures should have been identical. The multiple regression procedure quantified the similarity of each detection pattern to the baseline detection patterns. Three R^2 values were produced for each model and subject. R_E^2 was the proportion of variance accounted for when detection patterns with common envelopes were used as predictors; R_F^2 was the proportion of variance accounted for when detection patterns with common fine structures were used as predictors. R_{EF}^2 was the proportion of variance accounted for when *both* detection patterns with common envelopes and detection patterns with common fine structures were used as predictors. The R_E^2 is underlined in Tables VI and VII if the addition of envelope as a predictor along with fine structure significantly increased the proportion of variance explained (i.e., if R_{EF}^2 was significantly greater than R_F^2), and R_F^2 is underlined if R_{EF}^2 was significantly greater than R_E^2 (i.e., if addition of the fine structure as a predictor significantly increased the proportion of variance, as compared to that ex-

TABLE VII. Same as Table VI except data are presented for the N_0S_π condition. Computational models were tested using stimuli with the lowest (−17 dB) and with the highest (0 dB) threshold observed for the subjects. All R^2_{EF} values were statistically significant ($p<0.05$). Note that significant values of 0.00 in this table represent very small but significant R^2 values that were rounded to zero; these small but significant values were observed in cases when a single predictor explained nearly all of the variance in the data (e.g., R^2_F values of 0.99–1.0), such that even a very small value of R^2_E (<0.005) resulted in a significant increment in R^2_{EF}, as determined by the incremental F-test. Model abbreviations are the same as in Fig. 8 and Sec. III B 1.

			P(Y|T+N)			P(Y|N)		
N_0S_π		E_S/N_0	R^2_E	R^2_F	R^2_{EF}	R^2_E	R^2_F	R^2_{EF}
Subjects	S1	0	0.12	0.43	0.44	0.22	0.41	0.57
	S2	−10	0.13	0.56	0.57	0.14	0.55	0.59
	S3	−17	0.00	0.38	0.39	0.14	0.14	0.35
	S4	−1	0.41	0.30	0.68	0.40	0.26	0.63
	S5	−16.5	0.06	0.45	0.51	0.25	0.15	0.47
	S6	−10	0.39	0.27	0.50	0.32	0.15	0.52
Models	sT	−17	0.04	0.95	0.95			
	sI	−17	0.67	0.19	0.68			
	Wav	−17	0.06	0.99	0.99			
	Xav	−17	0.00	0.99	0.99			
	Lp	−17	0.01	0.30	0.32			
	FCc	−17	0.24	0.23	0.52	0.02	0.27	0.28
	FCn	−17	0.25	0.25	0.46	0.62	0.01	0.66
	BR	−17	0.02	0.55	0.56			
	sT	0	0.00	1.00	1.00			
	sI	0	0.99	0.16	0.99			
	Wav	0	0.00	1.00	1.00			
	Xav	0	0.22	0.90	0.91			
	Lp	0	0.05	0.34	0.32			
	FCc	0	0.30	0.26	0.38	0.27	0.85	0.98
	FCn	0	0.17	0.25	0.28	0.90	0.85	0.97
	BR	0	0.00	0.85	0.75			

plained by the envelope alone). For a more detailed description of the methods used here and in Study 3, see Davidson (2007) and Davidson *et al.* (2009).

Results are presented for simulations using a SNR of 10 dB E_S/N_0 for the N_0S_0 condition (Table VI). Model abbreviations are as in Fig. 3. All of the N_0S_0 models relied more heavily on envelope than fine structure ($R^2_E>R^2_F$) with the exception of the PO model, which made approximately equal use of envelope and fine-structure cues on noise-alone trials. In general, the patterns of model interactions between envelope and fine structure (i.e., R^2 values) were in stark contrast to the results of the human subjects presented in the same table, which indicated that subjects relied roughly equally on envelope and fine-structure cues. The only notable exception was for the PO model, which predicted a more equal utilization of envelope and fine-structure cues than the other models [but recall that the PO model captured at most about 40% of the variance in subjects' detection patterns (Fig. 2)].

Results are also presented for the N_0S_π condition (Table VII). Each model was evaluated twice: once with the SNRs set to the highest threshold observed for the human subjects in Study 3 and once at the lowest SNRs for the subjects. Every model except sI was dominated by the fine structure of the waveform, as would be expected for conventional ITD-based models of binaural detection at low frequencies. The interaction pattern for most models differs from the results of the subjects, which indicated a more equal reliance on envelope and fine structure. The Lp, FCc, FCn, and BR models produced R^2 values that were most similar to the human subjects for T+N stimuli. For some of the linear regression analyses, R^2 values are (much) higher than would be expected based on the subjects' data (e.g., all but the PO model for N_0S_0 conditions, and the Wav and Xav models at low SNRs for N_0S_π conditions). This result suggests that the dependence of the model detection patterns on the envelope and fine-structure cues was more straightforward, and thus more predictable, than for the human subjects. The subject data from Study 3 indicate that a linear combination of average cues derived from the envelope and fine structure should not account for all of the predictable variance in the E_1F_1 and E_2F_2 detection patterns and suggest instead models may need some form of running temporal combination of envelope and fine-structure cues. Thus, the reliance on envelope and fine structure is likely a necessary, but insufficient, condition for predicting subjects' detection patterns. Internal noise added at the decision stage would reduce the correlations between model results and the envelope or fine-structure cues, but addition of internal noise would not produce the patterns of correlations seen in the data.

IV. CONCLUSIONS AND FUTURE WORK

The results of the present study show that several existing diotic models that have successfully predicted subjects' thresholds for tone-in-noise detection tasks cannot explain diotic detection patterns for reproducible noise maskers. In particular, none of the temporal models examined in this work were able to predict significant proportions of variance in all subjects' data; this was true even when energy cues were made unreliable, forcing subjects to rely on cues other than overall energy. A model based on a linear combination of energies at the output of several filters surrounding the target frequency (MD; Ahumada and Lovell, 1971; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986) best predicted the N_0S_0 data for stimuli with level variations between noises (Study 2). A model based on envelope fluctuation (ES; Richards, 1992; Zhang, 2004) predicted N_0S_0 detection patterns estimated using equal-energy stimuli (Study 3) more accurately than either the Dau *et al.* (1996a) or Breebaart *et al.* (2001a) models.

Implementations of several models of binaural detection were also tested. The models that made the most significant predictions used linear combinations of the average absolute values or of the standard deviations of ITDs and ILDs (Wst, Wav, Xst, and Xav; Isabelle, 1995; Isabelle and Colburn, 2004; Goupell and Hartmann, 2007). The binaural version of the model of Breebaart *et al.* (2001a) made fewer significant predictions than Xav and Wav, but seemed to more appropriately weight the use of stimulus envelope and fine structure in the computation of the model decision variable. As for the diotic condition, none of the models tested were comprehensive enough to make significant predictions for every subject in every stimulus condition.

Although the template-based models examined here (Dau *et al.*, 1996a, 1996b; Breebaart *et al.*, 2001a, 2001b, 2001c) did not predict a large portion of the variability in the subjects' data, they are capable of predicting thresholds for a multitude of psychophysical tasks and will be examined more thoroughly in future work. Trial-by-trial responses were not simulated here, and a running template was not computed. Computation of a running template would be an interesting modeling exercise, which would more fully examine the potential of these specific models and would also provide an initial investigation of the general class of detection mechanisms that have the ability to change dynamically over time. Some of the subjects (including some of the authors) have reported being influenced by particular noise waveforms, or even feeling temporarily confused for brief periods (i.e., tens of trials) during an experiment. Individual responses and waveform identification numbers were recorded on each trial for the experiments presented here, providing data suitable for an interesting analysis of template-based models. Suppose that a template was constructed as the mean of several preceding trials of randomly generated noise. Suppose also that this memory was a buffer of a limited number of waveforms in a first-in, first-out configuration. Model predictions for the data in Studies 1–3 could be re-examined as a function of the buffer length (or the number of internal representations of the stimuli used to compute an average template). This analysis is possible because responses to each waveform can be used to sort waveforms into *perceived* tone-plus-noise and noise-alone groups, regardless of the stimuli used for each trial.

The models and analyses presented in this paper assume that normally-distributed internal noise is *added* at the decision stage. Implementing the models without complicated or model-specific internal noise had advantages for this particular study (as described in Sec. II). This decision was not without consequences. Many of the dichotic models cannot make predictions for N trials without implementing a more complicated internal noise model or the introduction of some form of processing asymmetry. Thus, implementing these models without internal noise is incomplete but informative. The internal noise for the diotic models seems likely to make only marginal changes in their predictions for individual samples, at least as revealed by the correlation analyses used here.

A possible drawback of this modeling approach (and for that matter, a drawback of any of the models used in this study) is that the potential use of short-term cues is not captured by the template mechanisms employed in the above models. Subjects have reported that relatively brief segments of stimuli were often the basis for decisions during dichotic detection tasks. This fact compounds the modeling problem because the temporal locations of these stimulus segments are unknown and may differ among waveforms. Cues that occur in brief segments of the stimuli are not well suited for detection with the temporal weighting scheme of the Breebaart model, which averages across waveforms. Further, a strategy based on short-term cues would likely require a rethinking (i.e., shortening) of the time constant used for smoothing the output of the binaural processor in the Breebaart model. Recent evidence suggests that the relatively long estimates of binaural temporal windows, 60–200 ms (e.g., Grantham and Wightman, 1979; Kollmeier and Gilkey, 1990; Culling and Summerfield, 1998), may, in fact, be too long, and estimates on the order of 50 ms or shorter might be more suitable for modeling the current data (Bernstein *et al.*, 2001; Kolarik and Culling, 2005). Researchers' testing temporal aspects of binaural processing have reported time constants as short as 10 ms (e.g., Akeroyd and Bernstein, 2001). Note that the variances of the interaural differences depend on the distributions of short-time estimates of interaural differences, even though it is only the spread of values that is used for a decision.

Another possibility is that subjects may employ more than one type of (potentially short-term) template. This strategy could be investigated by grouping waveforms by their respective hit and false-alarm rates, and then investigating the templates that result from training the model with waveforms corresponding to high, moderate, and low hit rates.

Another class of models worth further investigation includes those based on the spectrum of the envelope of amplitude fluctuations, such as modulation filter bank models (e.g., Berg, 2004; Dau *et al.*, 1997a, 1997b). These models have been successful at predicting average thresholds for

low-frequency diotic tone-in-noise detection tasks, but were outside the scope of this paper. Future studies will test these models with the reproducible noise data.

Several of the models examined in this study incorporated information from multiple frequency channels (e.g., the MD model and MDS, the MD model with suboptimal weights) using linear weighting schemes that were either sub-optimal or fit to the subjects' data. Preliminary simulations using optimal weighting strategies, such as linear discriminant analyses, produced improved model-data correlations with respect to sub-optimal weighting schemes. These results suggest that subjects may be exploiting correlations between frequency channels to perform the detection task. Future modeling efforts will investigate the use of optimal linear across-frequency weighting for the multi-channel models treated in this study.

Another suggestion for future modeling efforts is inspired by Hancock and Delgutte (2004). Results from the current study suggest that a single binaural delay/attenuation model cannot explain detection of tones masked by reproducible noise stimuli. The Hancock and Delgutte (2004) model was originally designed to predict ITD discrimination data and is based on recordings from the inferior colliculus of cat. The model employs a neuronal pooling strategy that combines responses across a population of model neurons tuned in best frequency and ITD according to distributions measured in cat. It is possible that responses of a population of channels tuned to a number of different ITD values are necessary to account for the current data.

ACKNOWLEDGMENTS

This work was supported by NIDCD F31 077798 (S.A.D.), NIDCD 00100 (H.S.C.), and NIDCD 01641 (L.H.C. and S.A.D.). We thank Dr. Marty Sliwinski, Dr. Yan Gai, and Junwen Mao for helpful comments. We also thank Dr. Scott Isabelle for providing his data and stimulus waveforms. Dr. Torsten Dau and Dr. Jeroen Breebaart provided very helpful input regarding implementation strategies for their respective models. We also thank Susan Early for her editorial comments.

APPENDIX A: IMPLEMENTATION OF THE MD MODELS

The MD and MDS models were implemented using a linear combination of the rms output of three or seven fourth-order gammatone filters, depending on the masker bandwidth (Fig. 1). Davidson *et al.* (2006) showed that filters exceeding the bandwidth of the stimulus noise do not significantly increase the predictive power of the model. Therefore, the center frequencies were selected to span 275–725 Hz (in 75-Hz increments) for the 2900-Hz noise bandwidth condition of Study 2 and 425–575 Hz for the 100-Hz noise bandwidth condition of Study 2. The bandwidth of all of the filters was set to 75 Hz to match Davidson *et al.* (2006). The MD model was not used to predict the data from Study 3, as the masker bandwidth in that study was only 50 Hz.

The weights (w_i) for the linear combination were established with two separate methods. For the first method (the standard MD model), the weights were fitted to the individual subjects' detection patterns using the reproducible stimuli from each study (Fig. 3). The MATLAB function fminsearch was used to minimize the quantity of 1 minus the correlation coefficient between the linear combination of the rms filter outputs and the z-scores of $P(Y|T+N)$ or $P(Y|N)$ for each subject in each condition in Studies 1 and 2.

A second variant of the MD model was also tested. This model (the MDS model) used a decision-theoretic suboptimal weighting scheme to compute model weights (rather than fitting the weights to each subject's data). Individual weights were computed for the MDS model using 1000 repetitions of randomly created (i.e., not reproducible) noise. Tones were added and weights were computed as

$$w(i) = \frac{\overline{F(i,m)_{(T+N)}} - \overline{F(i,m)_{(N)}}}{\frac{\mathrm{var}(F(i,m)_{(T+N)}) + \mathrm{var}(F(i,m)_{(N)})}{2}}, \tag{A1}$$

where F is the root-mean-squared filter output for frequency channel i and random-noise repetition m for T+N or N stimuli, and the means and variances were computed across repetition m within frequency channel i. [Note that this method would be optimal if the covariance of each channel was accounted for in Eq. (A1).] Both models' decision variables were given by

$$M(j) = \sum_i F(i,j)w(i), \tag{A2}$$

where j is the reproducible noise waveform index, using the weights computed with either method above.

APPENDIX B: IMPLEMENTATION OF THE ES MODEL

The implementation of the ES model (Fig. 1) was the same as that in Davidson *et al.* (2006). The ES model DV was computed as

$$E(j) = \frac{\sum_t |x[t-\Delta t,j] - x[t,j]|}{\sum_t x[t,j]}, \tag{B1}$$

where $x[t,j]$ is the Hilbert envelope of the output of a fourth-order gammatone filter centered at 500 Hz, with a 75-Hz ERB for stimulus waveform j, and Δt is the time resolution of the sampled waveform. To ensure that all fine structure was removed from the stimulus waveform, $x[t,j]$ was filtered with a tenth-order maximally flat infinite impulse response (IIR) filter with a cut-off frequency of 250 Hz before being processed with Eq. (B1). The statistic was normalized as suggested by Zhang (2004) to remove the effects of energy and duration. Upon addition of the tone to the noise waveform, the stimulus envelope flattens. As such, the DV decreases with increasing tone level.

APPENDIX C: IMPLEMENTATION OF THE DAU MODEL

The Dau model (DA, Fig. 1) consists of a third-order gammatone filter centered at the tone frequency (500 Hz)

with a bandwidth of 1 ERB, approximately 75 Hz at a center frequency of 500 Hz (Glasberg and Moore, 1990). The output is half-wave rectified and passed to a series of adaptation loops (Dau *et al.*, 1996a), designed to simulate adaptation in auditory-nerve responses by processing fast stimulus fluctuations almost linearly and compressing slowly fluctuating stimuli. The output of the adaptation loops is low-pass filtered with a time constant of 20 ms (8 Hz) to remove fine structure and leave envelope information. The output at this stage is referred to as the internal representation of the model.

The internal representation is passed to an optimal detector. The optimal detector uses a template derived from the normalized difference between the mean of 500 T+N internal representations and the mean of 500 N internal representations (Fig. 4). A large number of noises were used to simulate extensive subject training. The templates were computed using randomly generated noise with a signal added at 10 dB above each subject's threshold. On each trial, the optimal detector first subtracts the noise-alone template from the internal representation computed from the reproducible stimulus on that trial. The mean scalar product of the normalized difference template and the difference between the noise-alone template and the internal representation of the reproducible stimulus is then computed as a function of time. The model was originally designed to pick the interval (from a two-interval task) with the larger scalar product as the one containing the tone. For the purposes of this study, which focuses on single-interval tasks, the scalar product itself was used as the decision variable. This process is summarized with the following equation:

$$D(j) = \frac{1}{T_d}\int_0^{T_d} [\varphi_j(j,t) - \gamma_{\overline{N}}(t)] \frac{[\gamma_{\overline{T+N}}(t) - \gamma_{\overline{N}}(t)]}{\mathrm{rms}[\gamma_{\overline{T+N}}(t) - \gamma_{\overline{N}}(t)]} dt, \tag{C1}$$

where D is the Dau decision variable, φ_j is the internal representation of the current stimulus waveform j, $\gamma_{\overline{T+N}}$ is the mean of 500 internal representations of T+N stimulus waveforms (the T+N template), $\gamma_{\overline{N}}$ is the mean of 500 internal representations of N stimuli (the N template), T_d is the duration of the stimulus waveform, and rms is the root-mean-squared function.

The code used to implement this model is available at www.bme.rochester.edu/carney (last viewed August 19, 2009).

APPENDIX D: IMPLEMENTATION OF THE MONAURAL BREEBAART MODEL

The diotic version of the Breebaart model (Breebaart *et al.*, 2001a) is shown in Fig. 1 (BR). This model is similar to the Dau model; however, the Breebaart model was implemented as a bank of processors with increasing center frequencies. Two filters per ERB were implemented over the same bandwidths as the MD and MDS models. The low-pass filter from the Dau model was replaced with a double-sided exponential window with time constants of 10 ms each. The structure of the decision device is described in detail in Breebaart *et al.* (2001a) and is composed of a suboptimally weighted combination of internal representations at different frequency channels, which are then summed as a function of time and frequency. Like the Dau model, the Breebart model also uses both T+N and N templates (Fig. 5). The templates were established as the means of 50 internal representations[9] of randomly generated T+N and N waveforms at each subject's threshold. The detector first computes the DV B according to the following equation:

$$B(j) = \int_F \int_\tau \frac{\mu(i,t)}{\sigma^2(i,t)} U(j,i,t) di dt. \tag{D1}$$

The quantity $U(j,i,t)$ is the difference between the internal representation of the reproducible waveform j and the N template for each frequency (F) channel i. $U(j,i,t)$ is weighted across frequency and time by the difference between the T+N and N templates (μ) normalized by the variance of the N templates (σ^2).

The code used to implement this model is available at www.bme.rochester.edu/carney (last viewed August 19, 2008).

APPENDIX E: IMPLEMENTATION OF THE PO MODEL

The PO model was computed as described in Davidson *et al.* (2006) and was based on the model described by Carney *et al.* (2002) (PO, Fig. 1). Two model auditory-nerve fibers of Heinz *et al.* (2001) with spontaneous rates of 50 spikes/s converged upon a coincidence detector of the type described in Colburn (1977). The fibers' center frequencies were selected such that their phase responses differed by 180° at the tone frequency (which occurred for the two center frequencies of 459 and 542 Hz). The count at the output of the coincidence detector was used for the model DV as described by

$$G(j) = n_{\mathrm{fib}}^2 T_{\mathrm{CW}} \int_0^{T_d} \kappa_{459}(j,t)\kappa_{542}(j,t) dt, \tag{E1}$$

where n_{fib} is the number of auditory-nerve fiber inputs at each center frequency, T_{CW} is the time window for coincidence detection, t is time, T_d is the duration of the stimulus, and κ is the output of the auditory-nerve model of Heinz *et al.* (2001) at each of the two center frequencies.

The mechanism used by the PO model is as follows: as the level of the tone is increased, and the responses of the fibers become more phase locked to the tone. The count at the output of the coincidence detector decreases as tone level increases because the two model fibers progress to firing perfectly out of phase. The model detects the tone on the basis of a reduction in the coincidence detector's average rate with respect to its response to the noise alone. The simulations presented here were performed at SNRs matched to the subjects' thresholds; the model DV was the coincidence detector's average rate. Ten model fibers were used with a coincidence window of 20 μs. As in Davidson *et al.* (2006), the onsets and offsets of the auditory-nerve fiber responses were truncated because they exceeded realistic levels and did not produce decision variables correlated to the psychophysical data. Due to the use of relatively short-duration stimuli in the

present study, only the first and last 25 ms of the responses were truncated. The model DV (G) was computed for each reproducible stimulus j.

APPENDIX F: IMPLEMENTATION OF THE INDEPENDENT-CENTERS AND AUDITORY-IMAGE MODELS

Isabelle's (1995) decision variables based on ITDs and ILDs for waveform j are given by

$$T(j,t) = \frac{\phi_L(j,t) - \phi_R(j,t)}{\omega_c} \tag{F1}$$

and

$$I(j,t) = 20 \log \frac{A_L(j,t)}{A_R(j,t)}, \tag{F2}$$

where $\phi(j,t)$ is the instantaneous phase computed from the complex analytic signal for the right (R) or left (L) stimulus waveforms, ω_c is the center frequency of the noise band, $A(j,t)$ is the envelope of the complex analytic signal for either the right or left stimulus waveform, and j is the index of each reproducible stimulus waveform. Note that T is only approximately equal to the ITD because the frequency of the masker stimulus varies as a function of time. The complex analytic signals were computed using the Hilbert transform. A selection of several decision variables related to those in Isabelle (1995) (see Isabelle and Colburn, 2004) and Goupell and Hartmann (2007) is shown below. These included the standard deviations[10] of ITD and ILD computed for each reproducible stimulus as defined by

$$s_T(j) = \left(\frac{1}{T_d} \int_0^{T_d} (T(j,t) - \overline{T(j)})^2 dt \right)^{1/2} \tag{F3}$$

and

$$s_I(j) = \left(\frac{1}{T_d} \int_0^{T_d} (I(j,t) - \overline{I(j)})^2 dt \right)^{1/2}, \tag{F4}$$

where T_d is the duration of the stimulus, and T and I are as defined above in Eqs. (F1) and (F2). A weighted combination of the standard deviations of ITD and ILD [Eq. (F5)] and a combination of the average absolute values of ITD and ILD [Eq. (F6)] were also explored, as defined by

$$W_{ST}(j,a) = a s_T(j) + (1-a) s_I(j), \tag{F5}$$

and

$$W_{AV}(j,a) = \frac{1}{T_d} \int_0^{T_d} a|T(j,t)|dt + \frac{1}{T_d} \int_0^{T_d} (1-a)|I(j,t)|dt, \tag{F6}$$

respectively, where a is a weight determined by minimizing the sum of squared errors between W and $z\{P(Y|T+N)\}$ for each condition and subject in each study. Note that the decision variables described by Eqs. (F5) and (F6) would fit into the class of independent-center models of Goupell and Hartmann (2007) because the standard deviations (or average absolute values) of ITD and ILD were computed before the weighted combination of ITD and ILD was computed. The four metrics described in Eqs. (F3)–(F6) were compared to the standard deviation of a temporal combination of ITD and ILD [Eq. (F7)] as well as the average value of the absolute value of the temporal combination of ITD and ILD [Eq. (F8)] defined by

$$X_{ST}(j,b) = \left(\frac{1}{T_d} \int_0^{T_d} (\{bT(j,t) + (1-b)I(j,t)\} - \overline{\{bT(j,t) + (1-b)I(j,t)\}})^2 dt \right)^{1/2} \tag{F7}$$

and

$$X_{AV}(j,b) = \frac{1}{T_d} \int_0^{T_d} |bT(j,t) + (1-b)I(j,t)|dt, \tag{F8}$$

respectively, where b is a weight computed in the same manner as in Eqs. (F5) and (F6), and T and I are as defined in Eqs. (F1) and (F2). The decision variables in Eqs. (F7) and (F8) would fit into the class of auditory-image models in Goupell and Hartmann (2007) because ITD and ILD were combined before computing the standard deviation or averaging over time.

Isabelle's (1995) implementation of Hafter's (1971) lateral position model was also considered. This model is based on a combination of ITD and ILD using a trading ratio of 20 μs/dB defined by

$$L_p(j,a) = \frac{1}{T_d} \int_0^{T_d} |T(j,t) + aI(j,t)|dt, \tag{F9}$$

where T_d is the duration of the stimulus, a is the trading ratio, and T and I are as defined in Eqs. (F1) and (F2). The lateral position model is similar to Eqs. (F7) and (F8), except that a constant trading ratio was used for all computations. These models [Eqs. (F3)–(F9)] were of particular interest because they allow for the distinct interaction of statistics based on envelope and fine structure as a function of time. Such interactions are implied by the results of Study 3 (Davidson, 2007; Davidson *et al.*, 2009).

APPENDIX G: IMPLEMENTATION OF THE FC MODEL

The DV of the FC model (Fig. 8) (Marquardt and McAlpine, 2001) was computed based on a linear combination of differences and cross correlations between channels with different delays:

$$F_C(j) = \{w_1[\kappa_R(j,t) - \kappa_L(j,t-\tau)]\} + [w_2 \kappa_R(j,t)\kappa_L(j,t-\tau)] + [w_3 \kappa_R(j,t-\tau)\kappa_L(j,t)] + \{w_4[\kappa_L(j,t) - \kappa_R(j,t-\tau)]\}, \tag{G1}$$

where $F_C(j)$ is the model DV for the reproducible stimulus j, $\kappa(j,t-\tau)$ is the output of the auditory-nerve model of Heinz *et al.* (2001) delayed by τ s (250 μs, corresponding to a phase delay of 45° at the 500-Hz signal frequency), and w is the suboptimal weight computed for each delay channel, as shown in Fig. 8. Weights were computed using a strategy similar to that used for the MDS model (see Appendix A). Weights derived for each of the four channels rarely took on

a value of zero, as channels were tuned in increments of 90° of interaural phase. Two different types of weights were tested: FCc used an UCC (product) of the inputs for the channels tuned to ±45° (weighted by w_2 and w_3 in Fig. 8), while the FCn used a NCC (Colburn *et al.*, 1997). For a more complete description of the FC model weights, see Davidson (2007) and Davidson *et al.* (2009).

APPENDIX H: IMPLEMENTATION OF THE BINAURAL BREEBAART MODEL

The binaural processor in the Breebaart model (Fig. 8) (Breebaart *et al.*, 2001a) is described by

$$E(j,i,t) = [(\varphi_L(j,i,t) - \varphi_R(j,i,t))]^2 \quad \text{(H1)}$$

for N_0S_π stimuli, where $\varphi(j,i,t)$ describes the output of the adaptation loops for reproducible stimulus j, frequency channel i, at time t, for the left or right ear. (Only the 0.0 delay and 0.0 attenuation channel were included in these predictions, as explained in the main text.) The processor output $E(j,i,t)$ is then filtered with a double-exponential window with a time constant of 30 ms per exponential. The filtered signal, $E'(j,i,t)$, is then scaled, compressed with a logarithm, and then scaled again as follows:

$$E''(j,i,t) = a \log(bE'(j,i,t) + 1), \quad \text{(H2)}$$

with $a=0.1$ and $b=0.000\,02$. The two scale factors were calibrated by setting the model threshold to predict N_0S_π and $N_\rho S_\pi$ detection tasks, as described in Breebaart *et al.* (2001a). The detector stage for the binaural model is similar to that for the monaural models [Eq. (D1)]. However, for the binaural case, the temporally weighted internal representation of each waveform is integrated over both time and frequency to compute the decision variable. The templates used to compute the weights are computed using the compressed and filtered outputs of the binaural processor. This method differs from computing a difference between T+N and N templates and comparing to the double-exponential filtered output of the adaptation loops as in the monaural model; recall that the N template is identically zero for the binaural model.

[1]The z-score was set to 0.005 for P(Y|T+N) or P(Y|N) values that were equal to 0 and to 0.995 for P(Y|T+N) or P(Y|N) values that were equal to 1 in order to avoid infinite z-score values.

[2]Note that, although the parameters of most of the models examined here were set to fixed values suggested by the previous literature and were not fitted to each subject's data (exceptions are specifically identified in the text), all of the fits had two free parameters, the slope and intercept of the line relating the values of the model's DV to the z-scores of the subject. Note that the square of Pearson's product-moment correlation is a measure of the variance predicted by this linear statistical model. As such, the reports of proportion of predicted variance assume a linear model with slope and intercept fit to the data.

[3]When model d's were large (because the SNR was set to the threshold of a poor subject or simply because the models did not include internal noise), r^2 values for predictions of the detection patterns that included both hits and false alarms, referred to as P(Y|W) (Davidson *et al.*, 2006), were artificially high because of the separation between the distributions of P(Y|T+N) and P(Y|N). Thus, modeling analyses presented here were confined to predictions of P(Y|T+N) and P(Y|N). The net effect of analyzing hit and false-alarm rates separately was to lower the proportions of variance explained with respect to the variance that might be explained in P(Y|W).

[4]The values reported in this paragraph are based on the correlation between the first-half and last-half of the data in terms of P(Y|N) and P(Y|T+N), not z-scores. However, simulations indicate that similar values would be obtained if z-scores had been used.

[5]Inspection of Fig. 2 shows some significant predictions for the energy model (CB) under these equal-energy conditions. These predictions *appear* significant because no internal noise was used in the simulations. One might suspect that the peripheral filter included in this model recovered energy differences across stimuli. The largest difference between levels at the output of the gammatone filter for the stimuli in Study 3 was about 1 dB. For the CB model to explain these results, given the variability of the hit and false-alarm rates in the detection patterns and also the variability of the energy-based decision statistic, the subjects would have had to reliably measure the output of a CB filter with a resolution of about 0.04 dB (to correctly order 25 T+N or N stimulus waveforms in terms of level) in the presence of internal noise with an effective variance of approximately 1 dB across noises (estimated assuming the internal-to-external noise ratio is approximately 1 for the data from Study 2 in the conditions where the data are correlated to the CB model; see Evilsizer *et al.*, 2002).

[6]It was also of interest to determine whether the MD model predictions were significantly better than the MDS model predictions. For all subjects but S3 [see Fig. 2(B), P(Y|N)], the MD model made better predictions than the MDS model. Tests of significant differences between non-independent correlations were computed for each subject with each stimulus bandwidth to test the hypothesis that the MD model was significantly better at predicting detection patterns than the MDS model. Results indicated that the MD model predicted significantly ($p<0.05$) more variance in P(Y|T+N) for S2, S3, and S4 in the 100-Hz bandwidth condition and for S1 and S4 in the 2900-Hz bandwidth condition, and more variance in P(Y|N) for S1 and S2 in the 100-Hz bandwidth condition and for S2 in the 2900-Hz bandwidth condition. Thus, the MD weighting strategy did, for some subjects, make significantly better predictions than the MDS weighting strategy.

[7]Note that the ES model predictions shown here explain less variance than those in Davidson *et al.* (2006) because predictions were made separately for P(Y|T+N) and P(Y|N) in the present study, whereas Davidson *et al.* (2006) made predictions for the combined detection pattern, P(Y|W).

[8]These decision variables were also computed using fourth-order gammatone filters centered at 500 Hz. However, in all but the 2900-Hz case, predictions were poorer when peripheral filtering was used, and therefore, those results were not included in this document. These decision variables were also tested using the auditory-nerve models of Heinz *et al.* (2001) and Zilany *et al.* (2006). Poor results (i.e., worse than those achieved with no peripheral processing) were also encountered using the auditory-nerve models as a peripheral processing stage, but this was likely due to the fact that these decision variables rely on the complex-analytic signal, which is not well defined for the output of the auditory-nerve models (the outputs of which have nonzero dc components). Therefore, the predictions that used the peripheral model of Heinz *et al.* (2001) are not shown.

[9]This number was reduced from 500 for practical considerations. The sensitivity of model decision variables to the number of internal representations was not great; results were stable for 20 or more repetitions.

[10]Standard deviations were used in this study as they resulted in slightly, but not significantly, better predictions than did variances. Isabelle (1995) also mentioned that results based on standard deviation and variance were not significantly different.

Ahumada, A., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Ahumada, A., Marken, R., and Sandusky, A. (**1975**). "Time frequency analyses of auditory signal detection," J. Acoust. Soc. Am. **57**, 385–390.

Akeroyd, M. A., and Bernstein, L. R. (**2001**). "The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses," J. Acoust. Soc. Am. **110**, 2516–2526.

Berg, B. G. (**2004**). "A temporal model of level-invariant tone-in-noise detection," Psychol. Rev. **111**, 914–930.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). "Sensitivity to brief changes of interaural time and interaural intensity," J. Acoust. Soc. Am. **109**, 1604–1615.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001a**). "Binaural processing model based on contralateral inhibition I. Model structure," J. Acoust. Soc. Am. **110**, 1074–1088.

Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001b**). "Binaural processing model based on contralateral inhibition II. Dependence on spectral

parameters," J. Acoust. Soc. Am. **110**, 1089–1104.
Breebaart, J., van der Par, S., and Kohlrausch, A. (**2001c**). "Binaural processing model based on contralateral inhibition III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.
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.
Colburn, H. S. (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.
Colburn, H. S., Isabelle, S. K., and Tollin, D. J., **1997**, in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, New York), pp. 533–556.
Culling, J. F., and Summerfield, Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.
Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection," J. Acoust. Soc. Am. **102**, 2906–2919.
Dau, T., Kollmeier, D., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.
Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.
Davidson, S. A. (**2007**). "Detection of tones in reproducible noise: Psychophysical and computational studies of stimulus features and processing mechanisms," Ph.D. thesis, Syracuse University, Syracuse, NY.
Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (**2006**). "Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results," J. Acoust. Soc. Am. **119**, 2258–2275.
Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. **2009**. "Diotic and dichotic detection with reproducible chimeric stimuli," J. Acoust. Soc. Am. **126**, 1889–1905.
Edwards, A. L., **1979**, *Multiple Regression and the Analysis of Covariance* (Freeman, New York).
Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H. (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.
Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.
Gilkey, R. H., and Robinson, D. E. (**1986**). "Models of auditory masking: A molecular psychophysical approach," J. Acoust. Soc. Am. **79**, 1499–1510.
Gilkey, R. H., Robinson, D. E., and Hanna, T. E. (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.
Glasberg, B. R., and Moore, B. J. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Goupell, M. J. (**2005**). "The use of interaural parameters during incoherence detection in reproducible noise," Ph.D. thesis, Michigan State University, East Lansing, MI.
Goupell, M. J., and Hartmann, W. H. (**2007**). "Binaural models for the detection of interaural coherence III. Narrowband experiments and binaural models," J. Acoust. Soc. Am. **122**, 1029–1045.
Grantham, D. W., and Wightman, F. L. (**1979**). "Detectability of time-varying interaural correlation in narrow-band noise stimuli," J. Acoust. Soc. Am. **65**, 1509–1517.
Hafter, E. R. (**1971**). "Quantitative evaluation of a lateralization model of masking-level differences," J. Acoust. Soc. Am. **50**, 1116–1122.
Hancock, K., and Delgutte, B. (**2004**). "A physiologically based model of interaural time difference discrimination," J. Neurosci. **24**, 7110–7117.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). "Auditory-nerve model for predicting performance limits of normal and impaired listeners," ARLO **2**, 91–96.
Isabelle, S. K. (**1995**). "Binaural detection performance using reproducible stimuli," Ph.D. thesis, Boston University, Boston, MA.
Isabelle, S. K., and Colburn, H. S. (**1991**). "Detection of tones in reproducible narrow-band Noise," J. Acoust. Soc. Am. **89**, 352–359.
Isabelle, S. K., and Colburn, H. S. (**2004**). "Binaural detection of tones masked by reproducible noise: Experiment and models," BU-HRC Report No. 04:01, Boston University, Boston, MA.
Kolarik, A. J., and Culling, J. (**2005**). "Measuring the binaural temporal window," J. Acoust. Soc. Am. **117**, 2563.
Kollmeier, B., and Gilkey, R. H. (**1990**). "Binaural forward and backward masking: Evidence for sluggishness in binaural detection," J. Acoust. Soc. Am. **87**, 1709–1719.
Marquardt, T., and McAlpine, D. (**2001**). "Simulation of binaural unmasking using just four binaural channels," Assoc. Res. Otolaryngol. Abstr. **24**, 87.
McAlpine, D., Jiang, D., and Palmer, A. R. (**2001**). "A neural code for low-frequency sound localization in mammals," Nat. Neurosci. **4**, 396–401.
Pfafflin, S. M., and Mathews, M. V. (**1966**). "Detection of auditory signals in reproducible noise," J. Acoust. Soc. Am. **39**, 340–345.
Richards, V. M. (**1992**). "The detectability of a tone added to narrow bands of equal energy noise," J. Acoust. Soc. Am. **91**, 3424–3425.
Siegel, R. A., and Colburn, H. S. (**1989**). "Binaural processing of noisy stimuli: Internal/external noise ratios under diotic and dichotic stimulus conditions," J. Acoust. Soc. Am. **86**, 2122–2128.
Webster, F. A. (**1951**). "The influence of interaural phase on masked thresholds, I. The role of interaural time-deviation," J. Acoust. Soc. Am. **23**, 452–462.
Zhang, X. (**2004**). "Cross-frequency coincidence detection in the processing of complex sounds," Ph.D. thesis, Boston University, Boston, MA.
Zilany, M. S. A., and Bruce, I. C. (**2006**). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," J. Acoust. Soc. Am. **120**, 1446–1466.

Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise

Trevor R. Agus[a)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow G31 2ER, United Kingdom; Department of Psychology, University of Strathclyde, 40 George Street, Glasgow G1 1QE, United Kingdom; and CNRS, Universite Paris Descartes, and Ecole Normale Superieure, 29 rue d'Ulm, 75005 Paris, France

Michael A. Akeroyd and Stuart Gatehouse[b)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow, G31 2ER, United Kingdom

David Warden
Department of Psychology, University of Strathclyde, 40 George Street, Glasgow G1 1QE, United Kingdom

(Received 8 October 2008; revised 24 July 2009; accepted 25 July 2009)

Three experiments measured the effects of age on informational masking of speech by competing speech. The experiments were designed to minimize the energetic contributions of the competing speech so that informational masking could be measured with no large corrections for energetic masking. Experiment 1 used a "speech-in-speech-in-noise" design, in which the competing speech was presented in noise at a signal-to-noise ratio (SNR) of −4 dB. This ensured that the noise primarily contributed the energetic masking but the competing speech contributed the informational masking. Equal amounts of informational masking (3 dB) were observed for young and elderly listeners, although less was found for hearing-impaired listeners. Experiment 2 tested a range of SNRs in this design and showed that informational masking increased with SNR up to about an SNR of −4 dB, but decreased thereafter. Experiment 3 further reduced the energetic contribution of the competing speech by filtering it into different frequency bands from the target speech. The elderly listeners again showed approximately the same amount of informational masking (4–5 dB), although some elderly listeners had particular difficulty understanding these stimuli in any condition. On the whole, these results suggest that young and elderly listeners were equally susceptible to informational masking.
 [DOI: 10.1121/1.3205403]

PACS number(s): 43.66.Dc, 43.66.Sr, 43.71.Lz, 43.71.Ky [RLF] Pages: 1926–1940

I. INTRODUCTION

Elderly listeners generally have more listening difficulties than young listeners, particularly when listening to one of several simultaneous talkers (CHABA, 1988). This undoubtedly stems in part from the reduced functioning of their ears, but there could also be some "higher-level" hearing difficulties, for example, in confusing the simultaneous talkers. Such difficulties have often been demonstrated experimentally in normal-hearing listeners, as there is a greater amount of masking when competing speech is present compared to some acoustically equivalent condition without speech (e.g., Brungart, 2001; Freyman *et al.*, 2001; Arbogast *et al.*, 2002). This effect is termed "informational masking." Previously, we have demonstrated that informational masking may be factor in people's self-report of their listening ability, in that an analysis of responses to the "Speech, Spatial, and Qualities of Hearing" questionnaire (Gatehouse and Noble, 2004) showed lower self-reported performance in situations likely to involve informational masking than in control situations (Agus *et al.*, 2009). Here, we report direct measurements of the effect of age and hearing loss on the informational masking of speech by speech.

How best to define informational masking remains an open question (e.g., Durlach *et al.*, 2003; Watson, 2005; Durlach, 2006), but in order to obtain an experimentally tractable definition it is convenient to use Lutfi's (1990) equation

$$\text{total masking} = \text{informational masking} + \text{energetic masking}, \quad (1)$$

or, equivalently,

$$\text{informational masking} = \text{total masking} - \text{energetic masking}. \quad (2)$$

Lutfi (1990; Oh and Lutfi, 1998) took (A) the masked threshold for a "stimulus known statistically" (i.e., a greater-uncertainty condition) as the measure of "total masking" in Eq. (2) and (B) the masked threshold observed for a "stimulus known exactly" (i.e., a minimal-uncertainty condition) as the measure of "energetic masking." The amount of uncertainty in a signal can be quantified for experiments with tonal signals, but it is not clear how it can be measured for speech signals. As the focus in this paper is on the informational

a)Author to whom correspondence should be addressed. Electronic mail: trevor.agus@ens.fr
b)Stuart Gatehouse was closely involved in the design and analysis of this experiment before his death in February 2007.

masking of speech by speech, we used corresponding definitions instead: (A) the measure of total masking was taken as the masked threshold in a condition with competing speech as part of the masker, and (B) the measure of energetic masking was taken as the masked threshold in an acoustically equivalent control condition *without* competing speech. That is,

$$\text{informational masking} = \text{SRT}_{\text{speech}} - \text{SRT}_{\text{nonspeech}}, \quad (3)$$

where $\text{SRT}_{\text{speech}}$ and $\text{SRT}_{\text{nonspeech}}$ are the speech-reception thresholds in the presence or absence of competing speech. This equation underlies much of the research on informational masking in speech, as it makes no claims about the source of the informational masking—it could be due to similarity or uncertainty (Watson, 2005), modulation masking (Kwon and Turner, 2001), "semantic interference" (Carhart *et al.*, 1969a), or otherwise unknown facets of speech-on-speech masking—although it does require that the speech and nonspeech maskers generate the *same* amount of energetic masking.

A modulated, speech-shaped noise is often used as the nonspeech masker (e.g., Festen and Plomp, 1990; Brungart, 2001), as it can represent both the overall temporal fluctuations of speech and its overall long-term spectrum. However, it still has *some* acoustical differences to speech—as indeed would any noise masker—and if these differences affect the energetic masking compared to actual speech, then the amount of informational masking derived using Eq. (3) could be misleading.

An alternative to finding a perfect acoustically equivalent masker is instead to use a masking noise in *both* conditions. That is, the masker in the total-masking condition is a combination of competing speech plus a noise, and the masker in the energetic-masking condition is just the noise. The essence of this design is that the noise should contribute almost all the energetic masking, while the competing speech contributes the informational masking but only a minimal amount of additional energetic masking: thus energetic and informational masking are conceptually separated. This goal can be achieved by taking advantage of the well-known result that the speech-reception threshold for a sentence in noise is generally negative—when expressed as a speech-to-noise ratio in decibels—meaning that the sentence has less power than the noise when it is at threshold (e.g., −8 dB; Hawkins and Stevens, 1950). For example, if a sentence at a long-term average speech-to-noise ratio of −8 dB is added to a noise at 60-dB SPL, then the sentence can (just) be identified, and it might be expected to create a small amount of informational masking. But because the level of the sentence (52-dB SPL) is considerably below the level of the noise (60-dB SPL), the total long-term average power of the combination is almost entirely due to the power of the noise and is indeed is less than a decibel greater (60.6-dB SPL).[1] The crucial experimental comparison is thus between the SRTs in the combined masker and in the noise-alone masker: if the difference is also 0.6 dB in this example, then the entire effect can be attributed solely to the average increase in acoustical power. If, however, the difference is larger than 0.6 dB, then there must have been some additional masking that cannot be attributed purely to the acoustical power of the maskers. Any such additional masking should mostly be informational masking, and so a simple application of Eq. (2) allows it to be calculated.

Carhart *et al.* (1968) pioneered this design. They used a speech-to-noise ratio of 0 dB and so assumed that the increase in masker power due to the combination was 3 dB; their noise masker was a modulated noise, not a static noise. As part of a larger experiment, they measured SRTs of spondees in either a combination of competing speech plus modulated noise or in modulated noise alone. They observed 7.8 dB more masking in the combined maskers: of this, 3 dB was attributed to the increase in total power of the maskers and so the remaining 4.8 dB was attributed to "semantic interference," which would now be termed informational masking. However, since modulated noise was used, some of the additional masking could be attributed to the speech "filling in" the dips in the noise, which would lead to an overall reduction in the modulations; this effect was estimated by Carhart *et al.* (1969a) to be approximately 3 dB. Carhart *et al.* (1969a) reported a similar experiment, differing in that it included unmodulated noise and a speech-to-noise ratio of approximately −3 dB; they observed 6.6 dB of additional masking beyond that expected due to differences in power.

The method of Carhart *et al.* (1968) was also used in one of two experiments that have studied the effect of age on informational masking. Tillman *et al.* (1973) found 3 dB of additional masking for younger listeners and 4 dB for the elderly listeners. When a second competing talker was added to the masker, elderly listeners showed 2 dB more informational masking than the young listeners. Taken together, these results suggest that elderly listeners may be more susceptible to the informational masking of speech, although only by about 1–2 dB. Another experiment that compared amounts of informational masking for young and elderly listeners is that of Li *et al.* (2004), who measured the *release* from informational masking, using a "precedence-effect" method similar to Freyman *et al.* (1999). They showed that elderly listeners had the same release from informational masking as younger listeners, and thus they inferred that the total amount of informational masking, before any was released, was the same for elderly and younger listeners.

Many studies have measured the effect of competing speech for hearing-impaired listeners (e.g., Tun and Wingfield, 1999; Tun *et al.*, 2002; Hornsby *et al.*, 2006; Humes *et al.*, 2006). These experiments were mostly not designed to allow a clear distinction between energetic masking and informational masking, however, but one important study by Arbogast *et al.* (2005) did explicitly measure informational masking. They used pure-tone vocoded sentences (Arbogast *et al.*, 2002), with the target speech and competing speech formed from different sets of modulated pure tones so that their mutual energetic masking was minimized. Arbogast *et al.* (2002) found that hearing-impaired listeners had much higher SRTs for the target speech in noise (which had the same long-term average spectrum as the pure-tone vocoded competing speech), and therefore less of their speech-on-speech masking was attributable to informational masking.

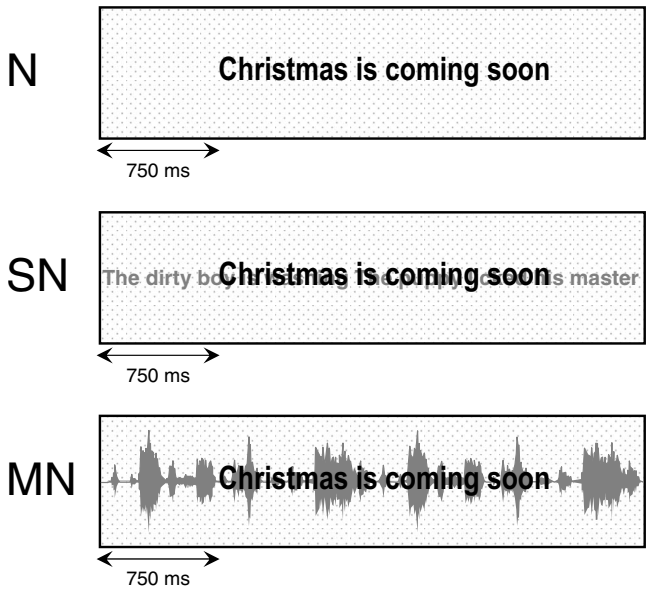


FIG. 1. A schematic illustration of the three stimulus conditions of experiment 1. The open black box represents the steady noise present in each condition. The text *Christmas is coming soon* represents the target speech, which starts 750 ms after the steady noise in each condition. In the SN condition (middle panel) the gray text represents the competing speech ("The dirty boy was washing the puppy licked his master"). In the MN condition (bottom panel) the gray waveform represents the modulated noise.

That is, their hearing-impaired listeners showed less informational masking than their normal-hearing listeners.

Taken together, these results are contrasting: those of Li *et al.* (2004) and Arbogast *et al.* (2005) imply that informational masking is no larger in elderly or hearing-impaired listeners than in younger normal-hearing listeners—and may even be less—but the results from Tillman *et al.* (1973) would suggest the opposite. The aim of the present experiments was therefore to compare the amount of informational masking observed in these groups and to further develop methodologies for comparing young and elderly listeners' susceptibility to informational masking of speech by speech. Experiments 1 and 2 used a "speech-in-speech-in-noise" design related to Carhart *et al.* (1968, 1969a), whereas experiment 3 used a design inspired by Arbogast *et al.* (2002, 2005). We attempted to increase the amount of informational masking obtained by having target- and competing-speech sentences that were similar to each other: they both used recordings from the same talker and were presented diotically. Thus the experiments omit some cues for distinguishing simultaneous sentences that would often be available in everyday conversation, such as talker differences and localization cues (Freyman *et al.*, 1999; Brungart, 2001). Unlike many informational-masking experiments, however, the target sentence started *after* the competing-speech sentence; this was done to give the listeners some marker as to which sentence to report from the mixture.

II. EXPERIMENT 1

Experiment 1 used the speech-in-speech-in-noise design to measure informational masking. The conditions are summarized in Fig. 1. In the baseline condition ("N"), SRTs for a target sentence (here "Christmas is coming soon") were measured against a masker consisting of a static noise alone (represented by the stippled box). In the experimental condition ("SN"), SRTs were measured against a masker made from both competing speech ("The dirty boy was washing the puppy licked his master") and a static noise. Note that the target sentence started 750 ms after the first of the competing sentences. In a control condition ("MN"), SRTs were measured against a masker made from both a modulated noise (shown by the gray waveform) and the static noise.

Both the modulated noise and the static noise had the same long-term average spectra as the target speech and competing speech. The modulated noise was also designed to be equivalent to the competing speech in terms of its overall fluctuations in power: it was constructed using a one-band vocoder, so lacked the spectral variations in the competing speech and was not recognizable as speech. The level of the competing speech and modulated noise was set to be 4 dB less than the level of the static noise, i.e., a long-term average speech-to-noise ratio of −4 dB. This value was chosen as a compromise between the requirement to minimize the level of the competing speech (so that it added only an insubstantial amount of extra acoustic power to the main noise), and the requirement that the competing speech should be intense enough to be intelligible (in order to generate some informational masking).

Given that the speech and modulated noise were presented at a long-term average speech-to-noise ratio of −4 dB, a simple calculation shows that they would be expected to lead to an increase in power of 1.5 dB over the power of the static noise, on average. The energetic masking in the SN and MN conditions should therefore be 1.5 dB higher than in the N condition. That is, if energetic masking was the only factor affecting SRTs in all three conditions, then the predictions for the SRTs are

$$\mathrm{SRT_{SN}} = \mathrm{SRT_{MN}} = \mathrm{SRT_N} + 1.5, \tag{4}$$

where the subscripts refer to the condition. But if the competing speech did indeed add some informational masking (here termed I), then

$$\mathrm{SRT_{SN}} = \mathrm{SRT_{MN}} + I = \mathrm{SRT_N} + 1.5 + I. \tag{5}$$

Thus, a comparison of the SRT in the SN and MN conditions enables the amount of informational masking to be calculated.

A. Stimuli

The stimuli were generated from combinations of target speech, static noise, competing speech, and modulated noise. The *target speech* was a sentence taken from the Audiovisual Sentence List (ASL) (MacLeod and Summerfield, 1990) spoken by a male talker. Each sentence had a simple syntactic structure, with three keywords for scoring (e.g., *the mother cooked the dinner*). Following MacLeod and Summerfield (1990) the full set of 270 sentences was divided into sets of 15 sentences. The *competing speech* was a pair of ASL sentences, concatenated without gaps. The pairs of sentences were selected pseudo-randomly so that they would be different from each other and different from the target sentence.

Note that the target speech and competing speech were spoken by the same talker. The *static noise* was a speech-shaped, unmodulated noise. It was constructed by cutting a random section (with 10-ms raised-cosine gates) from a 15-s burst of speech-shaped noise, which was in turn created by an inverse fast Fourier transform of the average spectrum of all the ASL sentences after randomizing the phases. Thus the static noise had the same long-term average spectrum as the target and competing speech. The *modulated noise* was formed from the same speech-shaped noise but was modulated by the envelope of the competing speech (cf. Festen and Plomp, 1990). This envelope was calculated by extracting its instantaneous amplitude (via the Hilbert transform; e.g., Hartmann, 1998) which was low-pass filtered with a cut-off frequency of 32 Hz (with a first-order Butterworth filter with 3-dB/octave slopes). All stimuli were digitally processed at a sample rate of 44.1 kHz.

The maskers were combined into three conditions, termed N, SN, and MN. The N condition consisted of the target speech and static noise alone; in the SN condition, competing speech was added; the MN condition paralleled the SN condition, but modulated noise was added instead of competing speech.

The competing-speech or modulated-noise pairs always started at the same time as the static noise, but the target speech started 750 ms later. The choice of 750 ms was made because it was approximately half the duration of the first of the competing sentences so that the target speech started halfway through the first competing-speech sentence. Thus, in the SN condition, the target speech could be distinguished as the second of three sentences. A short warning tone was presented 750 ms before the onset of the static noise. The static noise was presented at an overall level of 60-dB SPL, and, as noted, the competing speech and the modulated noise were presented 4 dB lower.

After some training, all listeners were able to correctly identify the target speech in competing speech in the absence of static noise, except for three of the most hearing-impaired listeners, who reported that the simultaneous talkers were "jumbled up." This inability to distinguish two talkers at equal levels was interpreted as a raised speech-reception threshold, rather than lack of training, because they were able to follow all other instructions as successfully as the other listeners, and they were all at the higher end of hearing impairment for this study.

Due to the limited availability of suitable experimental spaces, the experiment was run in three different locations: a sound-treated booth and two separate quiet offices. Although these locations had different equipment and different levels of background noise, tests showed that the choice of location did not affect speech recognition in the N condition, so it was assumed that the SN and MN conditions would also be unaffected by the choice of location. Three sets of equipment were used for experimental control and sound generation: in the sound-treated booth, an RME DIGI96/8 PAD soundcard was used with an Arcam A80 amplifier and Sennheiser HD580 Precision headphones, whereas in both offices AudioCapture UA-5 USB audio interfaces were used with EDIROL Audio Capture UA-5 D/A converters, SAMSON C-Que 8 headphone amplifiers, and Sennheiser HD580 Precision headphones. In the sound-treated booth, listeners were able to speak to the experimenter via a microphone; in the offices, the experimenter and listener were in the same room and could speak directly to each other.

B. Procedure

Six-point psychometric functions were measured for speech reception (i.e., the identification of the target speech) as a function of its level relative to that of the masking noise. All stimuli were presented diotically. Listeners were asked to identify the target speech in the stimulus as the second of three sentences heard, and repeat it to the experimenter. The number of keywords accurately repeated was scored, accepting as correct any responses that were homonyms and/or had the same stem as the keyword (MacLeod and Summerfield, 1990). Listeners were encouraged to repeat the sentence they thought they heard, even if it was little more than a guess and even if they thought it made little sense. The levels of the target speech for the six points in each psychometric function were chosen individually on the basis of training data, but always covered a range of 10 dB at 2-dB intervals. The aim was for few words to be correctly identified at the lowest target level, yet for most words to be correctly identified at the highest target level. If necessary, the levels were revised between blocks of trials.

The experiment was run in two sessions, each with nine blocks of trials. A block consisted of 15 trials at each of the six levels. The resulting 90 trials in a block were ordered randomly. The target-speech sentences used in a set of 15 trials were chosen at random, without replacement, from one of the ASL lists of 15 sentences. Within each block, a different list was used for each point of each psychometric function. The type of masker—N, SN, or MN—was fixed within a block but varied across blocks. Three blocks were run for each condition; thus there were 45 trials per point for each listener. This resulted in 135 keywords per point since there are three keywords in each ASL sentence.

A 45-min structured training session preceded data collection in the first session, in which listeners practiced responding to stimuli similar to those used in the experiment. Training began with the N condition, followed by the MN condition, then the SN condition with the competing speech spoken by a female talker, to make it easier to identify the target, and finally the SN condition with both the target and competing speech spoken by the same male talker. At each stage the signal-to-noise ratio of the target was set to a relatively easy level, and then progressively reduced. The training used the Bench–Kowal–Bamford (BKB) sentences (Bench and Bamford, 1979) instead of the ASL sentences for the target and competing speech in order to avoid any memorization of the sentences. The BKB sentences have comparable durations, syntaxes, semantic content, and scoring methods to the ASL sentences.

C. Analysis

An analytic function was fitted to the data using a standard least-squares method to determine quantitatively the

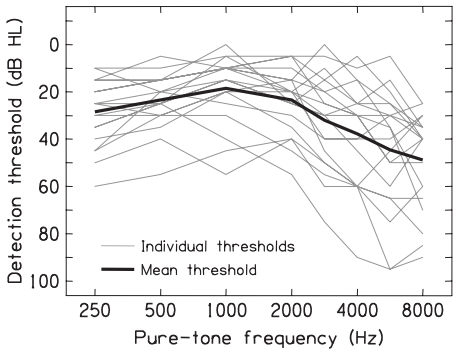


FIG. 2. The pure-tone audiograms of the 20 elderly listeners' better ears in experiment 1 (gray lines) and the corresponding mean audiogram (thick black line).

50% SRT (L_{SRT}, in decibels) and the peak gradient (m, in %/dB) of the psychometric function. The function was a two-parameter cumulative-normal function:

$$W = 100\Phi\left(\frac{\sqrt{2\pi}}{100}m(L - L_{SRT})\right), \quad (6)$$

where L is the level of the target sentence (decibels), W is the percentage of keywords correctly identified at that level, and $\Phi(z)$ is the cumulative standard normal distribution.[2]

D. Participants

Eight young and 20 elderly listeners took part in the experiment. Seven of the young listeners were students at the University of Strathclyde, Glasgow, and the other was a staff member at the MRC Institute of Hearing Research (IHR). All of the young listeners had self-reported normal hearing. The 20 elderly listeners were volunteers who had previously agreed to take part in research at IHR. They were originally recruited from hearing clinics, the electoral register, and other sources. Figure 2 shows the elderly listeners' audiograms. Their better-ear average ("BEA;" defined as the mean of their audiometric thresholds at 0.5, 1, 2, and 4 kHz) ranged from 6 to 53-dB hearing level (HL), and most of their audiometric functions showed some degree of the high-frequency hearing loss associated with presbycusis. The elderly listeners' BEA was correlated with their age [$r(18) = 0.54$, $p = 0.01$].

E. Results

Figure 3 shows the psychometric functions of each listener for each of the three conditions, with the elderly listeners ordered in terms of their hearing impairment. There was the expected overall effect of hearing loss, in that the psychometric functions of those with greater hearing impairments were generally shifted further to the right than those with lesser hearing impairments. For everyone, the leftmost psychometric function was that of the N condition (circles/solid lines); this was expected as the N condition had the least power of any of the maskers, by 1.5 dB. More importantly, the rightmost psychometric function for *all* listeners was the SN condition, with the MN condition in the middle. That is, there was more total masking in the SN condition than the MN condition. This result is consistent with the prediction that there would be informational masking in the SN condition that is not present in the MN condition.

Figure 4 shows the SRTs [derived from Eq. (6)] for the N condition (top panel), the difference between SRTs in the MN and N conditions (middle panel), and the difference between SRTs in the SN and MN conditions (bottom panel). Note that the scales of the ordinates are the same in all three panels, despite their different ranges. There was a strong positive relationship between SRT_N and hearing impairment [top panel; $r(18) = 0.902$, $p < 0.001$]. But there was no relationship between the $SRT_{MN} - SRT_N$ difference and hearing impairment [middle panel; $r(18) = -0.03$, $p = 0.91$]; across listeners, SRT_{MN} was between 0.9 and 2.8 dB higher than SRT_N, with a mean value of 1.7 dB (SD=0.5 dB). This mean value was close to the predicted value of 1.5 dB given the difference in power of the maskers. Finally, there was a clear negative correlation between $SRT_{SN} - SRT_{MN}$ and hearing impairment [bottom panel: $r(18) = -0.78$, $p < 0.001$]. Across listeners, SRT_{SN} was between 1.0 and 4.6 dB higher than SRT_{MN}, with a mean value of 3.2 dB (SD=0.9 dB). Given the design of the experiment, this $SRT_{SN} - SRT_{MN}$ difference of, on average, 3.2 dB is argued to represent the amount of informational masking due to the competing speech. The negative correlation between it and hearing impairment—which accounted for over half of the variance in the data ($r^2 = 0.61$)—indicates that the listeners with poorer audiometric thresholds showed less informational masking. The effect was relatively small in terms of decibels; however, the values of the $SRT_{SN} - SRT_{MN}$ difference calculated from the regression line at BEAs of 10 and 50 dB were, respectively, 4 and 2 dB.

The effect of age is shown in Fig. 5. The mean $SRT_{MN} - SRT_N$ difference (left pair of bars) was near identical across age group and was statistically insignificant [$t(26) = -0.33$, $p = 0.75$]. The $SRT_{SN} - SRT_{MN}$ difference (right pair of bars) differed by 0.5 dB across age group, with the younger group giving the higher difference. This difference did not quite reach statistical significance across age group [$t(23.52) = 1.97$, $p = 0.06$], although the power of the statistical test was undoubtedly reduced by the relatively large variability for the older group.

Figure 6 shows the individual SRTs from Fig. 4, but plotted against the listeners' ages (the plotted regression lines, and the correlations reported below, are for the older group only). For SRT_N (top panel), there was a trend for older listeners to have higher thresholds than the younger elderly listeners, as would be expected from Fig. 4 given the correlation between age and BEA. However, the scatter in the data was relatively high and the correlation was not significant [$r(18) = 0.38$, $p = 0.10$)]. There was, however, an effect of age on the $SRT_{MN} - SRT_N$ difference [top panel; $r(18) = 0.45$, $p = 0.05$], with the oldest listeners showing a greater $SRT_{MN} - SRT_N$ difference, although the differences observed were less than 2 dB. There was no significant effect of age amongst the elderly listeners on the amount of informational masking, $SRT_{SN} - SRT_{MN}$ [bottom panel; $r(18) = -0.37$, $p = 0.11$].

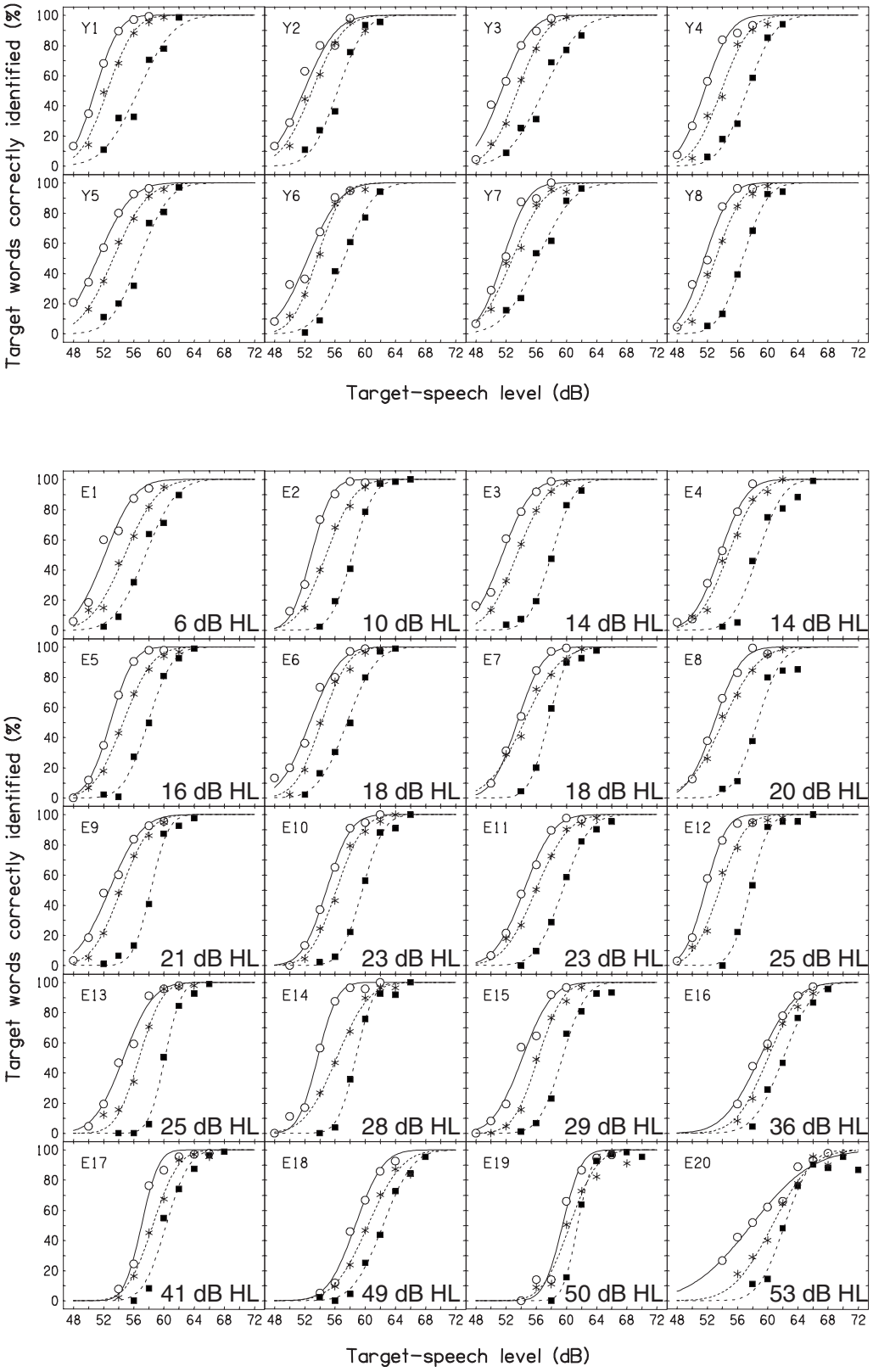


FIG. 3. The psychometric functions measured in experiment 1 for all young listeners (top panels) and elderly listeners (bottom panels). In each panel the symbols show the measured data and the lines show the best-fitting cumulative-normal functions [Eq. (6)]. The parameter is condition: N: open circles and solid lines; MN: asterisks and dashed lines; SN: filled squares and dashed lines. The elderly listeners are ordered by their BEAs, which are shown in the bottom-right corner of each panel.

F. Discussion

Speech-reception thresholds were measured in three conditions of masking: in static noise alone (termed N), in competing speech combined with static noise (SN), and in modulated noise combined with static noise (MN). It was found that the SRTs in the MN condition were, on average,

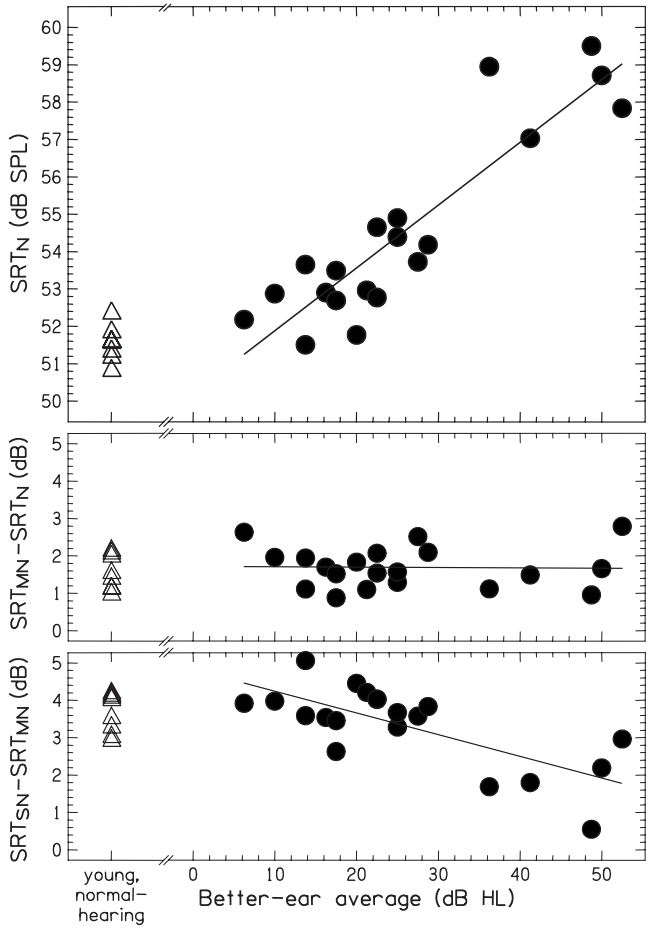


FIG. 4. The derived SRTs and differences in SRTs measured in experiment 1. The top panel shows the speech-in-noise thresholds (SRT_N) for each listener; the middle panel, the effect on SRT of adding modulated noise ($SRT_{MN}-SRT_N$), and the bottom panel, the amount of informational masking ($SRT_{SN}-SRT_{MN}$). The regression lines are fitted through the elderly listeners' data only (filled circles).

1.7 dB greater than in the N condition. This difference can be attributed to energetic masking, as it closely matches the prediction of 1.5 dB made on the basis of the difference in powers of the maskers. It was also found that the SRTs in the SN condition were in turn 3.2 dB greater, on average, than in the MN condition. This difference was attributed to informational masking from the competing speech. Although this was small in terms of decibels, it was large in terms of

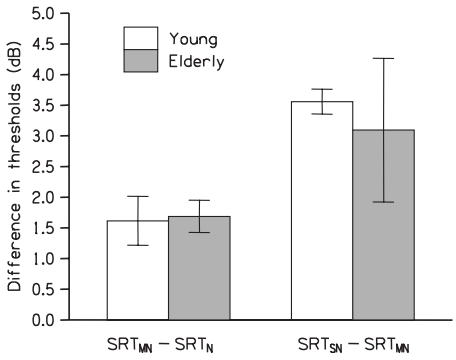


FIG. 5. The across-age-group averages of $SRT_{MN}-SRT_N$ and $SRT_{SN}-SRT_{MN}$ measured in experiment 1. The error bars are 95%-confidence intervals.

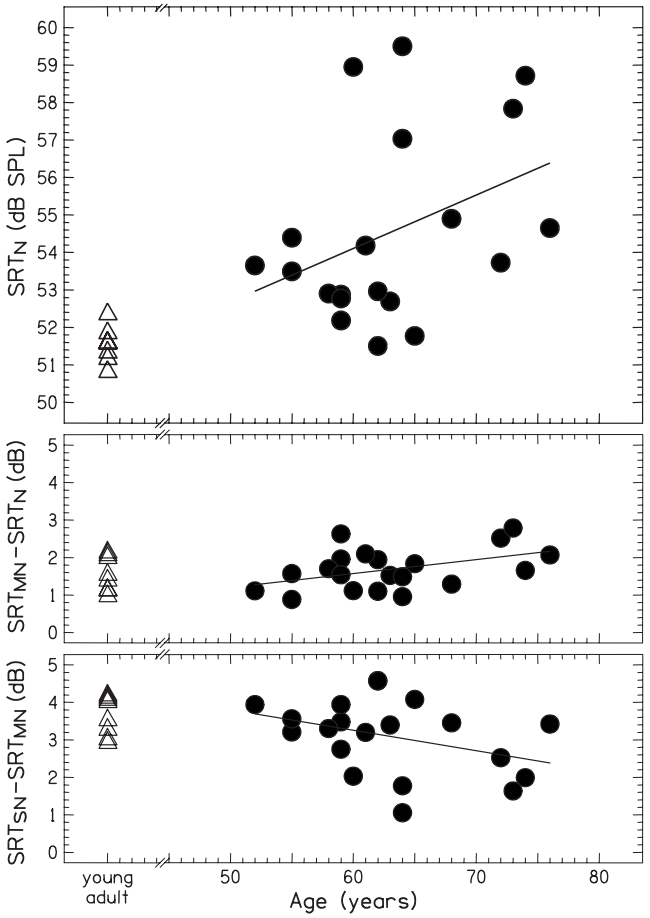


FIG. 6. As Fig. 4, but with the data plotted against age instead of BEA hearing loss.

speech recognition, given that the slope of the psychometric functions was about 10%–20%/dB in this experiment.

An average effect of about 3 dB is at the smaller end of the range of informational-masking effects reported by other researchers (Carhart *et al.*, 1968, 1969a; Freyman *et al.*, 1999; Brungart, 2001; Li *et al.*, 2004; Wu *et al.*, 2004). Of these, Carhart *et al.* (1968) deserves further discussion, as their design was similar to the current experiment although (1) they used a combined masker of speech and modulated noise, not speech and static noise and (2) they used equal-level competing speech and noise masker,[3] instead of reducing the level of the speech in order to minimize its contribution to the energetic masking. They found a difference of 7.8 dB between SRTs for their equivalents of the N and SN conditions. Of this, they attributed 3 dB of the effect to the addition of the competing speech, leaving 4.8 dB for informational masking. This estimate of the amount of informational masking is similar to that observed in the present experiment, despite the differences in sentence materials, choice of noise, and measurement of levels. The issue of the level of the competing speech relative to that of the noise is considered in experiment 2 below.

Attempted quantifications of informational masking have been found to be remarkably dependent on the exact experimental conditions used, and it must be remembered that our results are, strictly, only applicable to the present speech-in-speech-in-noise design. First, the presence of the

noise used in the SN condition could itself have affected the amount of informational masking generated by the competing speech: for example, Freyman *et al.* (2001) showed that adding a noise at the same location as the competing talker could reduce its informational masking, and Kidd *et al.* (2005) showed that adding a noise at the same frequencies as a competing talker reduced its informational masking. Second, there could be elements of informational masking induced by the noises themselves: for example, Kwon and Turner (2001) argued that any modulation of maskers could mask speech, which could be considered informational masking, and Lutfi (1990) argued that even masking by an unmodulated noise could include some informational masking due to uncertainty. Third, the amount of informational masking of speech depends on the differences the experimenter allows between the target speech and the competing speech (e.g., Carhart *et al.*, 1969b; Freyman *et al.*, 1999; Brungart, 2001) and may depend on which speech materials are used. It is a wider question as to what laboratory experiments would best represent real-life listening and whether competing speech should be mixed with some other form of masking sound, but in many ways the "worst-case" condition, though experimentally popular, may be unrealistic. It would likely use the same talker for the target and competing speech (or even multiple sources of competing speech; Brungart *et al.*, 2001; Freyman *et al.*, 2001), presented from the same location and at the same level, timed so that they start simultaneously, and would include confusable keywords that are themselves concurrent. In contrast, the antithetical condition to this—with different-gender talkers, from different locations, at different levels, with clear timing differences, and differing syntax and semantics—is likely to give less informational masking, but may also be more representative of real life.

The effects in this experiment were only weakly affected by age. The $SRT_{MN}-SRT_N$ difference did not differ across age group (Fig. 5, left panel), although within the older group it did increase slightly with age, at a rate of 0.37 dB/decade (Fig. 6, middle panel). In contrast, the $SRT_{SN}-SRT_{MN}$ difference was smaller for the older group, and within the older group it decreased at a rate of 0.55 dB/decade. But neither effect reached statistical significance at the usual $p=0.05$ level, which we attribute to the relatively large scatter in the relevant data points. Although suggestive then that informational masking reduces with age, the results remain somewhat inconclusive.

The effect of hearing impairment was clearer; however, the hearing-impaired elderly listeners were less susceptible to informational masking than the normal-hearing listeners. This result supports that of Arbogast *et al.* (2005), who also observed less informational masking for hearing-impaired listeners. They suggested that this effect may not necessarily be due to a fundamental difference in hearing-impaired listeners' susceptibility, but could instead be related to the design of the experiment. First, hearing-impaired listeners may have been less able to hear the competing speech due to the greater effects of energetic masking, and so it would generate less informational masking. Second, there may be a ceiling effect, such that there is less informational masking when the target speech is at a higher level than the competing speech (see also Brungart, 2001). Both of these applied to our experiment, as (1) we presented the *competing* speech at the same speech-to-noise ratio for both normal-hearing and hearing-impaired listeners, but not the same sensation level, and so the intelligibility of the competing speech may well have been reduced, but (2) we presented the *target* speech at a higher level to the hearing-impaired listeners, as otherwise they would not have reached the 50% criterion for the speech-reception thresholds. Experiment 3 below was, in part, designed to avoid this potential criticism, whereas experiment 2 explored the effect of competing-speech level for normal-hearing listeners.

III. EXPERIMENT 2

Experiment 2 again used the speech-in-speech-in-noise design but measured performance across a range of speech-to-noise ratios (−16 to +4 dB) rather than at just −4 dB. In each of these, the competing speech in the SN condition was predicted to add some informational masking and some additional energetic masking relative to the N condition, but it would be expected that both effects would be minimal when the competing speech was near its own threshold and both effects would be larger when it was far more intense. Also, to check the assumption that the additional energetic masking was the same as the increase in power due to the addition of competing speech, corresponding MN conditions were tested, and the N condition was included as a baseline.

A. Procedure

The stimuli were constructed in the same way as those of experiment 1. In the SN conditions, nine competing-speech levels were used, varied parametrically across blocks at speech-to-noise ratios of −16, −10, −8, −6, −4, −2, 0, +2, and +4 dB. In the MN conditions, five modulated-noise levels were used at modulated-noise-to-static-noise ratios of −10, −4, 0, +4, and +8 dB (these will be referred to below as "noise-to-noise ratios" to simplify the text). Finally, the N condition was the same as in experiment 1, with neither competing speech nor modulated noise. The target speech was again varied parametrically within each block, with target-speech-to-noise levels of −10, −8, −6, −4, −2, and 0 dB.

B. Results and discussion

The procedures and equipment were generally the same as for experiment 1, but rather than repeating the target sentence to the experimenter, the participants typed their responses. The responses were marked by the experimenter off-line after the experiment. There were again three blocks of 90 trials for each condition.

The SN and MN conditions were tested in two separate procedures. For the SN conditions, there were nine competing-speech levels which, with the N condition, resulted in a total of 30 blocks for each listener. The listeners took between four and ten sessions to complete the 30 blocks: the durations of the sessions were based on each listener's preference. Four young listeners participated, aged 24, 24, 32, and 35 years old. Three were students at the

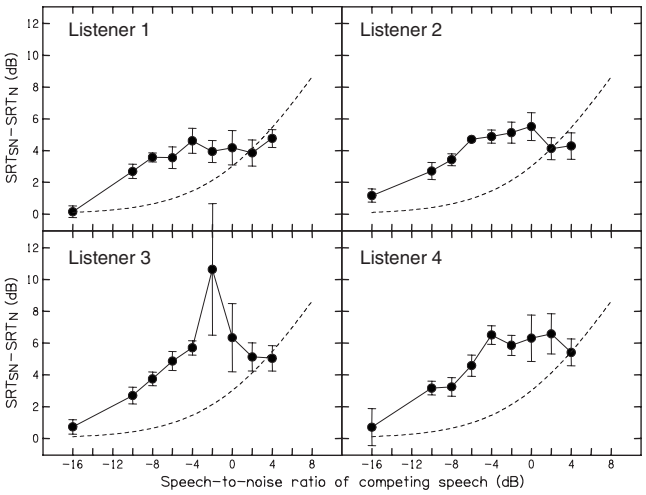


FIG. 7. The SRTs measured in the SN conditions of experiment 2. In each panel the symbols show the measured data and the lines show the predictions given by energetic masking alone, based on the long-term average power of the total masker. The four panels are for the four listeners. The error bars are 95%-confidence intervals.

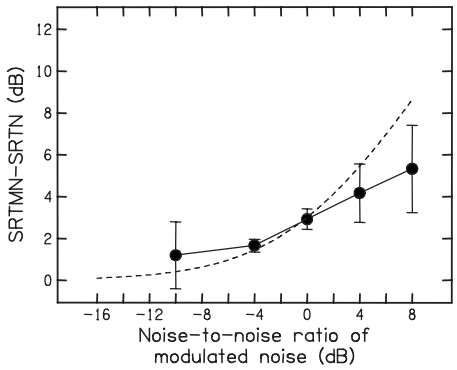


FIG. 8. As Fig. 7, but for the MN conditions of experiment 2. The data are averaged across the listeners. The error bars are 95%-confidence intervals.

University of Strathclyde, Glasgow, and the other was a staff member at the MRC Institute of Hearing Research. For the MN conditions, there were five modulated-noise levels. There were three listeners, who completed the resulting 18 blocks (including the N condition) in three to five sessions each. All three participants were staff members of the MRC Institute of Hearing Research. One also completed the SN conditions of the current experiment, and one was the first author. All listeners in this experiment had self-reported normal hearing.[4]

Figure 7 shows the results for each listener, plotted as the $SRT_{SN}-SRT_N$ difference (the values of SRT_N were −9.6, −8.5, −8.2, and −10.3 dB for the four listeners). The SRT difference gradually increased as the speech-to-noise ratio was increased, but then remained approximately constant for the highest ratios.[5] The dashed lines show the expected effect if the masking was due *solely* to the total power of the combined masker (e.g., 0.6 dB at −8 dB, 1.5 dB at a speech-to-noise ratio of −4 dB, or 3 dB at a speech-to-noise ratio of 0 dB, and so on). The results—except at the extremes of the range—were generally considerably higher than expected from this model.

Figure 8 shows the corresponding results from the MN conditions, plotted as the $SRT_{MN}-SRT_N$ difference. The results are averaged across all three listeners as they responded similarly. For noise-to-noise ratios up to 0 dB, the results were broadly similar to those predicted from the increases in the overall power of the combined masker (dashed line): thus the assumption that the increase in energetic masking was equal to the increase in power was validated. But for higher noise-to-noise ratios, however, the SRTs were *lower* than predicted. We believe that this effect was due to listeners being able to take advantage of the dips in the modulated noise (e.g., Howard-Jones and Rosen, 1993; Cooke, 2006); note that this argument would not be expected to apply at lower noise-to-noise ratios, as there the static noise would have been considerably more effective at filling in the dips of the modulated noise.

The results add to those of experiment 1 in clearly demonstrating that the presence of the competing speech gives additional masking, beyond that which would be expected from its energy, of up to 3–5 dB, depending on the listener. The informational masking at a given speech-to-noise ratio can be estimated as the difference between the values shown in Fig. 7 and those in Fig. 8. The largest amount of informational masking was observed for competing-speech-to-noise ratios in the region of −6 to −4 dB, suggesting that the choice of −4 dB in experiment 1—which was made so that the competing speech would only add a small amount of energetic masking while still being intelligible—was about optimal. Up to this point, the amount of informational masking increased with increasing levels of competing speech, suggesting that in this range the concomitant increase in the intelligibility of the competing speech may directly affect the amount of informational masking obtained. For speech-to-noise ratios greater than −4 dB, however, SRT_{SN} did not increase but SRT_{MN} did. Thus, at higher speech-to-noise ratios, the amount of informational masking (the difference) declined, despite the (presumably) continued intelligibility of the competing speech, to the extent that there was effectively no informational masking at a speech-to-noise ratio of +4 dB. The intelligibility of the competing speech cannot therefore be the *sole* determinant of informational masking.

Rather, the effects are consistent with the notion that informational masking was exacerbated by the similarity of target-speech and competing-speech intensities, suggesting an element of confusion between target speech and competing speech (e.g., Brungart *et al.*, 2001). This has been previously demonstrated by analyzing listeners' errors and counting the words from the competing speech reported instead of those of the target speech: Brungart (2001) and Arbogast *et al.* (2002) found that up to 90% of the incorrect responses involved keywords from the competing speech, although both of their experiments used the coordinate response measure (CRM) speech corpus (Bolia *et al.*, 2000), in which the set of possible responses is extremely limited. In contrast, a corresponding analysis of the present experiment (which used ASL sentences) showed that no more than 4% of the competing-speech keywords were reported at any combination of target-speech and competing-speech levels. Thus our listeners only occasionally confused the competing speech with the target speech to the extent that they would report the

wrong keywords. It is likely that part of this reluctance to confuse the target and competing speech is because the present ASL materials can be distinguished, at least in part, by their wide vocabulary and non-identical sentence structures, whereas a simple confusion between target and competing words can play a role in "closed-set" designs such as the CRM task, in which the keywords can be freely confused. If confusion contributes to informational masking for the ASL materials, this must occur at the acoustic or phonemic level rather than at the word level. Whatever the exact reason why the amount of informational masking varies with the level differences between two talkers, the effect leads to potential confounds for many designs when comparing different groups' susceptibility to informational masking. To avoid this, the final experiment used a design in which the target and competing sentences were always presented at the same level.

IV. EXPERIMENT 3

Both experiments 1 and 2 used the speech-in-speech-in-noise design to reduce the energetic contribution of the competing speech while still allowing some informational masking. In this experiment, the energetic contribution from the competing speech was further reduced by separating it in frequency from the target speech, using a method inspired by Arbogast *et al.* (2002). They processed CRM sentences using a 15-band pure-tone vocoder and allocated seven (randomly selected) bands to the target speech and seven others to the competing speech. Arbogast *et al.* (2002) found that the SRT in this condition was, on average, 22 dB greater than the SRT in an acoustically equivalent noise and attributed the difference to informational masking. A similar method was adopted here, except that (1) the stimuli here were filtered by a set of narrow bandpass filters instead of applying a pure-tone vocoder (to help preserve some of the natural speech cues) and (2) similarly filtered masking noises were added as well. The design is summarized in Fig. 9; note that the target speech or competing speech used alternate frequency bands (first and second panels). The masking noise underlying the target speech—termed the "target noise"—was made from the same bands as the target speech (third panel), and likewise for the "competing noise" (bottom panel). These four stimuli were then combined in various ways to make four conditions termed "N," "N+," "SN," and "SN+" (see Table I).

The target speech, target noise, and competing noise were present in all the conditions. In the SN and SN+ conditions, the competing speech was also present (in order to introduce informational masking) and the amount of informational masking was calculated as the difference between the SRTs with and without competing speech. In the N+ and SN+ conditions the competing noise was set to be 14 dB more intense than in the N and SN conditions. This was done in order to estimate any contribution to energetic masking from "cross-excitation" of the other set of frequency bands: because (1) the filtering was done with sharp, but not infinitely sharp, bandpass filters, and (2) listeners' auditory filters are not infinitely sharp either, so some overlap in excitation would be expected between bands. The increase in noise level in the N+ and SN+ conditions meant that the combined power of the competing-speech-plus-noise masker was much larger in the SN+ condition than in the SN condition. That is, if masking was entirely due to the power in the competing channel, without any effect of informational masking, then the SRT would be largest in the SN+ condition.

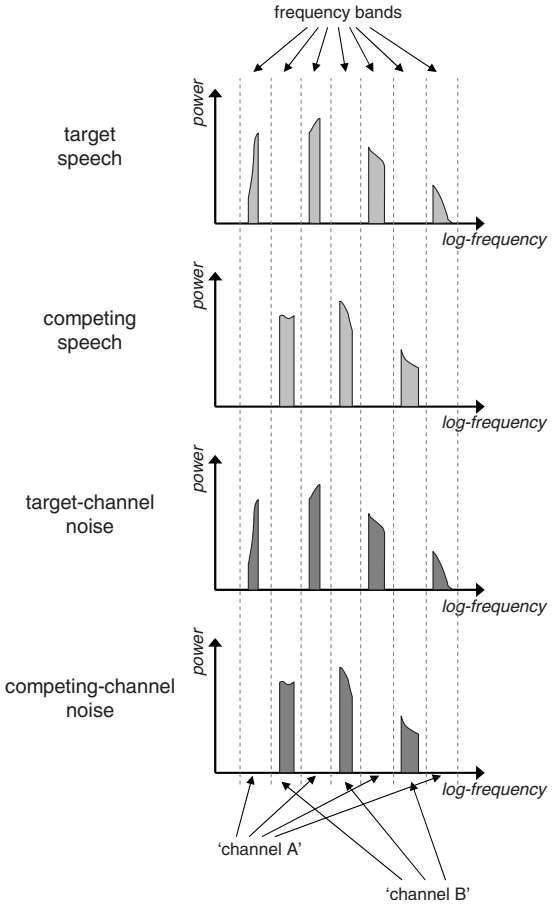


FIG. 9. Illustrative spectra of the stimuli used in experiment 3. A target sentence was filtered into seven frequency bands, of which the first, third, fifth, and seventh were kept and the rest removed (top panel). A competing sentence was similarly filtered but the opposing bands kept (second panel). Two speech-shaped noises were also filtered to form a noise masking the target speech (third panel) and a noise masking the competing speech (fourth panel). All four were then combined into one stimulus for presentation. Not illustrated is a complementary set of stimuli, in which the frequency bands used for the target and competitor were swapped.

TABLE I. The levels of the stimuli used in experiment 3. The levels of the target noise were varied to form the psychometric function. For the young listeners, the levels ranged between 50 and 60 dB SPL, at 2-dB intervals; for the elderly listeners, the levels were chosen individually on the basis of training data, but again always covered a range of 10 dB at 2-dB intervals.

	Levels (dB SPL)			
Label	Target speech	Competing speech	Target noise	Competing noise
N	52	···	Varied	52
SN	52	52	Varied	52
N+	52	···	Varied	66
SN+	52	52	Varied	66

It was noted earlier that it was likely that the relative intensities of the target speech and competing speech could affect the amount of informational masking. To control for this, the levels of target and competing speech were fixed to be the same throughout the current experiment, and psychometric functions were generated instead by varying the level of the masking noises.

A. Stimuli

The target speech and competing speech were processed versions of the ASL corpus (MacLeod and Summerfield, 1990), selected as for experiment 1; the competing speech was again formed by concatenating two of the ASL sentences, and the target speech started 750 ms after the start of the first competing sentence. The masking noises were always static and initially generated so that they matched the long-term spectra of the speech.

All the stimuli were filtered into seven non-overlapping frequency bands centered on 100, 212, 424, 849, 1697, 3394, and 6788 Hz. Each band was half an octave wide, except for the lowest one, which was widened to 50 Hz (0.74 octaves) to facilitate the design of an appropriate filter. The filters were fourth-order bandpass Butterworth filters with 12-dB/octave slopes, each with less than a 15-ms group delay. Neighboring frequency bands were each separated by half an octave. This separation was designed to minimize the energetic masking of each frequency band by its neighboring bands. Across trials, the bands used for the target speech and competing speech were randomly chosen: either the first, third, fifth, and seventh to the target speech and the second, fourth, and sixth to the competing speech (see Fig. 9) or vice versa (not illustrated). Table I reports the combinations of stimuli used and their levels. Note that these values reported in the table are the long-term average levels of the speech and noise components *before* filtering; the filtering process meant that the total power of what was in the first, third, fifth, and seventh bands was reduced by 5.1 dB, and what was in the second, fourth, and sixth bands was reduced by 9.6 dB. This does not affect the SRTs reported below, which were calculated as the difference in the power of the target speech at threshold and the power of the target noise alone.

B. Procedure

Psychometric functions of target-speech reception against target-noise level were measured. Listeners were asked to listen to the target sentence, which was always the second of three sentences: one of the competing sentences started before it and finished after it. Listeners were required to repeat the target sentence to the experimenter. Six levels of target noise were used to generate each psychometric function, chosen individually for each condition and each listener on the basis of training data. The scoring system, sentence randomization, training, and instructions to the participant paralleled experiment 1.

The experiment was presented in blocks of 90 trials, which lasted 10–15 min. Three such blocks (one per condition in a random order) were presented with short breaks. There were three repetitions of this procedure, spread over 3 days, preceded by a 30-min training session on the first day. Experimenter errors led to three blocks being discarded: for one young listener, the last block was run using the wrong conditions; for two elderly listeners, one psychometric function was measured over a different range to the other two blocks for the same condition. In all three cases, the erroneous blocks were omitted from the analysis, and the estimated performance was based on the average of just two blocks of trials for the three listeners in the affected conditions.

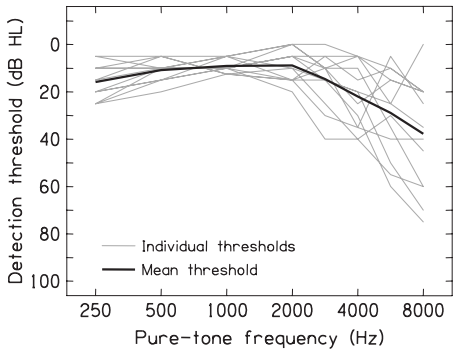


FIG. 10. The pure-tone audiograms of the 20 elderly listeners' better ears in experiment 1 (gray lines) and the corresponding mean audiogram (thick black line).

The stimuli were presented through a 24-bit D/A converter (an RME DIGI96/8 PAD soundcard) and an Arcam A80 amplifier. Participants listened through Sennheiser HD580 Precision headphones while in a sound-treated booth. Participants' responses were picked up by a microphone in the booth and presented over headphones to the experimenter outside the booth. The experimenter recorded the total number of target keywords correctly identified.

C. Participants

Eight young listeners and 13 elderly listeners participated in the experiment. Three of the young listeners were staff at the MRC Institute of Hearing Research and had previously taken part in experiments using the same speech corpus. The other five were recruited from the University of Strathclyde Psychology Department and had no previous experience with psychoacoustical experiments. The 13 elderly listeners were initially invited from population surveys based on the electoral register. They had participated in previous auditory experiments, but these experiments did not involve the ASL corpus. Figure 10 shows the elderly listeners' better-ear audiograms. Their mean age was 66 years (ranging from 51 to 80 years), and their mean BEA was 13-dB HL (ranging from 4- to 21-dB HL); thus most, if not all, would be classified as having normal hearing, though with some high-frequency hearing loss.

D. Results

Figure 11 shows the psychometric functions for six illustrative listeners. The four panels per listener are for the four conditions, and within each panel the two psychometric functions are for the case where the target speech used the even-numbered frequency bands (solid symbols) or where it

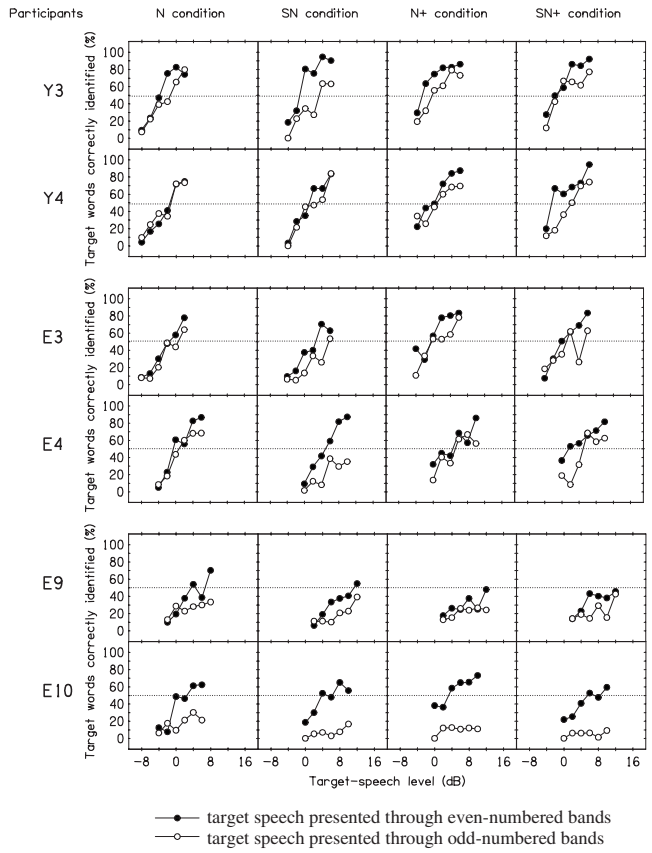


FIG. 11. The psychometric functions from experiment 3 for an illustrative selection of listeners (see text for choices). The conditions are ordered by column. In each panel one psychometric function is for when the target speech used the first, third, fifth, and seventh frequency bands, the other for when it used the second, fourth, and sixth bands.

used the odd-numbered bands (open symbols). For the young listeners (e.g., listeners Y3 and Y4 in Fig. 11), most of the psychometric functions were similar for the two target channels, and a good level of speech understanding was achieved at the higher speech-to-noise ratios. The same occurred in most of the elderly listeners (e.g., E3 and E4 in Fig. 11), but some others had relatively low scores for target speech in one or both of the target channels, not always reaching the 50% criterion for the speech-reception threshold (e.g., E9 and E10 in Fig. 11).

Figure 12 shows the SRTs derived from the psychometric functions via Eq. (6), plotted against hearing impairment (left panels) or age (right panels). The lines are regression lines and are calculated within the elderly group only. Note that some of the SRTs are extrapolations as performance never reached 50%. In addition, any extrapolated SRTs that were improbably large (i.e., derived from particularly shallow psychometric functions) have been capped at +12 dB. The mean SRTs for the younger group of listeners were −3.3, 1.0, −2.0, and −1.4 dB for the N, SN, N+, and SN+ conditions, respectively, and a repeated-measures ANOVA confirmed the overall effect of condition for these younger listeners [$F(3,21)=20.32$, $p<0.001$]. A comparison of the SN and N conditions showed that there was about 4 dB of informational masking ($SRT_{SN}-SRT_N=4.3$ dB). Moreover, the SN+ condition did not give the highest SRT despite having the highest overall power. In fact, raising the level of the competing noise in the presence of the competing speech actually reduced the SRT (i.e., $SRT_{SN+}-SRT_{SN}=-2.4$ dB), although on its own the change in level of the competing masker led to an increase in SRT of about 1 dB (i.e., $SRT_{N+}-SRT_N=1.3$ dB).

The calculations for the corresponding mean SRTs for the elderly listeners excluded four listeners who gave at least one capped SRT: without them, the means were 1.2, 6.2, 4.2, and 4.8 dB for the N, SN, N+, and SN+ conditions, respectively. Thus, the effect of informational masking ($SRT_{SN}-SRT_N$) was 5.0 dB, the effect of raising the competing noise on informational masking was −1.4 dB ($SRT_{SN+}-SRT_{SN}$), and the effect of increasing competing noise in the absence of competing speech was 3 dB ($SRT_{N+}-SRT_N$). All three differences were about 1–2 dB larger than the corresponding differences in the younger group. A mixed ANOVA on SRT_{SN}, SRT_{N+}, and SRT_{SN+}, each relative to SRT_N, confirmed an overall effect of age group [$F(1,15)=5.44$, $p<0.03$] and an effect of condition [$F(2,30)=18.01$, $p<0.001$]. The effect of age is supported by inspection of the regression lines within the elderly group, all of which had a positive gradient. The ANOVA did not show an interaction between age group and condition [$F(2,30)=0.83$, $p=0.45$]. However, the amount of informational masking was not itself significantly different between the two age groups [$SRT_{SN}-SRT_N$, $t(15)=1.34$, $p=0.20$]. The effect of increasing the level of the competing-channel noise was also insignificant across age group [$SRT_{N+}-SRT_N$, $t(15)=1.83$, $p=0.09$], but when combined with competing speech it was significant [$SRT_{SN+}-SRT_N$, $t(15)=2.17$, $p=0.05$]. As noted, four of the elderly listeners were excluded from these calculations, as they gave at least one capped SRT. It is intriguing that these people were primarily the oldest and more impaired of the group: they were the oldest four, included four of the five highest hearing levels, and returned three of the four highest values of SRT_N.

E. Discussion

The difference in SRTs between the N and SN conditions was about 4–5 dB. Given that the only experimental difference between the two conditions was that there was competing speech present in the SN condition but not in the N condition, and assuming that its presence did not affect the energetic masking of the target speech because of the separation of frequency bands, then this difference in SRT can be attributed to informational masking.

We did not find any evidence that, on average, the amount of informational masking increased with age. Nevertheless, it will be noted that it was the oldest listeners who gave high SRTs that required capping, and as some of the capped SRTs were in the SN condition but none were in the N condition, then these listeners may well have been particularly susceptible to informational masking.

The amount of informational masking observed here was about 1–2 dB larger than that measured in experiment 1. This may have been because the target speech and competing speech were presented at the same intensity (the pre-

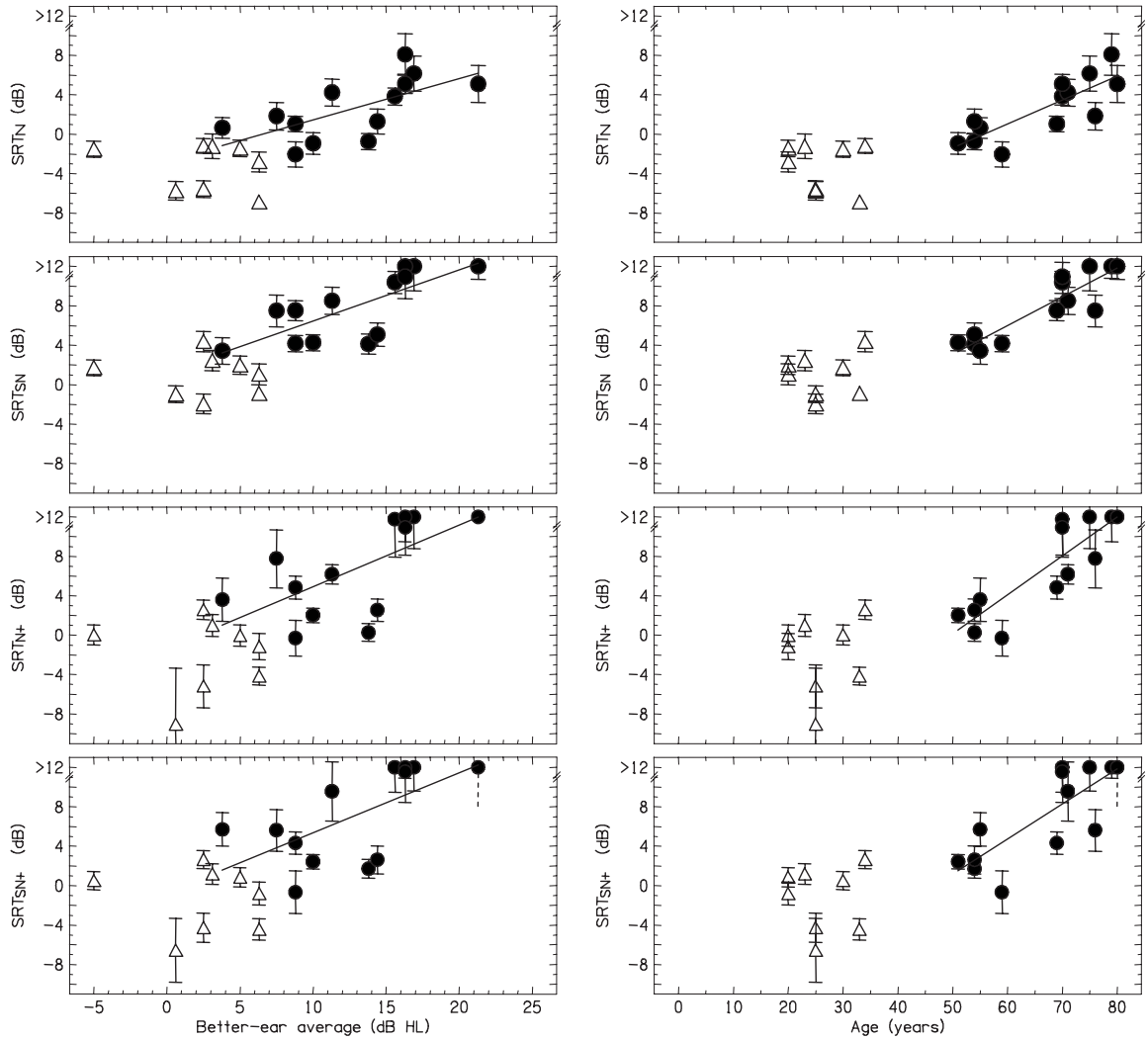


FIG. 12. The SRTs measured in experiment 3. The conditions are ordered by row. The data are plotted as a function of hearing level (left panels) or age (right panels; any SRT larger than 12 dB was capped at that value (see text). The data from the young listeners are shown as open triangles; the data from the elderly listeners as filled circles. The regression line was fitted to elderly listeners' data only (solid line). The error bars show 95%-confidence intervals; a dashed line indicates an error bar that was too large to show on the graph. Note that the range of BEAs is considerably smaller than that shown in Fig. 4.

vious experiment suggested that differences in intensity could serve as a potential cue to release informational masking). The speech-to-noise ratio in the competing channel was also fixed, but the speech-to-noise ratio in the target channel was varied. It seems unlikely that these differences in signal-to-noise ratio provided a cue with which to distinguish the target and competing speech, as informational masking was spread relatively widely across the range of speech-to-noise ratios tested: there were no obvious dips in the speech-recognition psychometric functions at 0 dB (cf. Egan *et al.*, 1954) or plateaus below 0 dB (cf. Brungart, 2001). It could be that much of the informational masking had already been released by other cues, such as the spectral differences between the target and competing speech that were introduced by the filtering processes, although that would suggest that the amount of informational masking occurring with unfiltered speech could be significantly greater than the 4–5 dB observed here or the 3–4 dB observed in experiment 1.

The amount of informational masking observed here was far less than the 22 dB observed by Arbogast *et al.* (2002), whose method we adapted. Nevertheless, the differences between the two experiments were substantial: (1) Arbogast *et al.* (2002) used pure-tone vocoded speech, not narrow-band filtered speech; (2) their vocoded speech was formed from many combinations of carrier frequencies, whereas our filtered speech was always formed from one of just two sets of frequency bands; (3) they estimated informational masking as the difference in SRTs between (in the present terminology) an N condition and an S condition (not an SN condition); and (4) they used the closed-set CRM corpus, not the open-set ASL corpus. Of these, we believe that the primary determinant of their 22-dB effect was their use of the CRM corpus, as large amounts of informational masking have also been observed in other experiments using this corpus (e.g., 12 dB by Brungart, 2001). Furthermore, in subsequent tests we conducted, using CRM sentences that were band-filtered in a similar fashion to the current experiment, we observed effects of up to 30 dB for normal-hearing listeners between the N and S conditions (Agus, 2008). These comparisons highlight the extent to which methodological differences affect measures of informational masking.

Our calculation of the amount of informational masking here is only strictly valid if the competing speech and target speech were ideally isolated from one another; that is, that they did not overlap in excitation. The comparison of SRT_{N+} and SRT_N demonstrated that increasing the level of the competing noise by 14 dB led to an increase in masking of 1–3 dB (SRT_{N+} versus SRT_N). This result suggests that there was indeed some overlap in excitation between the frequency bands: if there was none, then the $SRT_{N+}-SRT_N$ difference would have been zero. But if a 14-dB change in the level of the noise led to only about a 1–3-dB change in SRT, then the 3-dB increase in overall power due to adding the competing speech would be expected to have only a minimal effect on SRT (the effect was quantified by Agus, 2008 and found to be just 0.1 dB). The overlap in excitation can therefore be neglected in deriving the amount of informational masking. But the addition of the 14-dB change in level of the competing noise led to a *reduction* in informational masking of about 1–2 dB (i.e., SRT_{SN+} versus SRT_{SN}). This suggests that there was less informational masking in the SN+ condition than the SN condition, as the additional competing noise would only have increased the energetic masking of the target speech. This result parallels those of Kidd *et al.* (2005) and Freyman *et al.* (2001), who also found that the addition of a noise reduced the amount of informational masking induced by competing speech, presumably by energetically masking the competing speech. These results highlight an important contrast between the two kinds of masking: an increase in the power of a noise always raises energetic masking but can reduce informational masking.

V. SUMMARY

Three experiments were conducted to study the informational masking for target speech by competing speech. All used a speech-in-speech-in-noise method, designed to conceptually separate the energetic-masking and informational-masking contributions to the speech-reception threshold for target speech. The stimuli were sentences chosen from the ASL corpus. Across all experiments, the amount of informational masking observed was of the order of 3–5 dB. Both experiments 1 and 2 used unprocessed sentences for both the target and competing speech: Experiment 1 demonstrated that elderly listeners were no more susceptible to informational masking than younger listeners, but hearing-impaired listeners showed typically 1–2 dB less informational masking, while experiment 2 showed that the amount of informational masking observed was largest when the levels of the target speech and competing speech were about equal. Experiment 3 used sentences that were narrow-band filtered into separated, non-overlapping frequency bands. Again there was no significant effect of age group on informational masking, although there were some indications that the oldest listeners may have been the most susceptible to informational masking. The present results are consistent with previous experiments (Tillman *et al.*, 1973; Li *et al.*, 2004; Arbogast *et al.*, 2005), showing that any detrimental effects of age and hearing impairment on informational masking of speech by speech are, at most, small.

ACKNOWLEDGMENTS

T.R.A. was funded by a Ph.D. studentship from the Medical Research Council. The Scottish Section of IHR is supported by intramural funding from the Medical Research Council and the Chief Scientist Office of the Scottish Government. We thank Richard Freyman (Associate Editor) and two anonymous reviewers for their insightful comments on the submitted manuscript.

[1]Note that the addition of the static noise also reduces the scale of the fluctuations in power compared to speech alone, and so reduces the opportunity for dip-listening (e.g., Howard-Jones and Rosen, 1993; Cooke, 2006).

[2]This equation is the same as Qin and Oxenham (2003), except that it has been recast so that the slope parameter represents the peak gradient in %/dB.

[3]Carhart used the "frequent peaks" of a VU meter, not a long-term rms average, to compare the levels of speech to noise.

[4]Practical constraints prevented us from running this experiment on hearing-impaired or elderly listeners, but a comparison of similar data for different listeners would clearly be of interest.

[5]There was one outlier, namely, listener 3's higher SRT at the speech-to-speech ratio of −2 dB. In the underlying psychometric function for this speech-to-speech ratio, the listener never scored above 40%. Thus the SRT, estimated at 50% speech reception, was an unreliable extrapolation, leading to a large 95%-confidence interval.

Agus, T. R. (**2008**). "Informational masking of speech for elderly listeners," Ph.D. thesis, University of Strathclyde, Glasgow, UK.

Agus, T. R., Akeroyd, M. A., Nobel, W., and Bhullar, N. (**2009**). "An analysis of the masking of speech by competing speech using self-report data," J. Acoust. Soc. Am. **125**, 23–26.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G., Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bench, J., and Bamford, J. (**1979**). *Speech-Hearing Tests and the Spoken Language of Hearing-Impaired Children* (Academic, London).

Bolia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969a**). "Perceptual masking in multiple background sounds," J. Acoust. Soc. Am. **45**, 694–703.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969b**). "Release from multiple maskers: Effects of interaural time disparities," J. Acoust. Soc. Am. **45**, 411–418.

Carhart, R., Tillman, T. W., and Johnson, K. R. (**1968**). "Effects of interaural time delays on masking by two competing signals," J. Acoust. Soc. Am. **43**, 1223–1230.

CHABA (**1988**). "Speech understanding and aging," J. Acoust. Soc. Am. **83**, 859–895.

Cooke, M. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Durlach, N. I. (**2006**). "Auditory masking: Need for improved conceptual structure (L)," J. Acoust. Soc. Am. **120**, 1787–1790.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003**). "Note on informational masking (L)," J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. **26**, 774–782.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.
Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). "The role of perceived spatial separation in the unmasking of speech," J. Acoust. Soc. Am. **106**, 3578–3588.
Gatehouse, S., and Noble, W. (**2004**). "The speech, spatial and qualities of hearing scale (SSQ)," Int. J. Audiol. **43**, 85–99.
Hartmann, W. M. (**1998**). *Signals, Sound, and Sensation* (Springer-Verlag, New York).
Hawkins, J. E., Jr., and Stevens, S. S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Hornsby, B. W. Y., Ricketts, T. A., and Johnson, E. E. (**2006**). "The effects of speech and speech-like maskers on unaided and aided speech recognition in persons with hearing loss," J. Am. Acad. Audiol **17**, 432–447.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Acoust. Soc. Am. **93**, 2915–2922.
Humes, L. E., Lee, J. H., and Coughlin, M. P. (**2006**). "Auditory measures of selective and divided attention in young and older adults using single-talker competition," J. Acoust. Soc. Am. **120**, 2926–2937.
Kidd, G., Jr., Mason, C. R., and Gallun, F. J. (**2005**). "Combining energetic and informational masking for speech identification," J. Acoust. Soc. Am. **118**, 982–992.
Kwon, B. J., and Turner, C. W. (**2001**). "Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference?," J. Acoust. Soc. Am. **110**, 1130–1140.
Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (**2004**). "Does the informational content of an irrelevant source differentially affect spoken word recognition in younger and older adults?," J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091.
Lutfi, R. A. (**1990**). "How much informational masking is informational masking?," J. Acoust. Soc. Am. **88**, 2607–2610.
MacLeod, A., and Summerfield, Q. (**1990**). "A procedure for measuring auditory and audiovisual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use," Br. J. Audiol. **24**, 29–43.
Oh, E. L., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Tillman, T. W., Carhart, R., and Nicholls, S. (**1973**). "Release from multiple maskers in elderly persons," J. Speech Hear. Res. **16**, 152–160.
Tun, P. A., O'Kane, G., and Wingfield, A. (**2002**). "Distraction by competing speech in younger and older listeners," Psychol. Aging **17**, 453–467.
Tun, P. A., and Wingfield, A. (**1999**). "One voice too many: Adult age differences in language processing with different types of distracting sounds," J. Gerontol. B Psychol. Sci. Soc. Sci. **54B**, 317–327.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta Acust. **91**, 502–512.
Wu, X., Wang, C., Chen, J., Hongwei, Q., Li, W., Wu, Y., Schneider, B. A., and Li, L. (**2005**). "The effect of perceived spatial separation on informational masking of Chinese speech," Hear. Res. **199**, 1–10.

Effects of external noise on detection of intensity increments[a)]

Walt Jesteadt,[b)] Kim S. Schairer,[c)] Lance Nizami, Samar Khaddam, and Stephen T. Neely
Center for Hearing Research, Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 4 March 2009; revised 18 July 2009; accepted 21 July 2009)

The detection of an intensity increment in a longer duration sinusoid or pedestal is often used as a measure of intensity resolution, but the decision processes underlying this measure are poorly understood. Thresholds were obtained for detection of an increment in a 370-ms, 4-kHz pedestal in quiet or in noise to determine the relative contributions of background noise level and pedestal level, the effect of increment duration, and the effect of different noise spectra. Increment detection thresholds expressed in units of $\Delta L[10\log(1+\Delta I/I)]$ decreased as pedestal levels increased. At low pedestal levels, increment detection was limited by the masking effect of the noise and was similar across noise conditions for pedestals of equal sensation level. At high pedestal levels, the noise had no effect and increment detection was determined by the pedestal level in dB SPL (sound pressure level). Increment detection improved with increasing increment duration and was altered less by a noise band above the pedestal/increment frequency than by a broadband noise that produced equal masking at the pedestal/increment frequency. The quadratic-compression model described by Neely and Jesteadt [(2005). Acta Acust. Acust. **91**, 980–991] provided a better approximation to the data than a model based on excitation patterns.
 [DOI: 10.1121/1.3203994]

PACS number(s): 43.66.Fe, 43.66.Dc, 43.66.Ba [MW] Pages: 1941–1953

I. INTRODUCTION

Increment detection (detection of an intensity increment when the standard or pedestal is longer in duration than the increment or signal) and intensity discrimination (when the increment and pedestal are the same duration) are two well-established measures of intensity resolution that appear to rely on different decision processes. One line of evidence indicating different underlying processes is that interval-to-interval variation in overall level has little effect on increment detection (Jesteadt *et al.*, 2003) but has a predictable effect of degrading intensity discrimination (Jesteadt *et al.*, 2005). A second line of evidence indicating different underlying processes is that background noise has been shown to have a significant effect in limiting increment detection (Glasberg *et al.*, 2001; Oxenham, 1997) but less effect on intensity discrimination (Jesteadt *et al.*, 1977; Neff and Jesteadt, 1996). The present study explores the effects of noise on increment detection over a wide range of levels to determine the factors that contribute to those effects.

Two previous studies have examined increment detection in quiet and in noise. Oxenham (1997) obtained thresholds for detection of both increments and decrements in a 4-kHz, 500-ms pedestal. The pedestal was presented at 55 dB SPL (sound pressure level), with noise at −20, −10, 0, 10, or 20 dB spectrum level, plus a no-noise condition. Increments or decrements had steady-state durations ranging from 2 to 198 ms with 2-ms raised-cosine ramps. Only noise levels of 0-dB spectrum level and higher had an impact on increment detection, and the effect was greater at shorter increment durations. In a second experiment, Oxenham obtained thresholds for detection of increments and decrements using the same pedestal, with 10-dB spectrum level noise that was either low-pass filtered at 4.4 kHz, high-pass filtered at 3.6 kHz, or band-reject filtered to create a notched noise with cutoff frequencies of 3.6 and 4.4 kHz. Increment detection was best in low-pass noise, suggesting that the primary effect of the noise was to mask upward spread of excitation of the pedestal and pedestal plus increment. The effect of low-pass noise compared to no noise, however, suggested that noise played some role beyond limiting excitation. Additional conditions ruled out any effect of spectral splatter.

Oxenham (1997) evaluated a multi-stage temporal window model that has been used frequently in modeling increment and decrement detection (e.g., Peters *et al.*, 1995) and found that it predicted little or no effect of noise, although the data indicated that noise had a significant effect on increment detection and an even larger effect on decrement detection. Oxenham noted that the noise might interfere with modulation detection as well as limiting spread of excitation.

Comparing increment detection in quiet and in noise, Glasberg *et al.* (2001) repeated some of the detection-in-noise conditions used in earlier studies (Peters *et al.*, 1995; Moore *et al.*, 1996, 1999; Moore and Peters, 1997) and included corresponding conditions for increment and decrement detection in quiet. The pedestal was a 0.25-, 1-, or 4-kHz tone presented 18 dB above masked or quiet threshold [i.e., at 18 dB SL (sensation level)]. The noise maskers were chosen such that the output of an auditory filter centered on

[a)] A portion of this work was presented at the 147th Meeting of the Acoustical Society of America.
[b)] Author to whom correspondence should be addressed. Electronic mail: jesteadt@boystown.org
[c)] Present address: University of Wisconsin, 1975 Willow Drive, Room 373, Madison, WI 53706.

the pedestal would be 56.6 dB. Thus, only one noise level (and one corresponding signal-to-noise ratio) was tested in each masking condition. As in earlier studies, increment detection thresholds for conditions with noise maskers decreased with increasing increment duration and with increasing pedestal frequency. In conditions without noise, increment detection thresholds showed a similar effect of duration but were either constant or increased with increasing pedestal frequency. Increment detection at 4 kHz was poorer in quiet than in noise for pedestals that were 18 dB SL in both conditions. Glasberg *et al.* suggested that this unusual result might indicate that external noise had a smaller effect on performance than fluctuations in internal noise. A second experiment used only 4-kHz, 70-dB SPL pedestals presented in quiet or in three noise conditions: a narrow-band noise centered at either 4 or 7 kHz or with both bands together. To determine whether the interaction of noise and frequency effects in the first experiment was due to differences in the bandwidth of the noise at the output of auditory filters at 0.25, 1, and 4 kHz, the bandwidth of the noises in the second experiment varied from 0.05 to 0.4 kHz. This range was chosen to include the equivalent rectangular bandwidth (ERB) of the auditory filter at 0.25 kHz but to stop short of the 0.456 kHz ERB at the 4-kHz pedestal frequency used in the experiment. The effect of noise at 7 kHz was small, either alone or in combination with noise at 4 kHz, and the effect of noise bandwidth was small as well, with slightly better performance at the widest bandwidth.

The model used by Glasberg *et al.* (2001) was similar to the model described by Oxenham (1997), with the addition (after the temporal integrator) of a log transform, an internal noise, and a decision device based on a template-matching mechanism (Moore *et al.*, 1999). The decision process was based on a comparison (cross correlation) of the internal representation of the incoming stimulus with a template or pattern of levels across time that was created by averaging internal representations of more audible stimuli. Moore *et al.* (1999) and Glasberg *et al.* (2001) noted that template matching has elements in common with modulation detection, as described by Dau *et al.* (1996, 1997).

Oxenham (1997) and Glasberg *et al.* (2001) found evidence that background noise influenced increment detection both by limiting spread of excitation and by introducing modulation that might interfere with template matching or modulation detection. Oxenham emphasized the first interpretation, whereas Glasberg *et al.* emphasized the second. The effect of spread of excitation was reduced in these studies by the use of a limited range of levels. When pedestal level is varied over a wide range in quiet, the effect of level is larger than the effects of frequency, of increment duration, or of the other parameters that have been manipulated in studies of increment detection (e.g., Viemeister and Bacon, 1988). Thus, it is important to understand the origins of the level effect and how to control it.

To obtain a more complete picture of the effect of spread of excitation on increment detection and of the role of noise in reducing that effect, thresholds for increment detection were obtained at a number of different noise spectrum levels, with multiple pedestal-to-noise ratios for each noise level. The effect of increment duration was explored across three experiments. A fourth experiment explored the effect of varying the bandwidth of the noise masker.

II. EXPERIMENT 1: EFFECT OF PEDESTAL SL AND NOISE SPECTRUM LEVEL (60-MS INCREMENT)

In experiment 1, the range of conditions was expanded beyond previous studies. Increment detection thresholds were obtained over a range of noise and pedestal levels to compare increment detection as a function of pedestal level in both dB SPL and dB SL. The goal was to determine if the effects of noise on increment detection across a range of levels could be accounted for by assuming that the noise limited the spread of excitation or whether it was necessary to assume that an additional factor such as envelope fluctuation was involved.

A. Subjects

Four paid volunteers (1 male and 3 females), ages 19–21 years, participated. All were college students and two have previous experience with an intensity discrimination task. Hearing had been screened at 0.5, 1, 2, and 4 kHz within the past year using a two-interval-forced-choice (2IFC) adaptive procedure. Thresholds were $\leq$18 dB SPL for all test frequencies bilaterally for all subjects.

B. Apparatus

Stimuli were generated digitally at a sampling rate of 50 kHz using an array processor [Tucker Davis Technologies (TDT) AP2] and 16-bit digital-to-analog converters (TDT DD1). The pedestal and noise were generated on one channel of the DD1; the increment was generated on the other. The output of each channel was low-pass filtered at 20 kHz (TDT FT6) and attenuated (TDT PA4), and then the outputs were combined (TDT SM3) and presented to the listener through a headphone buffer (TDT HB6), a remote passive attenuator in the sound-treated chamber, and a Sennheiser HD 250 Linear II headphone. Stimuli were delivered to the left ear. With parallel attenuators, summers, and headphone buffers, up to four listeners were tested simultaneously. The same equipment was used in all experiments.

C. Procedure

Adaptive thresholds were estimated in a 2IFC task. Each trial consisted of a 300-ms warning interval, two 390-ms observation intervals separated by 560 ms, and a 100-ms feedback interval following the response of the final subject. A 400-ms pause preceded the beginning of the next trial. Visual markers for the warning and observation intervals and correct-interval feedback were given on a message window at the top of the keypad used to indicate responses. The increment was presented at an initial level well above threshold, decreased after two consecutive correct responses, and increased after one incorrect response to estimate the 71% correct point on the psychometric function (Levitt, 1971). Initial step sizes were 4 dB until the fourth reversal, and then 2 dB. Threshold was calculated as the average level of the

TABLE I. Parameters of stimuli used in the experiments.

Expt.	Pedestal level	Signal duration (ms)	Noise level (dB N_0)	Noise band
1	15, 25, 35, 45, 55, 65, and 75 dB SPL	60	0, 10, or 20	8 kHz low pass
2	−5, 0, 5, 10, 15, 20, 25, 30, 40, 55, 70, and 85 dB SL	10	0, 10, or 20	8 kHz low pass in all conditions; 16 kHz low pass in 40 dB SL pedestal, 0 dB N_0 condition
3	0, 5, 10, 20, 40, and 80 dB SL	20, 40, or 60	20	8 kHz low pass
4	10, 20, 30, 40, and 50 dB SL	10	20	4.8–16 kHz bandpass; 5.6–16 kHz bandpass; 7.2–16 kHz bandpass; 16 kHz low pass

Note: Pedestal and signal were 4 kHz and pedestal and noise durations were 370 ms in all conditions. The 0, 10, and 20 dB N_0 values correspond to 39, 49, and 59 dB overall level. For pedestal level specified in SL, the reference is the mean threshold across subjects for the pedestal in quiet or in noise.

reversal points after the fourth reversal for 100-trial blocks. The number of blocks per condition is specified below.

D. Stimuli and conditions

The pedestal was a 4-kHz, 370-ms tone with 5-ms $\cos^2$ ramps, presented in different conditions at 15, 25, 35, 45, 55, 65, and 75 dB SPL (not all pedestal levels were presented in each noise condition). The increment was a 4-kHz, 60-ms tone with 5-ms $\cos^2$ ramps, adapted in level to estimate threshold. The onset of the increment was 205 ms after the onset of the pedestal (0-voltage points). These temporal parameters resulted in a 200-ms steady-state segment of the pedestal preceding and a 100-ms segment following the increment, the same stimulus configuration used by Glasberg *et al.* (2001). The background noise, when present, was low-pass filtered at 8 kHz, was turned on and off with the pedestal, and was presented at 0, 10, and 20 dB spectrum level (N_0) in different conditions. These spectrum levels correspond to 39, 49, and 59 dB overall level. The stimulus parameters are summarized in Table I.

In addition to estimates of the threshold of the increment in the presence of the pedestal, for pedestals presented in quiet and in noise, estimates of thresholds were obtained for the increment without the pedestal and for the pedestal itself (i.e., a 60-ms, 4 kHz tone and a 370-ms, 4 kHz tone presented in quiet and in each of the three levels of noise). When the increment and pedestal were presented together, initial pedestal phase was fixed at 0°, and increment phase was tied to the pedestal in quadrature. When the tones were presented individually, the initial phase was random. In increment detection conditions for pedestals presented in noise, a different range of pedestal levels was used for each noise level. In 0-dB N_0 conditions, increment detection was measured using a range of pedestal levels from 25 to 55 dB SPL. The ranges were 35–65 dB SPL in 10-dB N_0 noise and 45–75 dB in 20-dB N_0 noise. In the absence of noise, increment detection was measured for pedestal levels from 15 to 75 dB SPL. Two estimates of threshold were obtained for the short and long duration 4-kHz tones in quiet after all other data collection was complete. Threshold estimates in each of the remaining conditions were based on two repetitions of two blocks each in a counterbalanced order.

E. Results and discussion

Mean thresholds for the pedestal in quiet and in noise are shown in Table II for this experiment and for conditions used in later experiments. Increment detection results for experiment 1 are shown in Fig. 1, where the pedestal level ($10 \log I$) and increment level at threshold ($10 \log \Delta I$) have been converted to $\Delta L(10 \log(1+\Delta I/I))$ and plotted on a log axis (Buus and Florentine, 1991). Geometric mean ΔL across subjects is shown as a function of pedestal level in dB SPL in the left panel and in dB SL (re thresholds for detection of the pedestal in quiet or in noise, from Table II) in the right panel. Because threshold for detection of the pedestal in quiet is close to 0 dB SPL, there is little shift in the data for the

TABLE II. Mean threshold for detection of the pedestal in quiet and in noise for conditions in experiments 1–4. Threshold values are means across subjects in dB SPL. Four subjects participated in experiment 1. Another group of three subjects participated in experiments 2–4.

	Threshold in dB SPL	Standard deviation
Experiment 1		
Quiet	5.1	6.32
0 dB N_0, 8 kHz low pass	25.1	2.15
10 dB N_0, 8 kHz low pass	34.0	1.81
20 dB N_0, 8 kHz low pass	43.9	2.40
Experiments 2 and 3		
Quiet	−0.6	5.42
0 dB N_0, 8 kHz low pass	19.4	0.58
10 dB N_0, 8 kHz low pass	33.7	1.29
20 dB N_0, 8 kHz low pass	42.4	1.13
0 dB N_0, 16 kHz low pass	21.8	4.07
Experiment 4		
20 dB N_0, 4.8–16 kHz	21.0	9.14
20 dB N_0, 5.6–16 kHz	10.3	2.05
20 dB N_0, 7.2–16 kHz	4.5	5.75
20 dB N_0, 16 kHz low pass	43.5	2.67

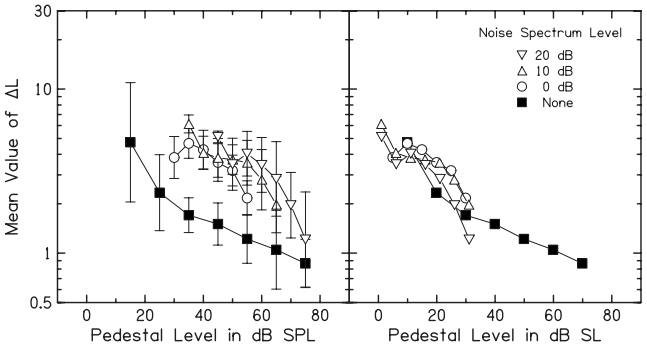


FIG. 1. Geometric mean values of ΔL across four subjects for experiment 1. ΔL was calculated as $10\log(1+(\Delta I/I))$. The 4-kHz, 60-ms increment was presented 205 ms after the onset of a 4-kHz, 370-ms pedestal. Increment detection was measured in conditions without noise (filled squares) or with 0- (circles), 10- (up triangles), or 20-dB (down triangles) spectrum level (N_0) noise that was low-pass filtered at 8 kHz and gated on and off with the pedestal. The data are plotted as a function of pedestal level in dB SPL on the left and in dB SL (based on thresholds for detection of the pedestal in quiet and in noise) on the right. For clarity, error bars representing ± 1 standard deviation are shown only on the left panel.

no-noise condition, but the effects of noise are different in the two panels. In dB-SPL coordinates, mean values of ΔL are higher in noise for a given pedestal level than in the no-noise condition, and mean ΔL increases with noise level. In dB SL coordinates, results are similar across noise levels, and the difference between noise and no-noise conditions is small. Compared to the no-noise condition, there is an improvement in threshold in the highest pedestal-level and noise-level conditions. Also, there is the appearance of a notch in the data around 10 dB SL. Error bars have been plotted in the left panel but omitted in the right panel to make the data points more visible.

The use of different pedestal-level ranges for different noise conditions was efficient but made statistical analyses more difficult. For this and later experiments, a within-subjects analysis of variance (ANOVA) was conducted on the log ΔL values for pedestal levels that were in common across noise levels. Effects are reported as significant for $\alpha < 0.05$. An ANOVA for the four subjects that was restricted to the three pedestal levels (45, 50, and 55 dB SPL) used at all three noise levels indicated significant effects of pedestal level $[F(2,6)=9.085]$ and noise level $[F(2,6)=22.656]$ but no interaction. A second ANOVA of six pedestal levels approximately equated in dB SL indicated a significant effect of pedestal level $[F(5,15)=46.319]$, a marginally significant effect of noise level $[F(2,6)=5.470]$, and a significant interaction. The large effect of noise level evident in the first analysis and in the left panel of Fig. 1 is all but eliminated when the pedestal levels are grouped by SL as in the right panel of Fig. 1.

As suggested by Oxenham (1997), the main effect of the noise is to restrict the spread of excitation. The area under the excitation pattern is relatively constant across noise levels for a constant-SL pedestal, and increment detection is relatively constant as well. Increment detection improves as the signal-to-noise ratio of the pedestal increases at a given noise level. This appears to be due to increased spread of excitation rather than increased distance from the fluctuations in the noise wave form because the same degree of improvement with pedestal level can be seen in the absence of noise. The notch at 10 dB SL is reminiscent of the notch in intensity discrimination near threshold (Hanna *et al.*, 1986) but is small and occurred at a higher SL than the notch reported by Hanna *et al.*

III. EXPERIMENT 2: EFFECT OF PEDESTAL SL AND NOISE SPECTRUM LEVEL (10-MS INCREMENT)

Results of experiment 1 suggested that the effect of noise on increment detection was to restrict spread of excitation rather than to introduce modulation that would interfere with detection of a brief change in the level of a longer duration pedestal. Because the increment duration was relatively long, however, any independent effects of the noise alone, such as effects due to envelope fluctuations, might have been missed. This issue was addressed in experiment 2 by use of a shorter, 10-ms increment. To explore the possible near-threshold notch in greater detail, pedestal levels were equated in dB SL and were extended to -5 dB SL in the noise conditions.

A. Subjects

Three volunteers (1 male, 2 females, ages 31–33 years) participated in experiment 2 and in the two subsequent experiments. One subject was paid for his participation. Two of the subjects (one of the authors and a college student) had previous experience with an intensity discrimination task, while the third subject (a new laboratory assistant) did not. Hearing thresholds were tested using standard clinical procedures within the 6 months prior to the experiment. Thresholds were ≤ 15 dB HL (hearing level) at all audiometric test frequencies from 0.5 to 8 kHz bilaterally for all subjects. Air-bone gaps were no greater than 5 dB.

B. Stimuli and procedures

The 4-kHz pedestal was the same as in experiment 1, except that pedestal levels in quiet ranged from 10 to 85 dB SL (re mean threshold across subjects for the pedestal-in-quiet condition) and pedestal levels in noise ranged from -5 to 40 dB SL (re the mean threshold across subjects for the pedestal in the presence of each of the three noise levels). The background noise properties were the same as in experiment 1. An additional single condition was run with a 16-kHz low-pass noise in the 0-dB N_0, 40-dB SL pedestal condition to determine whether additional noise in the 8–16 kHz range would influence increment detection thresholds by masking off-frequency listening or spread of excitation at higher frequencies. The increment was the same as in experiment 1, except that the duration was 10 ms with 5-ms $\cos^2$ ramps (i.e., no steady state). Stimulus parameters are summarized in Table I.

Two consecutive 100-trial blocks were completed for conditions in which the pedestal was presented without noise at 85, 70, 55, 25, and 10 dB SL and for the 16-kHz low-pass noise conditions. Four 100-trial blocks were completed for all other conditions. Masked thresholds were obtained for the increment and the pedestal in the 0- and 20-dB N_0, 8-kHz

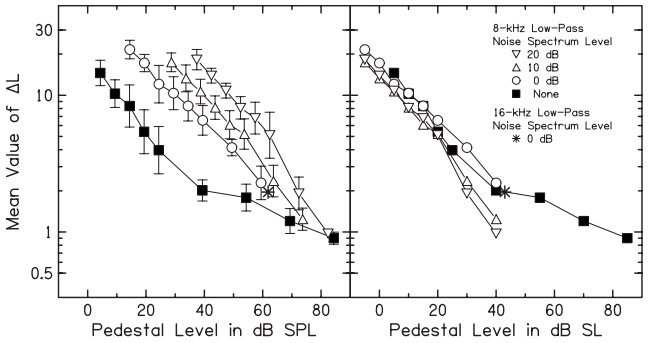


FIG. 2. Geometric mean values of ΔL across three subjects for experiment 2 plotted as in Fig. 1. Error bars representing ± 1 standard deviation are shown only on the left panel. The increment duration was 10 ms. All other stimulus parameters, except for specific pedestal levels, were the same as in experiment 1. The asterisk in each panel represents increment detection measured in the presence of a 0-dB spectrum level noise that was low-pass filtered at 16 kHz, with the pedestal at 40 dB SL (re mean threshold across subjects for the pedestal in the 16-kHz low-pass noise).

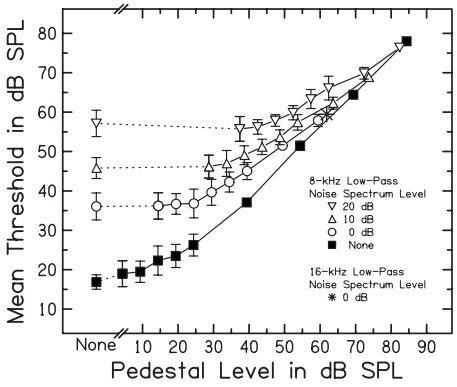


FIG. 3. Mean thresholds for the increment in dB SPL across three subjects as a function of pedestal level in dB SPL for experiment 2 with error bars representing ± 1 standard deviation. Symbol definitions are the same as in Fig. 2. Data points on the far left show thresholds for detection of the increment in the absence of the pedestal in quiet and in the three levels of noise. Increment detection thresholds are limited by the noise at low pedestal levels and converge on a single value in dB SPL at high pedestal levels.

low-pass noise conditions, followed by increment detection thresholds in those two noise conditions. Absolute thresholds for the increment and the pedestal and increment detection thresholds in quiet were then measured, followed by masked thresholds and increment detection thresholds in 10-dB N_0, 8-kHz low-pass conditions. Finally, masked thresholds and increment detection thresholds were measured in the 20-dB N_0, 16-kHz low-pass noise.

C. Results and discussion

The results are shown in Fig. 2 in a form that parallels Fig. 1. The asterisk in each panel indicates the data point obtained at the lowest noise level and highest pedestal level using noise low-pass filtered at 16 kHz. Results for this condition were consistent with the other data, suggesting that subjects were not using information above the 8-kHz cutoff of the low-pass noise in the main set of conditions.

As in experiment 1, the large effect of noise that is apparent when the data are plotted as a function of pedestal level in dB SPL is greatly reduced when the data are plotted in dB SL. Unlike experiment 1, there is no notch at low levels. Results are similar across noise levels. The difference in results between no-noise and noise conditions is small, except at high SLs where ΔL for increment detection in quiet is larger than ΔL in noise for pedestals equal in dB SL. The shapes of the functions differ for no-noise and noise conditions. The no-noise function becomes shallower at higher levels, whereas the noise functions become steeper with a greater effect of noise at higher noise levels.

Data were available for the quiet conditions and the three noise levels at pedestal levels of 5, 10, 15, 20, and 40 dB SL. An ANOVA of the log ΔL vales for the three subjects showed significant effects of pedestal SL $[F(4,8)=441.432]$, noise level $[F(3,6)=7.104]$, and a significant interaction $[F(12,24)=11.502]$. ANOVAs conducted at individual pedestal levels showed a significant effect of noise level only at 40 dB SL $[F(3,6)=23.676]$. *Post hoc* comparisons of means at 40 dB SL did not show a significant difference between any two conditions. The analyses support the conclusion that there is an effect of noise level at high pedestal levels even when pedestal levels are equated in SL.

Given that both the noise and the pedestal can be viewed in some sense as maskers, it is informative to re-plot the data as masked thresholds for the increment, as shown in Fig. 3. Thresholds for the increment in quiet and in noise are shown on the left of the figure. At low pedestal levels, thresholds for the increment approach thresholds for the pedestal as a limit. At high pedestal levels, the noise ceases to have any effect on increment threshold. As a result of these two constraints, the masking functions become shallower with increasing noise level. To estimate the slope of masking, each set of data points was fitted with a linear regression function after excluding points near threshold for the pedestal. Slopes, intercepts, and the range of pedestal levels included in each fit are given in Table III. In the absence of noise, the increment threshold, $10\log(\Delta I)$, increases 0.87 dB for every decibel increase in the pedestal, $10\log(I)$. This pattern has been described as a "near miss" to Weber's law (McGill and Goldberg, 1968a, 1968b). In the presence of noise, the deviation from Weber's law increases with noise level. This is true regardless of whether the data are plotted in dB SPL or in dB SL.

A model of loudness and partial loudness proposed by Moore *et al.* (1997) that is based on excitation patterns can be used to make predictions for increment detection thresholds in quiet and in noise to determine the extent to which the results in Figs. 1 and 2 can be accounted for on the basis

TABLE III. Parameters for linear functions describing the increase in pedestal level with increment level in Fig. 3. The correlation with the data was greater than 0.995 in all cases.

Noise level	Slope	*Y* intercept	Min SPL	Max SPL	No. of points excluded
None	0.87	4.0	24	84	4
0 dB N_0	0.60	21.7	24	59	2
10 dB N_0	0.55	27.6	34	74	1
20 dB N_0	0.51	34.1	42	82	1

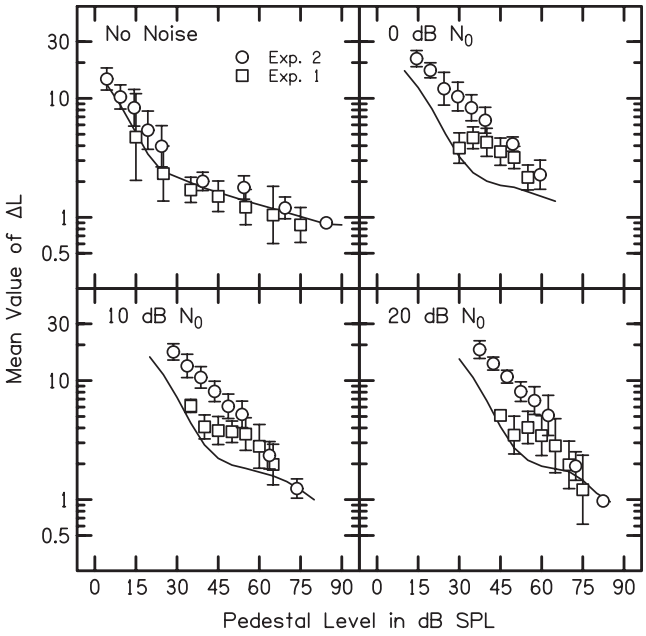


FIG. 4. A comparison of geometric mean values of ΔL from experiment 1, where the increment duration was 60 ms, with those from experiment 2, where the increment duration was 10 ms, with noise conditions across panels. Error bars represent ± 1 standard deviation. The solid line shows the values of ΔL predicted by the partial loudness model using an 8-phon criterion. In the absence of noise, the two sets of data are in good agreement, showing an expected effect of increment duration, and are well fitted by the model. With noise present, the data for the two studies diverge at low pedestal levels, where the model predicts better than observed performance.

of excitation patterns alone. The model has no information about the temporal properties of stimuli but makes accurate predictions of thresholds for a number of tasks involving detection of tones in noise and intensity discrimination (Jesteadt *et al.*, 2007). The only free parameter in the model is the threshold level of loudness (in phons) associated with detection of the increment. Moore *et al.* (1997) proposed a criterion value of 2 phons for prediction of masked thresholds but used higher levels to approximate results for brief stimuli to compensate for the fact that the model is not sensitive to duration. Data from experiments 1 and 2 are compared with model predictions using an 8-phon criterion in Fig. 4. To compare data and model predictions at each noise level, pedestal levels have been plotted in dB SPL. The pattern within panels would not change if the levels were plotted in dB SL. Data from the two experiments for the no-noise condition are in good agreement, with values of ΔL for experiment 2, where the increment duration was 10 ms, uniformly above those for experiment 1, where the increment duration was 60 ms. The partial loudness model provides a good account of data obtained in the absence of noise. A comparison across experiments for the three noise conditions suggests that there is an effect of increment duration at low pedestal levels in noise but not at high pedestal levels. The model predicts better performance in all noise conditions than is actually observed, with predicted values that parallel those for increment detection in the no-noise conditions, becoming shallower at higher levels. The data, particularly for the longer duration increment in experiment 1, show a curvature that is not predicted by the model. Although the major effect of the noise is eliminated by expressing pedestals levels in dB SL, as shown in the right panels of Figs. 1 and 2, a comparison of the data with the detailed predictions of a model based on excitation patterns suggests that the effects of noise cannot be accounted for on the basis of excitation patterns alone.

IV. EXPERIMENT 3: EFFECT OF PEDESTAL SL AND INCREMENT DURATION

The goals of experiment 3 were (1) to determine whether the notch in thresholds observed at low pedestal levels in experiment 1 for the 60-ms increment was repeatable in another set of subjects, and if so, at what duration it disappeared; (2) to further explore the increment-duration parameter space over the range of durations used in experiments 1 and 2 to relate the data from these two experiments; and (3) to compare effects of duration with the results from the Glasberg *et al.* and Oxenham studies.

A. Subjects

The same three subjects from experiment 2 participated in experiment 3.

B. Stimuli and procedures

The stimulus parameters, summarized in Table I, were the same as in experiments 1 and 2 wherever possible. The 4-kHz, 370-ms pedestal was presented at 0, 5, 10, 20, and 40 dB SL, with an 80-dB SL condition included when the pedestal was presented without noise. The 8-kHz low-pass noise was presented only at 20 dB N_0. The increment duration was 20, 40, or 60 ms with 5-ms $\cos^2$ ramps. These subjects had just completed a larger set of conditions with 10-ms increments in experiment 2, and a subset of those data was included in the analyses rather than repeating those conditions. Pedestal SLs for experiment 3 were based on the thresholds in quiet and in noise from experiment 2. Two consecutive 100-trial blocks were obtained in each condition, beginning with the longest increment and ending with the shortest increment, no-noise conditions preceding noise conditions at each increment duration. For each increment duration and noise condition, pedestal levels were tested in increasing order from 0 to 80 dB SL.

C. Results and discussion

Geometric mean values of ΔL obtained with a 60-ms increment in experiment 3 are compared with data for the same increment duration and noise level obtained in experiment 1 in Fig. 5. Given that the two experiments used independent groups of subjects and relatively few trials per data point, the agreement is remarkably good. To quantify the degree of agreement between the two sets of data, predicted values of ΔL in experiment 3 were obtained for each pedestal SPL used in experiment 1 using linear interpolation. The correlation between the 14 predicted and observed values of ΔL was 0.97. There is some indication of a notch in the new data, but it is not as clear as the notch observed at all noise levels in experiment 1.

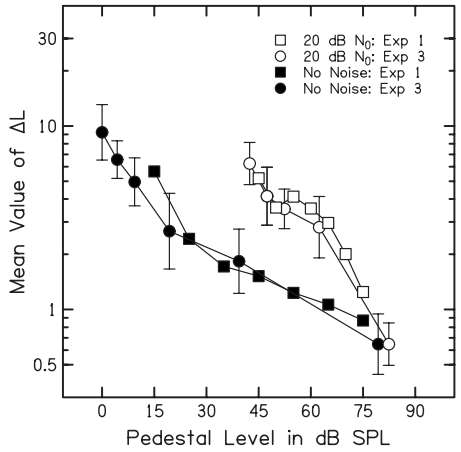


FIG. 5. A comparison of geometric mean values of ΔL from experiment 1, where the increment duration was 60 ms, with a replication of the 60-ms condition in experiment 3 in the condition without noise and with the 20-dB N_0 noise level used in experiment 3. Error bars represent ± 1 standard deviation. There were no subjects in common between the two groups.

Results for all four increment durations in experiments 2 and 3 are shown in Fig. 6, with data for the 10-ms increment re-plotted from Fig. 2. Although pedestal levels were specified in dB SL, the data have been plotted in dB SPL to avoid total overlap of the data obtained at low pedestal levels with and without noise. The noise shifts the threshold for the pedestal by 42.4 dB. The two sets of functions converge at pedestal levels of 40 dB SL (82.4 dB SPL) in noise and 80 dB SL (79.4 dB SPL) in quiet. Consistent with the pattern observed in Fig. 3, the noise ceases to have an effect on performance at a pedestal level of 40 dB SL.

A within-subjects ANOVA of the log ΔL values for the three subjects for noise and no-noise conditions, increment durations of 10, 20, 40, and 60 ms, and pedestal levels of 5, 10, 20, and 40 dB SL (0-dB SL was not available for the 10-ms increment duration) indicated significant effects of noise $[F(1,2)=18.835]$, increment duration $[F(3,6)$ $=34.870]$, and pedestal level $[F(3,6)=209.708]$. There were significant interactions of noise (no noise vs 20 dB N_0) and pedestal level $[F(3,6)=19.695]$ and increment duration and pedestal level $[F(9,18)=8.951)$. The significant effect of noise reflected a geometric mean ΔL across subjects and conditions of 4.55 dB in the absence of noise and of 2.91 dB with noise present. This difference is a result of the physical difference in pedestal levels between noise and no-noise conditions and is barely significant. The other two main effects and the two significant interactions are clearly observable in Fig. 6. There is a curvature, if not a clear notch, for data at low pedestal levels in noise for the two longer increment durations but not the shorter durations. The effect is not sufficient to result in a significant interaction of noise, increment duration, and pedestal level.

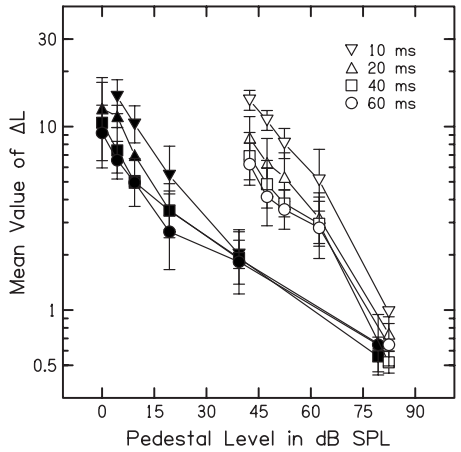


FIG. 6. Geometric mean values of ΔL across subjects for the 20-, 40-, and 60-ms increment durations tested in experiment 3, with data for 10-ms increments for these subjects re-plotted from Fig. 2 (experiment 2). Error bars represent ± 1 standard deviation. The increments were presented 205 ms after the onset of a 370-ms, 4-kHz pedestal. Filled and open symbols show increment detection in quiet and detection in an 8-kHz low-pass, 20-dB N_0 noise equal in duration to the pedestal.

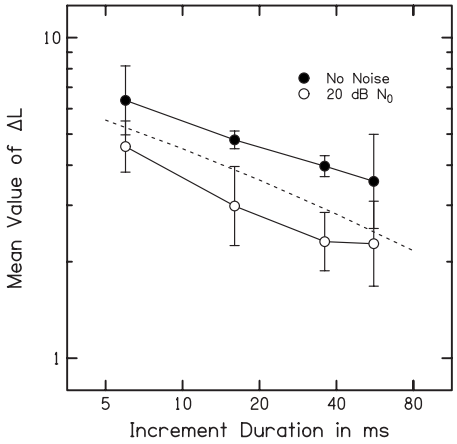


FIG. 7. Geometric mean values of ΔL for all data shown in Fig. 6, as a function of duration. Geometric means were computed across pedestal levels of 5, 10, 20, and 40 dB SL, and then a geometric mean and standard deviation were computed across subjects for each increment duration in quiet and in 20-dB N_0 noise. Increment durations have been corrected to equivalent rectangular durations. Thresholds are lower in noise because the equal-SL conditions correspond to higher dB SPL in noise, but there is an effect of duration with or without noise. The dashed reference indicates increment thresholds in units of 10 log(ΔI) that decrease by 1.5 dB for every doubling of increment duration.

The effect of increment duration is larger at low pedestal levels, but it is generally the case that values of ΔL decrease as increment duration increases. The lack of a duration-by-noise interaction suggests similar effects of duration in quiet and in noise, as shown in Fig. 7. Geometric mean values of ΔL are lower in the noise conditions only because the pedestals were presented at higher physical levels. To correct for the use of 5-ms onset and offset ramps, durations have been plotted as equivalent rectangular durations. Slopes of the best fitting lines in quiet and in noise would be -0.26 and -0.32, respectively. These values are comparable to those reported for intensity discrimination with tones by Florentine (1986) and are within the range reported for increment detection with noise stimuli by Heinz and Formby (1999). If values of ΔL are converted to units proportional to increment power, the improvement in increment threshold is 1.53 dB per doubling of duration or 5.1 dB per tenfold increase. Viemeister and Wakefield (1991) noted that detection thresholds would be expected to improve at a rate of 1.5 dB per doubling rather than 3 dB per doubling if the slope of the psychomet-

ric function were steep or if the integration period of an energy detector were matched to the signal duration. Although either condition might be met for conditions in experiment 3, the effect of duration is larger than that observed at 4 kHz by Moore and Peters (1997) or Oxenham (1997). Given the wide variability in duration effects reported in the literature [see Moore *et al.* (1999) and Wojtczak and Viemeister (1999) for summaries], the similarity of the effect in quiet and in noise in Fig. 7 may be more noteworthy than its size.

The results of experiment 3 are consistent with those of experiment 1 in that increment detection improves slowly with increasing pedestal level when the pedestal is near masked threshold in the noise, and the increment is 40 or 60 ms. For increments shorter than 40 ms, the improvement with increasing pedestal level is more uniform at low and high pedestal levels. The results of experiment 3 are also consistent with experiment 2 in showing full recovery from the effects of the noise for conditions where the pedestal is 40 dB above threshold in the noise. The effect of increment duration is similar in quiet and in noise, with an improvement in detection as a function of increment duration in the range observed for detection of tones in noise.

V. EXPERIMENT 4: EFFECT OF PEDESTAL SL AND NOISE BANDWIDTH

The broadband noise conditions in the first three experiments had a detrimental effect on increment detection that could not be entirely explained on the basis of excitation patterns. Values of ΔL were comparable in quiet and in noise when low-level pedestals were equated in dB SL. At high pedestal levels, however, values of ΔL were comparable when pedestals were equated in dB SPL. The purpose of experiment 4 was to explore the effect of high-pass (HP) noise with cutoff frequencies at various points above the pedestal frequency, chosen to restrict spread of excitation while having less effect than broadband noise on threshold for the pedestal. Noise maskers that restrict spread of excitation have been shown to restore Weber's law under conditions in which pedestals and increments are equal in duration (Viemeister, 1972).

A. Subjects

The same three subjects participated as in experiments 2 and 3.

B. Stimuli and procedures

Stimulus parameters are summarized in Table I. The 4-kHz pedestal was the same as in experiments 1–3, except that it was presented at 10, 20, 30, 40, or 50 dB SL re mean threshold across subjects for the pedestal in the presence of each noise. The noise was presented at 20 dB N_0. Across conditions. the noise bandwidth was 4.8–16 kHz (4.8 HP), 5.6–16 kHz (5.6 HP), 7.2–16 kHz (7.2 HP), or 0–16 kHz (broadband). The 50-dB SL pedestal was used in all but the broadband noise condition, where the pedestal would have exceeded 90 dB SPL. The increment was the same as that in experiment 2, with a duration of 10 ms.

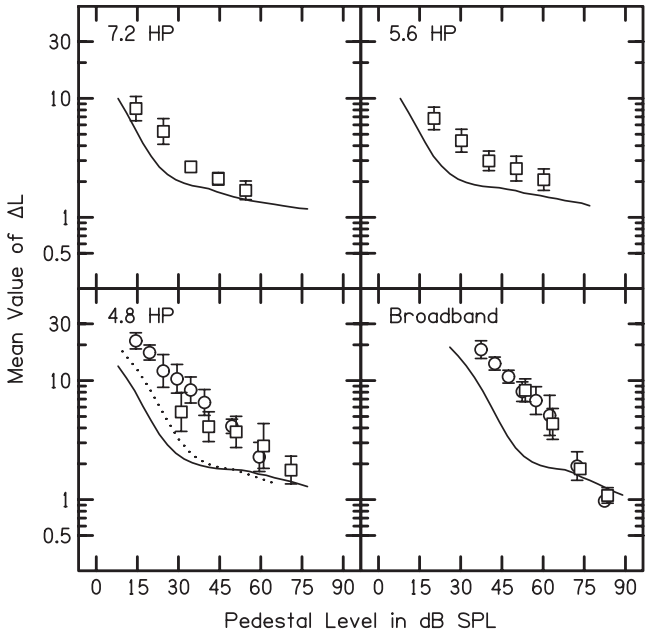


FIG. 8. Geometric mean values of ΔL across subjects for experiment 4 (shown by squares). The 10-ms, 4-kHz increment was presented 205 ms after the onset of a 370-ms, 4-kHz pedestal. Noise conditions are shown across panels. Noise spectrum level for all bandwidths was 20 dB. The solid lines show the values of ΔL predicted by the partial loudness model for each noise condition. Circles in the lower two panels represent data for the 10-ms increment obtained in experiment 2, where the maskers were broadband, but differed in level. Circles in the lower left panel are from the upper right panel in Fig. 4, where the 0-dB N_0 broadband masker produced the same amount of masking as the 4.8 HP masker. The dotted line shows the values of ΔL predicted by the partial loudness model for that broadband masker. Circles in the lower right panel are from the lower right panel in Fig. 4, where the broadband masker was 20 dB, as in experiment 4.

Two consecutive 100-trial blocks were obtained in each condition, first to determine thresholds for the pedestal alone and the increment alone in the presence of a given noise masker and then to determine thresholds for increment detection at each pedestal level in the presence of that masker. Data were obtained for the broadband noise condition first, followed by the 7.2 HP, 5.6 HP, and 4.8 HP noise conditions.

C. Results and discussion

Mean thresholds for the pedestal in each noise condition are given in Table II. Although the ERB of the auditory filter centered on 4 kHz is 0.456 kHz (Glasberg and Moore, 1990), noise bands with low-frequency cutoffs of 4.8 and 5.6 kHz overlap with the high-frequency side of the filter, resulting in some threshold elevation. Note that threshold for the pedestal in the 16-kHz, low-pass noise was comparable, as expected, to the threshold obtained in the 20-dB N_0, 8-kHz low-pass noise in experiment 2 and that the threshold for the pedestal in the 4.8–16 kHz band-pass noise was comparable to the threshold obtained in the 0-dB N_0, 8-kHz low-pass noise in experiment 2.

Increment detection results for experiment 4 are shown in Fig. 8, along with predictions from the partial loudness model (Moore *et al.*, 1997) in a format identical to that of Fig. 4. Data from the 0- and 20-dB N_0 conditions from ex-

periment 2 have been re-plotted in the lower two panels in Fig. 8 for conditions with comparable threshold shifts.

A within-subjects ANOVA of the log ΔL values for the three subjects for the four noise conditions and pedestal levels of 10, 20, 30, and 40 dB SL (50-dB SL was not available for the broadband noise condition) indicated significant effects of noise band [$F(3,6)=6.340$], pedestal level [$F(3,6)=112.073$], and a significant noise-band by pedestal-level interaction [$F(9,18)=16.972$]. The significant effect of noise band and the interaction both result from the difference between the broadband noise condition and the narrower bandwidths. Thresholds for increment detection for tones presented in broadband noise decreased more rapidly with increasing pedestal level than for tones presented in high-pass noise.

The data for the broadband noise condition in experiment 4 represent a partial replication of the 20-dB N_0 condition in experiment 2. Although a broader bandwidth was used in experiment 4, the two sets of data are in excellent agreement. Data for the 4.8 HP condition appear to show lower values of ΔL at low pedestal levels and a more gradual shift in ΔL with increasing pedestal level than data obtained in the earlier 0-dB N_0 noise condition, even though the two maskers produced similar shifts in pedestal threshold. Because the same subjects participated in both experiments, it was possible to conduct a within-subjects ANOVA of results obtained in the two noise conditions at four pedestal levels that were equal in dB SL in the two experiments. Results indicated a significant effect of pedestal level [$F(3,6)=167.779$] but no effect of noise band or noise-band by pedestal-level interaction. The dotted line in the lower left panel of Fig. 8 shows the partial-loudness-model predictions for the 0-dB N_0 broadband noise condition. Although the model predicted lower values of ΔL than the data, it showed a similar shift to a shallower function in high-pass noise. Despite the lack of a statistically significant interaction, the general agreement between model and data for the high-pass conditions provides support for the assumption that noise influences increment detection by restricting spread of excitation.

VI. DISCUSSION

Previous studies of increment detection in noise have used a limited number of pedestal-level and noise-level combinations. Either the noise level was constant in all conditions at a given frequency (Glasberg *et al.*, 2001; Moore *et al.*, 1993; Moore *et al.*, 1999; Moore and Peters, 1997), was a constant number of decibels below the pedestal (Moore *et al.*, 1996; Peters *et al.*, 1995), or the pedestal-to-noise ratio was varied for a fixed pedestal level (Oxenham, 1997). One of the primary goals of the present study was to explore the effect of varying pedestal-to-noise ratio at different noise levels. Because recent studies have focused on use of increment detection as a measure of temporal processing, they have typically used a range of pedestal frequencies and increment durations. Several studies included decrement detection as well in an effort to include increment and decrement detection in a single theoretical framework. The current studies explored the interaction of pedestal level and noise level as two factors limiting the ability to detect the increment, as in Henning (1969). The pedestal frequency, 4 kHz, was chosen to limit the influence of spectral splatter or off-frequency listening (Leshowitz and Wightman, 1971), and multiple increment durations were included only when it appeared that the interaction of pedestal and noise level might vary as a function of increment duration. The pattern of results in the first three experiments indicated that the threshold for increment detection for pedestal levels near the threshold for detection of the increment itself in noise was limited by that noise threshold. For pedestal levels more than 40 dB above the noise threshold, increment detection is determined by the physical level of the pedestal, independent of threshold. This pattern, in which the noise has an effect only for pedestals within 40 dB of the threshold for detection in noise, is similar to that observed for judgments of the loudness of a tone in noise (Humes and Jesteadt, 1991) and for thresholds reported in studies of additivity of masking (Humes and Jesteadt, 1989).

In the studies of increment detection reported by Moore and colleagues, noise spectrum levels in conditions with 4-kHz pedestals were generally 40 or 45 dB below the level of the pedestal. The data in Table II, averaged across all broadband noise conditions, indicate that a pedestal that is 40 dB above the spectrum level of the noise is at 17.1 dB SL. The data are in good agreement with Glasberg *et al.* (2001), who described their condition with a 70-dB SPL pedestal and 30-dB spectrum level noise as a condition where the pedestal was 18 dB above masked threshold. The data in Fig. 2 indicate that this is a condition where threshold for the pedestal is a controlling factor, and increment detection is uniform across noise levels for pedestal levels equated in dB SL. Increment detection would not be expected to be uniform across noise levels for pedestal levels greater than 30 dB SL where the spread of excitation grows nonlinearly. Equal-SL comparisons are useful at low SLs, but at high levels they can lead to the misleading conclusion that increment detection is better in noise than in quiet. This is, of course, never the case when the increments are considered in absolute physical units. Glasberg *et al.* (2001) reported a smaller value of ΔL in noise than in quiet with a pedestal level of 18 dB SL in the 4-kHz condition in their experiment 1. This effect was observed only at higher pedestal SLs in the current data.

Increment detection results from experiments 2 and 4 were compared to predictions from a partial loudness model (Moore *et al.*, 1997). Leibold and Jesteadt (2007) and Buss (2008) reported that this model predicts thresholds for intensity discrimination under some conditions and that it provides an excellent account of increment detection in quiet, as shown in Fig. 4. Two arguments against using the model in this way are that the relation between loudness and intensity resolution is tenuous (Schlauch and Wier, 1987; Zwislocki and Jordan, 1986) and that subjects might find it difficult, if not impossible, to judge the partial loudness of a tone added to another tone of the same frequency. There is, however, a relation between loudness and some measures of intensity resolution (Allen and Neely, 1997), and it is clear that the

model makes accurate predictions under some conditions. This is no doubt because the model is based on excitation patterns, and the improvement in increment detection with level is due to spread of excitation. Loudness per se may have little to do with the specific predictions for these stimuli.

The improvement in intensity discrimination with level, often referred to as the near miss to Weber's law (McGill and Goldberg, 1968a), is generally attributed to spread of excitation to higher frequencies (Florentine and Buus, 1981; Viemeister, 1988; Zwicker, 1956, 1970). Strong effects of level on increment detection were observed in the present data, both in quiet and in noise. Increment detection in quiet for the first three experiments could be described by two straight lines with a breakpoint between 20 and 40 dB SL and a steeper function for pedestal levels below the breakpoint. Results have been presented in units of ΔL on a log axis, but the pattern would be similar if shown in units of $10\log(\Delta I/I)$. Published functions show a comparable breakpoint for increment detection and a more pronounced breakpoint at a lower SL for intensity discrimination (Viemeister and Bacon, 1988). The partial loudness model has no temporal information and therefore predicts the same function for both increment detection and intensity discrimination. Given the observed differences between the functions for the two tasks, the model may provide a better account of increment detection than of intensity discrimination. The reason for this is unclear.

The results of experiment 4 demonstrate that, given two noises that produce equal threshold shifts for the pedestal, the noise with energy in the same frequency region as the pedestal will have a greater negative impact on increment detection. The effect may be larger than would be predicted on the basis of the excitation patterns incorporated in the partial loudness model. Given that the model is based entirely on spectral information and that increment detection is known to rely on temporal information (e.g., Glasberg *et al.*, 2001; Gallun and Hafter, 2006; Moore *et al.*, 1999; Oxenham, 1997), failure of the model to account for all aspects of the data is not surprising. Noise that falls within the critical band of the pedestal and increment may have a greater effect on temporal processing than noise located above the test frequency. The template models proposed by Moore *et al.* (1999) and Glasberg *et al.* (2001) predict such an effect because they assume that the decision process is based on the temporal properties of a single auditory filter. Glasberg *et al.* noted that noise at the output of the auditory filter centered on the pedestal and increment would have the temporal properties of a narrow-band noise, even if the input were broadband.

Modulation-filter-bank models (e.g., Dau *et al.*, 1996, 1997; Jepsen *et al.*, 2008) assume filters in the modulation frequency domain that receive input from a broad range of carrier frequencies. This assumption has been confirmed in a study of frequency selectivity for amplitude modulation (Ewert and Dau, 2000). Gallun and Hafter (2006) used a modulation filter bank model to account for increment detection as a function of increment duration and showed that modulation of a carrier at 2.013 kHz can interfere with increment detection at 0.477 kHz, a separation of more than two octaves. All of the noise conditions in experiment 4 included noise less than two octaves from the pedestal and increment, but broadband noise was clearly the most effective. More complete data obtained with high-pass or notched-noise maskers are necessary to test the assumptions of modulation-filter-bank and auditory-filter models.

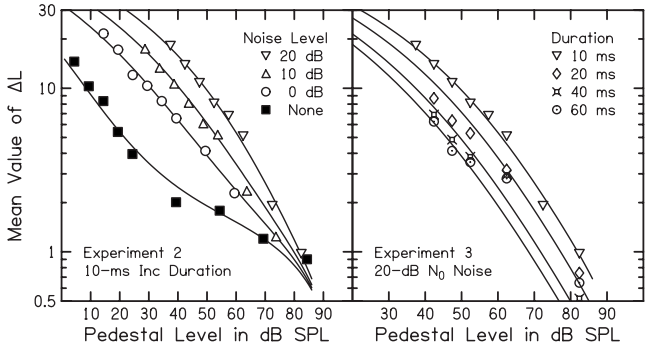


FIG. 9. Data from experiments 2 and 3 fitted by a model where the decision is based on the difference in perceptual intensity created by presentation of the increment. Performance is limited by a temporal integration process and by internal and external noise, where the internal noise includes both an additive and a multiplicative component. The model uses a three-parameter quadratic nonlinearity, a two-parameter temporal integration process, and a two-parameter internal noise. Parameter values were held constant in plotting the functions shown in the left and right panels, except that one of the quadratic-nonlinearity parameters decreased (by as much as 20%) with increasing noise.

Early multi-stage models of increment detection (Moore *et al.*, 1993, 1996; Oxenham, 1997) assumed a decision based on the maximum difference observed at the output of the temporal window, a measure referred to as ΔO. The emphasis here has been on the later models (Glasberg *et al.*, 2001; Moore *et al.*, 1999) that assumed a decision based on correlation with a temporal template for the increment, a measure more sensitive to the properties of the noise envelope. Neely and Jesteadt (2005) described a model with a decision process based on the overall difference between the two intervals, similar to ΔO. This model assumes a combination of multiplicative and additive internal noise and a quadratic nonlinearity based on loudness-matching data (Neely *et al.*, 2003). Use of a mathematically tractable nonlinearity and a temporal integration time constant rather than a temporal window allowed the authors to develop a single equation for d' that could be solved for signal level at threshold by setting $d'=1$. Variance in the decision variable due to internal noise is assumed to depend on the pedestal level and to add to the variance due to the noise component of the stimulus. Fits to the data from experiments 2 and 3 using this model are shown in Fig. 9. Note that the model predicts a different shape for the functions that describe the effect of pedestal level in quiet and in noise, whereas the partial loudness model predicts incorrectly that these functions will be similar in shape. Details of the Neely and Jesteadt (2005) model are described in the Appendix.

The lack of a strong effect of noise level at low SLs, the lack of a strong interaction between increment duration and the effect of noise, and the success of the model proposed by Neely and Jesteadt (2005), where decisions are based only on

ΔO rather than on the entire template, all argue that any template-matching or modulation detection cues are proportional to ΔO. Of the models discussed here, only the partial loudness model explicitly assumes use of the information available in higher-frequency channels, but the model described in the Appendix essentially lumps the output of the entire cochlea into a single channel. In experiment 4, the authors attempted to establish the relative contributions of higher-frequency channels to increment detection in noise by varying the frequency cutoff and bandwidth of the noise. The influence of noise bandwidth has not been explored in the model. A useful feature of the model is its potential to predict the influence on d' of any stimulus parameter. This would permit specific predictions concerning the effects of internal and external noises on the form of psychometric functions.

VII. CONCLUSIONS

At low pedestals levels, increment detection is limited by the masking effect of the noise and is similar across noise conditions for pedestals of equal SL. At high pedestal levels, the noise has no effect and increment detection is determined by the pedestal level in dB SPL.

A partial loudness model based on excitation patterns (Moore *et al.*, 1997) provides a good account of the improvement in increment detection with level observed in the absence of a noise masker but underestimates the detrimental effect of noise on increment detection thresholds.

When results are averaged across pedestal levels, the improvement in increment detection as a function of duration is similar in quiet and in noise.

Results suggest that off-frequency listening does not affect increment detection performance and that background noise is not necessary to control such effects.

A model proposed by Neely and Jesteadt (2005) provides a good fit to the data from experiments 2 and 3 by assuming that variance due to external noise adds to the variance due to internal, multiplicative noise.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. R01 DC006648, T32 DC000013, and P30 DC004662 from the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health. This work benefited from discussions with Frederick Gallun, Jessica J. Messersmith, Lori J. Leibold, Harisadhan Patra, and Donna L. Neff.

APPENDIX: QUADRATIC-COMPRESSION MODEL

Consider the transformation from physical intensity into perceptual intensity in three steps: (1) a logarithmic transformation $L=\mathcal{L}[I]$ from physical intensity I to signal level L (in decibels), (2) a quadratic-compression transformation $E=\mathcal{E}[L]$ from physical signal level L to a perceptual signal level called excitation E, and (3) an anti-logarithmic transformation $N=\mathcal{L}^{-1}[E]$ from excitation E to perceptual intensity N. The first and third steps are easily described as logarithmic expressions,

$$\mathcal{L}[I] = 10\log_{10}\left(\frac{I}{I_t}\right), \tag{A1}$$

$$\mathcal{L}^{-1}[E] = N_t 10^{E/10}, \tag{A2}$$

where I_t will be defined as the physical intensity of the signal at threshold and the perceptual intensity threshold N_t is arbitrarily set to 1. The fact that the second step is quasi-logarithmic makes the composite three-step transformation have properties similar to a single logarithmic transformation. However, all three steps are needed to allow perceptual intensity to serve as a decision variable for basic psychoacoustics tasks.

The middle step in the physical-to-perceptual intensity transformation represents the compressive response growth of the peripheral auditory system. If compression of the peripheral nonlinearity is modeled as a quadratic function of signal level (Neely and Jesteadt, 2005), then the corresponding excitation is related to signal level by

$$\mathcal{E}(L) = C - \frac{2}{q}\tanh^{-1}\left(\frac{2cL+b}{q}\right), \tag{A3}$$

where $q=\sqrt{b^2+4ac}$ and C is an arbitrary integration constant selected to make $\mathcal{E}[0]=0$. A useful feature of this nonlinear representation is that an explicit expression is available for the inverse of $\mathcal{E}$,

$$\mathcal{E}^{-1}(E) = \left(\frac{q}{2cL+b}\right)\tanh\left[\frac{q}{2}(C-E)\right]. \tag{A4}$$

The complete physical-to-perceptual intensity transformation is represented by the composite function $\mathcal{I}[I]=\mathcal{L}^{-1}[\mathcal{E}(\mathcal{L}[I])]$. The inverse transformation is simply $\mathcal{I}^{-1}[I]=\mathcal{L}^{-1}[\mathcal{E}^{-1}(\mathcal{L}[N])]$. Expressions for these two composite functions are derived by combining Eqs. (A1)–(A4). Besides the signal threshold I_t, this quadratic-compression model (QCM) has only three parameters: a, b, and c.

The QCM provides a foundation for modeling increment detection in noise. Neely and Jesteadt (2005) suggested that performance on an increment detection task may be described by

$$d' = \frac{\mathcal{I}[I_2]-\mathcal{I}[I_1]}{\sigma}, \tag{A5}$$

where I_1 is the intensity of the pedestal, I_2 is the incremented intensity, and σ^2 is the variance associated with the decision variable $\mathcal{I}[I_2]-\mathcal{I}[I_1]$. This variance represents the combined influence of internal noise, external noise, and temporal integration,

$$\sigma^2 = \frac{n_0^2 + n_1^2\mathcal{I}[I_1] + n_3^2\mathcal{I}^2[I_3]}{1-a_i e^{-\tau_s/\tau_i}}. \tag{A6}$$

In this equation, I_3 is the intensity of the external noise and τ_s is the duration of the increment. Equation (A6) introduces six additional model parameters: Two parameters (n_0 and n_1) describe the influence of internal noise on detection, two parameters (a_i and τ_i) represent temporal integration and describe the influence of increment duration on detection, and two parameters (L_{3t} and n_3) specify the external noise thresh-

TABLE IV. Estimated values of parameters for the model proposed by Neely and Jesteadt (2005) used to generate functions in Fig. 9.

Model parameters	
QCM in quiet	$a=0.6$, $b=0.1$, $c=-0.000\,64$
Temporal integration	$a_i=0.99$, $\tau_i=450$ ms
Internal noise	$n_0=2$, $n_1=0.25$
External noise	$L_{3t}=-42$, $n_3=0.224$
CGR with external noise	$b=0.090$, 0.085, 0.080

old level (for the purpose of determining its perceptual intensity) and how the external noise combines with internal noise.

If $\Delta L=\mathcal{L}(I_2/I_1)$ is defined and Eq. (A5) is used to derive an expression for I_2, then

$$\Delta L=\mathcal{L}[\mathcal{I}^{-1}(\mathcal{I}[I_1]+\sigma d')]-\mathcal{L}[I_1]. \quad \text{(A7)}$$

The three QCM parameters may be independently determined by requiring consistency with an intensity discrimination task (Jesteadt *et al.*, 1977). The two temporal integration parameters may be independently determined by requiring consistency with a single-tone detection task (Florentine *et al.*, 1988). The pedestal threshold was selected as the reference intensity for all three stimulus intensities. This leaves only two model parameters to describe increment detection in quiet and one additional parameter to describe the strength of the external noise contribution. To provide a better fit to the data for pedestal levels >60 dB SPL, the QCM model that describes compression growth rate (CGR) was decreased (by as much as 20%) in the presence of external noise. Estimated parameter values are summarized in Table IV.

Figure 9 shows ΔL as a function of I_1 (when $d'=1$) for several values of I_3 on the left and several values of τ_s on the right. In the absence of external noise (lowest line in left panel), n_0 determines threshold for pedestal levels below 30 dB SPL, while n_1 determines threshold for pedestal levels above 30 dB SPL. In the presence of external noise (upper lines in left panel), the main effect of decreasing parameter b is to decrease threshold more rapidly for pedestal levels above 60 dB SPL. Increment duration had less effect on model threshold (right panel) than observed in the data. Overall, the QCM model reproduces all of the major trends observed in the data.

Allen, J. B., and Neely, S. T. (**1997**). "Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise," J. Acoust. Soc. Am. **102**, 3628–3646.

Buss, E. (**2008**). "The effect of masker level uncertainty on intensity discrimination," J. Acoust. Soc. Am. **123**, 254–264.

Buus, S., and Florentine, M. (**1991**). "Psychometric functions for level discrimination," J. Acoust. Soc. Am. **90**, 1371–1380.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). "Modeling auditory processing of amplitude modulation I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Puschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Florentine, M. (**1986**). "Level discrimination of tones as a function of duration," J. Acoust. Soc. Am. **79**, 792–798.

Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.

Florentine, M., Fastl, H., and Buus, S. (**1988**). "Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking," J. Acoust. Soc. Am. **84**, 195–203.

Gallun, F. J., and Hafter, E. R. (**2006**). "Amplitude modulation sensitivity as a mechanism for increment detection," J. Acoust. Soc. Am. **119**, 3919–3930.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Glasberg, B. R., Moore, B. C., and Peters, R. W. (**2001**). "The influence of external and internal noise on the detection of increments and decrements in the level of sinusoids," Hear. Res. **155**, 41–53.

Hanna, T. E., von Gierke, S. M., and Green, D. M. (**1986**). "Detection and intensity discrimination of a sinusoid," J. Acoust. Soc. Am. **80**, 1335–1340.

Heinz, M. G., and Formby, C. (**1999**). "Detection of time- and bandlimited increments and decrements in a random-level noise," J. Acoust. Soc. Am. **106**, 313–326.

Henning, G. B. (**1969**). "Amplitude discrimination in noise, pedestal experiments, and additivity of masking," J. Acoust. Soc. Am. **45**, 426–435.

Humes, L. E., and Jesteadt, W. (**1989**). "Models of the additivity of masking," J. Acoust. Soc. Am. **85**, 1285–1294.

Humes, L. E., and Jesteadt, W. (**1991**). "Models of the effects of threshold on loudness growth and summation," J. Acoust. Soc. Am. **90**, 1933–1943.

Jepsen, M. L., Ewert, S. D., and Dau, T. (**2008**). "A computational model of human auditory signal processing and perception," J. Acoust. Soc. Am. **124**, 422–438.

Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.

Jesteadt, W., Schairer, K. S., and Neff, D. L. (**2005**). "Effect of variability in level on forward masking and on increment detection," J. Acoust. Soc. Am. **118**, 325–337.

Jesteadt, W., Tan, H., Khaddam, S., and Leibold, L. J. (**2007**). "Prediction of behavioral thresholds using a model of partial loudness," in 30th Midwinter Research Meeting of the Association for Research in Otolaryngology.

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). "Intensity discrimination as a function of frequency and sensation level," J. Acoust. Soc. Am. **61**, 169–177.

Leibold, L. J., and Jesteadt, W. (**2007**). "Use of perceptual weights to test a model of loudness summation," J. Acoust. Soc. Am. **122**, EL69–EL73.

Leshowitz, B., and Wightman, F. L. (**1971**). "On-frequency masking with continuous sinusoids," J. Acoust. Soc. Am. **49**, 1180–1190.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

McGill, W. J., and Goldberg, J. P. (**1968a**). "A study of the near-miss involving Weber's law and pure-tone intensity discrimination," Percept. Psychophys. **4**, 105–109.

McGill, W. J., and Goldberg, J. P. (**1968b**). "Pure-tone intensity discrimination and energy detection," J. Acoust. Soc. Am. **44**, 576–581.

Moore, B. C., 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., and Peters, R. W. (**1997**). "Detection of increments and decrements in sinusoids as a function of frequency, increment, and decrement duration and pedestal duration," J. Acoust. Soc. Am. **102**, 2954–2965.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1993**). "Effects of frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **94**, 3190–3198.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1996**). "Detection of decrements and increments in sinusoids at high overall levels," J. Acoust. Soc. Am. **99**, 3669–3677.

Moore, B. C., Peters, R. W., and Glasberg, B. R. (**1999**). "Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise," J. Acoust. Soc. Am. **106**, 3539–3552.

Neely, S. T., and Jesteadt, W. (**2005**). "Quadratic-compression model of auditory discrimination and detection," Acta. Acust. Acust. **91**, 980–991.

Neely, S. T., Schairer, K. S., and Jesteadt, W. (**2003**). "Estimates of cochlear compression from measurements of loudness growth," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigné, S. McAdams, and L. Collet (Springer, New York), pp. 42–56.

Neff, D. L., and Jesteadt, W. (**1996**). "Intensity discrimination in the pres-

ence of random-frequency, multicomponent maskers and broadband noise," J. Acoust. Soc. Am. **100**, 2289–2298.
Oxenham, A. J. (**1997**). "Increment and decrement detection in sinusoids as a measure of temporal resolution," J. Acoust. Soc. Am. **102**, 1779–1790.
Peters, R. W., Moore, B. C., and Glasberg, B. R. (**1995**). "Effects of level and frequency on the detection of decrements and increments in sinusoids," J. Acoust. Soc. Am. **97**, 3791–3799.
Schlauch, R. S., and Wier, C. C. (**1987**). "A method for relating loudness-matching and intensity discrimination data," J. Speech Hear. Res. **30**, 13–20.
Viemeister, N. F. (**1972**). "Intensity discrimination of pulsed sinusoids: The effects of filtered noise," J. Acoust. Soc. Am. **51**, 1265–1269.
Viemeister, N. F. (**1988**). "Psychophysical aspects of auditory intensity coding," in *Auditory Function: Neurobiological Bases of Hearing*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 213–241.
Viemeister, N. F., and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**, 172–178.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
Wojtczak, M., and Viemeister, N. F. (**1999**). "Intensity discrimination and detection of amplitude modulation," J. Acoust. Soc. Am. **106**, 1917–1924.
Zwicker, E. (**1956**). "Die elemenataren grundlagen zur bestimmung der informationskapazitat des gehors," Acustica **5**, 365–381.
Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (A.W. Sijthoff, Leiden), pp. 376–396.
Zwislocki, J. J., and Jordan, H. N. (**1986**). "On the relations of intensity jnd's to loudness and neural noise," J. Acoust. Soc. Am. **79**, 772–780.

Detection probability of vocalizing dugongs during playback of conspecific calls

Kotaro Ichikawa[a)]
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

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

Tomio Shinke
R&D Center, System Intech Co., Ltd., 424-8610 Shizuoka, Japan

Kotoe Sasamori
Marine Wildlife Center of Japan, 093-0084 Hokkaido, Japan

Yukio Miyauchi and Yuki Abe
Choshi Ocean Institute, 288-0014 Chiba, Japan

Kanjana Adulyanukosol
Phuket Marine Biological Center, 83000 Phuket, Thailand

Nobuaki Arai
Graduate School of Informatics, Kyoto University, 606-8501 Kyoto, Japan

(Received 7 February 2009; revised 30 June 2009; accepted 14 July 2009)

Dugongs (*Dugong dugon*) were monitored using simultaneous passive acoustic methods and visual observations in Thai waters during January 2008. Chirp and trill calls were detected by a towed stereo hydrophone array system. Two teams of experienced observers conducted standard visual observations on the same boat. Comparisons of detection probabilities of acoustic and visual monitoring between two independent observers were calculated. Acoustic and visual detection probabilities were 15.1% and 15.7%, respectively, employing a 300 s matching time interval. When conspecific chirp calls were broadcast from an underwater speaker deployed on the side of the observation boat, the detection probability of acoustic monitoring rose to 19.2%. The visual detection probability was 12.5%. Vocal hot spots characterized by frequent acoustic detection of calls were suggested by dispersion analysis, while dugongs were visually observed constantly throughout the focal area ($p<0.001$). Passive acoustic monitoring assisted the survey since detection performance similar to that of experienced visual observers was shown. Playback of conspecific chirps appeared to increase the detection probability, which could be beneficial for future field surveys using passive acoustics in order to ensure the attendance of dugongs in the focal area. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3203805]

PACS number(s): 43.66.Gf, 43.80.Ev, 43.80.Ka, 43.80.Nd [WWA] Pages: 1954–1959

I. INTRODUCTION

Dugongs, *Dugong dugon*, are endangered herbivorous marine mammals inhabiting tropical and subtropical shallow waters that include the Indian and Pacific Oceans. Although the world's largest dugong population is found in Australian waters, the local populations in many other locations are becoming endangered (e.g., Marsh *et al.*, 2002; Preen, 1998; Preen, 2004). Dugongs feed on shallow sea grass beds, spending 72% of their day in waters less than 3 m deep (Chilvers *et al.*, 2004; De Iongh *et al.*, 2007). Underwater dugongs are visible in shallow areas such as feeding grounds. Previously in Thailand, several visual surveys were conducted to estimate the total population of about 120 animals in Trang Province (Hines *et al.*, 2005), and acoustic observations of the dugongs suggested tidal and diel effect on their acoustic activities (Ichikawa *et al.*, 2006; Tsutsumi *et al.*, 2006). Although those previous works provided important information of the dugongs, the detection performances of the visual and acoustic observation remain to be evaluated.

Aerial surveys are an efficient method for surveying dugongs (Pollock *et al.*, 2006; Marsh *et al.*, 2004; Hines *et al.*, 2005; Adulyanukosol *et al.*, 2007) and are commonly used because they are quick and cover a wide range. Aerial surveys have some limitations. First, aerial observers can miss dugongs in deep water areas; this could cause a depth-dependent bias. Dugongs have often been observed offshore where the bottom was not visible. Second, independent observations are difficult from small airplanes that prevent *g*-zero estimation. *g*-zero is the detection probability of target

[a)]Author to whom correspondence should be addressed. Electronic mail: ichikawa@bre.soc.i.kyoto-u.ac.jp

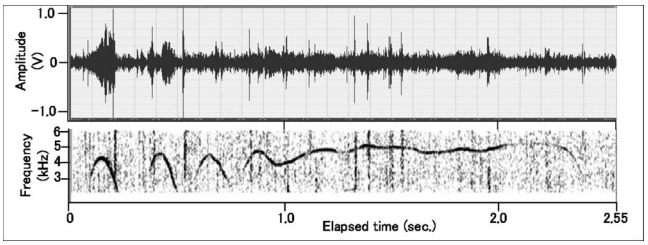


FIG. 1. A sonogram (bottom inset) and a wave form (upper inset) of typical dugong calls including three chirp calls and a trill call recorded in the study area. Chirp calls, which were the most frequent in the study area, had an average frequency of 4521 ± 1615 Hz and an average duration of 126 ± 87 ms ($n=704$). Trill calls were characterized by a frequency of 4152 ± 1111 Hz and a duration of 1737 ± 1049 ms ($n=74$) (Ichikawa *et al.*, 2003).

species on a survey line and is commonly assumed to be 1 since most targets can be observed. For marine mammal observations, this assumption is not usually satisfied. To compensate for this bias, *g*-zero can be estimated by comparing primary and independent visual observations (Buckland *et al.*, 1993). Third, habitats of dugongs are not always accessible by airplane. Thus, a long-term methodology that is less costly is preferable in remote areas with little financial support for research.

Ship-based visual observation is a standard method for surveying surfaced marine mammals. These surveys can be conducted in very deep waters since marine mammals must return to the surface to breathe. Primary and independent observers are commonly used to estimate abundance in ship-based observations. Moreover, ship- or boat-based observation is well established and widely used even in remote areas. The problem with this method for surveying dugongs lies in the difficulty in observing them from either land or ship. They surface only once or twice at a time, exposing only their nose and keeping the rest of their body underwater. Their dive duration is approximately 150 s (Chilvers *et al.*, 2004).

Over the past several years, passive acoustic methods have been widely applied for monitoring dolphins (Philpott *et al.*, 2007), porpoises (Wang *et al.*, 2005), baleen whales (Stafford *et al.*, 2007), and even for sirenians such as manatees (Phillips *et al.*, 2004) and dugongs (Ichikawa *et al.*, 2006; Tsutsumi *et al.*, 2006). Dugongs produce several types of sounds such as chirps, trills, and barks. Chirps are frequency-modulated narrow-band sounds that are often produced in sequence (Anderson and Barclay, 1995; Ichikawa *et al.*, 2006). Chirps start in an up-sweep and then turn into down-sweeps within a couple of hundred milliseconds. Trill sounds are long in duration and often last for more than 1 s, with an up and down frequency modulation that sounds like a vibration. Barks are a low frequency broadband sound that is not observed as frequently as the other two vocalization types. Chirp calls, which were the most frequent in the study area, had an average frequency of 4521 ± 1615 Hz and an average duration of 126 ± 87 ms ($n=704$). Trill calls were characterized by a frequency of 4152 ± 1111 Hz and a duration of 1737 ± 1049 ms ($n=74$) (Ichikawa *et al.*, 2003). Chirps and trills are appropriate targets for passive acoustic monitoring methods for dugongs (Fig. 1).

Data obtained from passive acoustics can be considered to be independent from visual detection. Pollock *et al.* (2006) compared visual detection of dugongs by two individual observers on the same platform and developed a model that estimated detection probabilities. Identical observation conditions resulted in overestimation of detection probability due to high correlation between detection probabilities of the two observers, thus resulting in an underestimation of the population size. Acoustics can be used as an independent observation method to estimate the detection probability of dugongs. Simultaneous observation of porpoises with visual and acoustic methods enables the calculation of detection probabilities for both techniques when both types of detection occur simultaneously (Akamatsu *et al.*, 2008). It should be noted that the constant vocalization as of the porpoises in Akamatsu *et al.* (2008) enhances accurate calculation of the detection probability. The authors modified this method for low frequency tonal sounds of wild dugongs, compared the detection probability to that of visual detection in Thai waters, and attempted to improve the detection probability by using a playback of the dugong vocalizations. The authors also evaluated constancy of both of the detections.

II. MATERIALS AND METHODS

A. Acoustic observations

A towed stereo hydrophone system (Towed Aquafeeler, System Intech Co. Ltd., Japan) was operated off of Trang Province, Thailand. The study area for the towed system was from Muk Island (7, 12′.28″N, 99, 23′.56″E) to the southwest end of Talibong island (7, 24′.31″N, 99, 20′.11″E) via Hat Yao port (7, 18′.4″N, 99, 24′.0″E). The authors operated a wooden boat (11.3 m in length and 2.4 m in width) for daily trips around the focal area. The distance of this trip was 60 km and took 6 h a day at a towing speed of 10 km/h. The cruise lines from north to south were separated by approximately 1 km in order to cover a wider range of the focal area, specifically inshore and offshore lines. Each day, the authors randomly selected either an inshore or offshore line to begin surveys.

The Towed Aquafeeler consisted of a 10 m draw, a 4 m flexible polyvinyl chloride rubber tube with two hydrophone elements (100 Hz–100 kHz) inside, a 60 m towed electric cable, a receiving unit, and a two-channel conditioning amplifier. The towed cable eliminated interfering noise from the towing boat. The cable had a neutral buoyancy that enabled towing even in shallow waters of 1 m depth, which is close to the minimum depth of 0.8 m that dugong forage in (Tsutsumi *et al.*, 2006). The stereo hydrophone (two hydrophone elements) was separated by 2 m from the preamplifier near the hydrophone element. The receiving sensitivity of the hydrophone was −193 dB (re 1 V/μPa). The amplifier had a variable high-pass filter (cutoff frequencies of 200 Hz, 1 kHz, and 4 kHz). In the present study, the authors selected a 1 kHz high-pass filter to eliminate flow noise interference. Stereo signals were recorded using a hard disk recorder (R-4 pro, Roland, Japan).

Two experienced audio listeners (T.A. and T.S., authors of this paper) monitored underwater sounds using head-

phones (MDR-Z600, SONY, Japan). These listeners had experience in listening to several dozen hours of recorded sounds and in identifying dugong vocalizations. The two listeners took turns listening, alternating every 30 min. The time at which dugong calls were received, with a 1 s resolution, was recorded by the listener. Two successive chirps or one long trill over a 1 s period were defined as a detection, and if these sounds were not heard, no record was made. Automatic track logging of the GPS (GPS 76s, Garmin, USA) was used to assess the location with respect to the detection time. Acoustic observers sat below the tower that the visual observers surveyed from so that they were out of sight of each other. Underwater sounds were amplified to a volume that prevented the listener from hearing the voice of the other observers on board. No exchange of cues between acoustic and visual observers was permitted. Onboard detections were confirmed by off-line listening. T.A. checked the recording of each detection with COOL EDIT PRO software (Syntrillium Software Corp., AZ) by listening to the sound and looking at the sonogram to exclude false positive detections. The detection threshold level was 90 dB rms re 1 μPa using a 1024-point fast Fourier transform.

Dugongs are known to call back to recorded conspecific chirp sounds, as indicated in an unpublished reference. The authors broadcasted the chirp sound from the observation boat using an underwater speaker (U.S.300, FOSTEX, Japan). Playback and non-playback experiments were conducted every other day. Vocal rates of dugongs have been suggested to be affected by the tidal cycle (Ichikawa *et al.*, 2006). To control this effect, playback of the chirps was continued for the whole day round trip from Muk Island to Talibong Island. Playback sounds were recorded from Talibong Island at the west end of the focal area of the present study. A digital bandpass filter that passes all frequency components between 2 and 5 kHz was applied to reduce the noise of the test sound in advance of the playback. The source level was set at 150 dB rms re 1 μPa.

Ichikawa (2007) estimated the source level of dugong calls to be approximately 141.6 ± 4.6 dB with a mode value of 138 dB using spherical spreading model assumption for the transmission loss. Given the mode of the source level of dugong calls and the detection threshold level, the acoustic detection range was calculated to be 251.2 m.

B. Visual observations

Four experienced visual observers performed visual observations, conducted without magnification. Two were professional dolphin watching guides, and the other two were researchers who have been conducting visual observations of dugongs in Thai waters for many years. Observers focused on the sector 90° to the left of the bow during first 30 min and then focused on the 90° sector on the right side for another 30 min, followed by 1 h of rest. The order of the observers was determined randomly every day. The observation platform was constructed on the boat and was 1.84 m in height with a shade-providing roof to prevent dazzling. Observers recorded species, distance, the angle relative to the bow, and the group size. The reference angle of 0° was provided by the direction of the bow, which is consistent with the end fire direction of the towed hydrophone. Humpback dolphins were occasionally observed in the study area, and records of their presence were used to exclude contamination by dolphin whistles in audio recordings. Dolphin whistles did not show a regular sequence like the calls of dugongs. Detection of non-sequential sounds corresponded with the visual detection of dolphins. Therefore, the exclusion of dolphin sounds from the acoustic data was easy.

C. Calculation of detection probabilities

The authors applied the method of Akamatsu *et al.* (2008) to estimate the detection probabilities by comparing the detections by two individual observations. All detections fell into a time interval ranging from 20 to 1000 s during the entire observation period. If more than one vocal sequence was detected within the time interval, that time interval was categorized as acoustically positive. If any visual observation of dugongs was made in a particular time interval, that time interval was also categorized as being visually positive. The acoustical detection probability (P_a) was defined as the ratio between the number of acoustically positive intervals (N_a) and the number of intervals when the dugong was actually present (N). The latter factor N could not be measured directly. However, if the authors assume an entire set of the visual detections (N_v) as another sampling population, detection probability of acoustic observations (P_a_estimate) can be estimated from the number of both acoustically and visually positive detections (N_m) over the total number of visual detections (N_v). Based on the idea of sampling statistics, if the distribution of individuals were completely uniform, P_a_estimate would be equal to P_a. The detection probability of visual observations (P_v_estimate) can also be obtained as N_m divided by N_a [Eqs. (1) and (2)]; see details in Akamatsu *et al.* (2008),

$$P_a\_\text{estimate} = N_m/N_v \tag{1}$$

and

$$P_v\_\text{estimate} = N_m/N_a. \tag{2}$$

D. Analysis of dispersion of detections

The accuracy of the estimation of the detection probability depends on the distribution pattern of detections. The authors segmented the transecting line into 833 m sample blocks, equivalent to 300 s time intervals, and tallied both acoustic and visual detections for each block. Analysis of dispersion using an I_δ-index was performed to examine the spatial distribution of locations where detections were made. The I_δ-index was proposed by Morisita (1962) as a measure of the dispersion of individuals in a population. The statistical significance of the index value was tested by F $(b-1,\infty;0.01)$. The index value and F value for a given group of N individuals was computed as

$$I_\delta = \sum x_i(x_i-1)/N(N-1)$$

and

$$F = (I_\delta(N-1) + b - N)/(b-1),$$

where x_i is the number of individuals in the ith sample block of the total b blocks ($i=1,2,3,4,\ldots,b$). If the index is greater (smaller) than 1, then the distribution is concentrated (uniform). An index equal to 1 indicates a random distribution.

III. RESULTS

Surveys were conducted for 12 days from January 11 to 23 in 2008. In total, 85 dugongs were observed visually. On average, seven animals per day were observed along the 60 km cruise. The maximum detection distance of visual observations was 220 m from the observer to the animal. In addition, three dugongs were observed by off-duty observers. These observations were not used for the analysis.

Acoustic surveys detected sequences of dugong vocalizations 237 times. Once a dugong started to vocalize, the sequence of chirps and trills was recognized, which is consistent with former observations in these waters. Unlike observations of finless porpoises by Akamatsu *et al.* (2008), counting of the number of animals was not possible because the bearing angle of the sound source was not clear in most cases due to a low signal to noise ratio.

A. Detection probability affected by playback

Observations with and without playbacks were conducted for 5 and 7 days, respectively. The number of acoustically and visually positive time intervals depended on the duration of the interval. Using control data without playback sessions, the numbers of acoustically and visually positive intervals were 111 and 54, respectively, when the duration of the interval was 20 s. These values were reduced to 28 and 49 with a 1000-s interval because the total number of intervals was reduced.

Estimated detection probabilities are shown in Fig. 2. The estimated detection probability increased as the time interval became longer. The detection probability of the passive acoustic method with a playback signal was about 1.3 times greater than that without a playback sound in any given time interval [Fig. 2(a)]. For both acoustic and visual detection, comparisons of regression slopes with and without playback showed statistical differences ($p<0.01$, $t=10.32$ and 6.36, $n=50$ for both, and two-tailed t test).

B. Dispersion of detections

Areas surveyed with and without playbacks were segmented into 33 and 34 sample blocks, respectively. Acoustically positive intervals both with and without playback were concentrated in two specific areas that were located on the east side of Muk Island and on the south side of Talibong Island ($p<0.01$, $I_\delta=2.73$ and 3.52, $F=6.88$ and 21.08, and $n=33$ and 34, respectively). The distribution of visually positive intervals with playbacks was not significantly concentrated ($p>0.05$, $I_\delta=1.43$, $F=1.34$, and $n=33$), and intervals without playback showed an almost random distribution ($p>0.05$, $I_\delta=0.97$, $F=1.07$, and $n=34$). The I_δ-index of the visual detections increased slightly during playback.

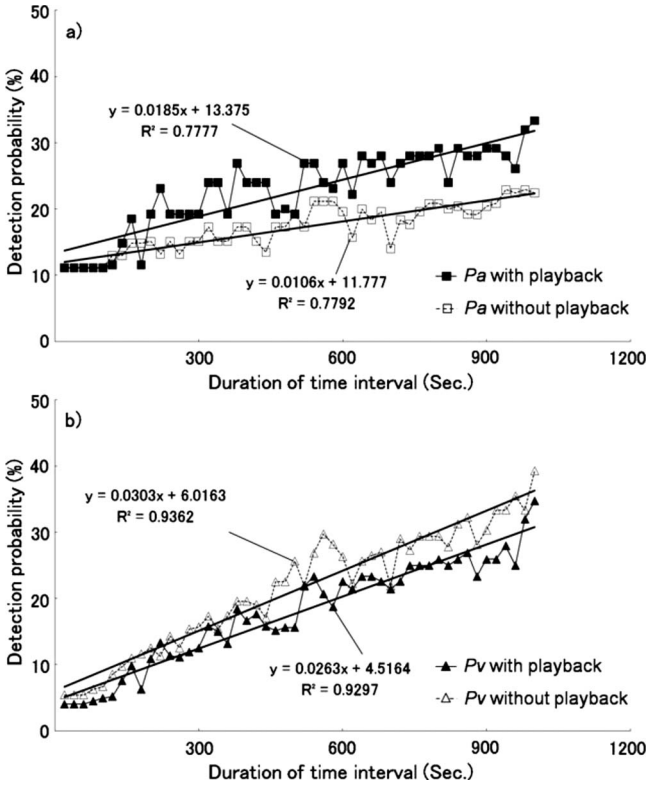


FIG. 2. Detection probabilities of (a) acoustic and (b) visual observations. The solid and empty squares (triangles) indicate acoustic (visual) detection probabilities with and without playback, respectively. For both acoustic and visual observations, statistically significant differences were found between the slopes of detection probabilities with and without playback ($p<0.01$). The playback of a conspecific chirp was effective for improving the performance of acoustic detection.

IV. DISCUSSION

Detection probabilities for both acoustic and visual observations of dugongs were estimated, with the detection ranges of both observations being well-matched. Playback significantly improved the acoustic detection probability. As indicated in an unpublished reference, dugongs responded by calling back to the playback of conspecific calls. The authors' result is consistent with the former experiment conducted with stationary playbacks. Detection probabilities of dugongs in the present study depended on the length of the time interval. Longer time intervals provided a larger detection probability. In the present study, the appropriate time interval was considered to be twice the average respiration interval of dugongs. Using a 300 s time interval, the detection probabilities of acoustic and visual observations were 15.1% and 15.7%, respectively. These probabilities were 19.2% and 12.5% during playback sessions. Passive acoustic methods and visual observations from experienced observers showed similar performance, and the acoustic detections were enhanced by the playback of conspecific calls.

Two major limitations should be noted for passive acoustic surveys of dugongs. The first is the uncertain matching of animals by visual and acoustic methods. There is also the possibility that matching visual and acoustic detections may not actually detect the same animal. The 300 s time interval corresponds to an 833 m interval on the survey line

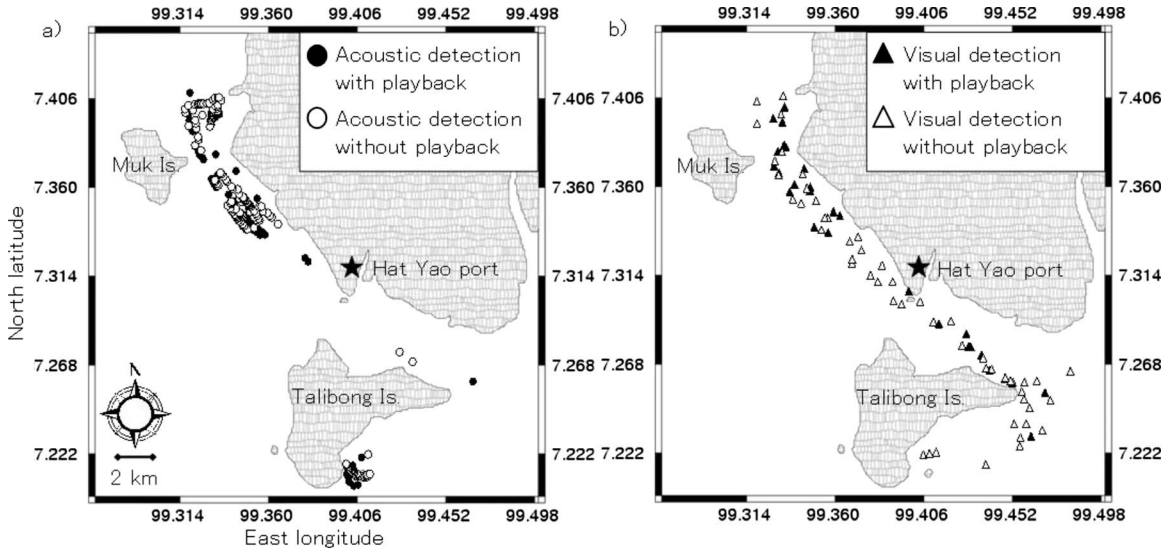


FIG. 3. Dispersion of (a) acoustic and (b) visual detections. The solid and empty squares (triangles) indicate acoustic (visual) detection with and without playback, respectively. Acoustic detections were concentrated in two specific areas, east of Muk Island and south of Talibong Island ($p<0.01$). Visual detections were not significantly concentrated ($p>0.05$) and were almost random when playbacks were not utilized.

at a speed of 10 km/h Thus, detection within the 300 s time interval used in the present study is interpreted as the probability that one or more dugongs are present in that 833 m of space. If the authors considered a longer time interval of 6 h, which was the entire observation duration of each day, the detection probability would be 1 since they detected at least one or more animals in a day during a 60 km cruise. Thus, at least one or more dugongs were in the surveyed area. Using acoustically/visually positive intervals is useful in discriminating absence or presence, which is equivalent to detecting the presence of individuals within the interval. Unlike previous results with finless porpoises (Akamatsu *et al.*, 2008), the authors could not count the number of animals using a sound bearing angle due to the low signal to noise ratio. The dugongs were found to call back to conspecific calls by an unpublished reference. Then, because the vocalization rate of an individual is not always constant, the number of detected calls is not proportional to the number of individuals. Therefore, acoustic detection only shows the presence of one or more animals. To apply a system with low detection performance, a relatively long section can be used to establish the presence/absence of animals. In the present study, each 833 m section was used as the unit section and associated with the presence or absence of dugongs as determined by acoustic or visual observation. Thus, the use of passive acoustic surveys for dugongs was more suitable for assessing the presence of animals rather than counting the number of animals in the focal area. Assessing presence or absence is the first step in the conservation of highly endangered aquatic animals. However, it should be noted that passive acoustic surveys should not be used to confirm the absence of endangered animals.

The second limitation of the passive acoustic method used is the locality of acoustical detection. The analysis of dispersion suggests distinctive vocal "hot spots" despite uniform visual detection (Fig. 3). Pollock *et al.* (2006) noted that the availability of animals for observation should be taken into account. It should thus be noted that acoustic availability may depend on the location, especially at the east end of Talibong Island. The visual observers covered distances of up to 220 m from the cruise line, and the estimated acoustic detection range was 251.2 m from the hydrophone. Although the detection performance of both methods should be low close to the limit of the observable range, the detection range of the acoustic method was likely wider than that of visual observation.

ACKNOWLEDGMENTS

The authors are grateful for the kind acceptance of this study by the National Research Council of Thailand and National Park Service of Thailand. They also thank Anuwat Prasitthipornkul, Sahdee Yakol, Sumen, Surasak Thongsukdee, and Mark from the Phuket Marine Biological Center, Akihiko Matsuda from the National Research Institute of Fisheries Engineering, and Takeshi Hara and Hideharu Tamori from the Japan Fisheries Resource Conservation Association. Nanako Amamoto and Rika Shiraki, students of Kyoto University, have also provided us with great support. The acoustic research portion of the expedition was supported by the Research and Development Program for New Bioindustry Initiatives, Grants-in-Aid for Scientific Research (B) 19405005, and the Fisheries Research Agency of Japan.

Adulyanukosol, K., Thongsukdee, S., Hara, T., Arai, N., and Tsuchiya, M. (**2007**). "Observations of dugong reproductive behavior in Trang Province, Thailand: Further evidence of intraspecific variation in dugong behavior," Mar. Biol. (Berlin) **151**, 1887–1891.

Akamatsu, T., Wang, D., Wang, K., Li, S., Dong, S., Zhao, X., Barlow, J., Stewart, B. S., and Richlen, M. (**2008**). "Estimation of the detection probability for Yangtze finless porpoises (*Neophocaena phocaenoides asiaeorientalis*) with a passive acoustic method," J. Acoust. Soc. Am. **123**, 4403–4411.

Anderson, P. K., and Barclay, R. M. R. (**1995**). "Acoustic signals of solitary dugongs: Physical characteristics and behavioral correlates," J. Mammal. **76**, 1226–1237.

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (**1993**).

Distance Sampling: Estimating Abundance of Biological Populations (Chapman and Hall, London), pp. 200–217.
Chilvers, B. L., Delean, S., Gales, N. J., Holley, D. K., Lawler, I. R., Marsh, H., and Preen, A. R. (**2004**). "Diving behavior of dugongs, *Dugong dugon*," J. Exp. Mar. Biol. Ecol. **304**, 203–224.
De Iongh, H. H., Kiswara, W., Kustiawan, W., and Loth, P. E. (**2007**). "A review of research on the interactions between dugongs (Dugong dugon Müller 1776) and intertidal seagrass beds in Indonesia," Hydrobiologia **591**, 73–83.
Hines, E. M., Adulyanukosol, K., and Duffus, D. A. (**2005**). "Dugong (*Dugong dugon*) abundance along the Andaman coast of Thailand," Marine Mammal Sci. **21**, 536–549.
Ichikawa, K. (**2007**). "Studies of behavioral ecology and protection measures of the dugongs based on analyses of marine acoustical information," Ph.D. dissertation, Department of Informatics, Kyoto University, Kyoto, Japan.
Ichikawa, K., Akamatsu, T., Shinke, T., Arai, N., Hara, T., and Adulyanukosol, K. (**2003**). "Acoustical analyses on the calls of dugong," in Proceedings of the Fourth SEASTAR 2000 Workshop, pp. 72–76.
Ichikawa, K., Tsutsumi, C., Arai, N., Akamatsu, T., Shinke, T., Hara, T., and Adulyanukosol, K. (**2006**). "Dugong (*Dugong dugon*) vocalization patterns recorded by automatic underwater sound monitoring systems," J. Acoust. Soc. Am. **119**, 3726–3733.
Marsh, H., Eros, C., Penrose, H., and Hugues, J. (**2002**). "Dugong status report and action plans for countries and territories," UNEP Early Warning and Assessment Report Series 1, Thailand, 59-67, p. 162.
Marsh, H., Lawler, I. R., Kwan, D., Delean, S., Pollock, K., and Alldredge, M. (**2004**). "Aerial surveys and the potential biological removal technique indicate that the Torres Strait dugong fishery is unsustainable," Anim. Conserv. **7**, 435–443.
Morisita, M. (**1962**). "Iδ-index, a measurement of dispersal of individuals," Res. Popul. Ecol. (Kyoto) **4**, 1–7.
Phillips, R., Niezrecki, C., and Beusse, D. O. (**2004**). "Determination of West Indian manatee vocalization levels and rate," J. Acoust. Soc. Am. **115**, 422–428.
Philpott, E., Englund, A., Ingram, S., and Rogan, E. (**2007**). "Using T-PODs to investigate the echolocation of coastal bottlenose dolphins," J. Mar. Biol. Assoc. U.K. **87**, 11–17.
Pollock, K. H., Marsh, H. D., Lawler, I. R., and Alldredge, M. W. (**2006**). "Estimating animal abundance in heterogeneous environments: An application to aerial surveys for dugongs," J. Wildl. Manage. **70**, 255–262.
Preen, A. (**1998**). "Marine protected areas and dugong conservation along Australia's Indian Ocean coast," Environ. Manage. (N.Y.) **22**, 173–181.
Preen, A. (**2004**). "Distribution, abundance and conservation status of dugongs and dolphins in the southern and western Arabian Gulf," Biol. Conserv. **118**, 205–218.
Stafford, K. M., Mellinger, D. K., Moore, S. E., and Fox, C. G. (**2007**). "Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002," J. Acoust. Soc. Am. **122**, 3378–3390.
Tsutsumi, C., Akamatsu, T., Shinke, T., and Adulyanukosol, K. (**2006**). "Feeding behavior of wild dugongs monitored by a passive acoustic method," J. Acoust. Soc. Am. **120**, 1356–1360.
Wang, K., Wang, D., Akamatsu, T., Li, S., and Xiao, J. (**2005**). "A passive acoustical monitoring method applied to observation and group size estimation of finless porpoises," J. Acoust. Soc. Am. **118**, 1180–1185.

Integration of auditory and vibrotactile stimuli: Effects of phase and stimulus-onset asynchrony

E. Courtenay Wilson, Charlotte M. Reed, and Louis D. Braida
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Speech and Hearing Bioscience and Technology Program, Harvard-MIT, Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

(Received 5 February 2009; revised 22 July 2009; accepted 22 July 2009)

The perceptual integration of 250 Hz, 500 ms vibrotactile and auditory tones was studied in detection experiments as a function of (1) relative phase and (2) temporal asynchrony of the tone pulses. Vibrotactile stimuli were delivered through a single-channel vibrator to the left middle fingertip and auditory stimuli were presented diotically through headphones in a background of 50 dB sound pressure level broadband noise. The vibrotactile and auditory stimulus levels used each yielded 63%–77%-correct unimodal detection performance in a 2-I, 2-AFC task. Results for combined vibrotactile and auditory detection indicated that (1) performance improved for synchronous presentation, (2) performance was not affected by the relative phase of the auditory and tactile sinusoidal stimuli, and (3) performance for non-overlapping stimuli improved only if the tactile stimulus preceded the auditory. The results are generally more consistent with a "Pythagorean Sum" model than with either an "Algebraic Sum" or an "Optimal Single-Channel" Model of perceptual integration. Thus, certain combinations of auditory and tactile signals result in significant integrative effects. The lack of phase effect suggests an envelope rather than fine-structure operation for integration. The effects of asynchronous presentation of the auditory and tactile stimuli are consistent with time constants deduced from single-modality masking experiments.
 [DOI: 10.1121/1.3204305]

PACS number(s): 43.66.Lj, 43.66.Wv [RLF] Pages: 1960–1974

I. INTRODUCTION

Multisensory interactions commonly arise in everyday exploration of the environment, and numerous examples can be cited to demonstrate the influence of one sensory modality over another. For example, the presence of an auditory signal can alter judgments regarding the intensity, numerosity, and motion of visual signals (Stein *et al.*, 1996; Bhattacharya *et al.*, 2002; Sekuler *et al.*, 1997), and the location of a visual stimulus can modify the perceived location of an auditory signal (as in the ventriloquism effect; Woods and Recanzone, 2004). In the area of speech perception, for example, the McGurk effect (McGurk and MacDonald, 1976) provides a powerful demonstration of the ability of visual cues derived from lip-reading to influence the perception of auditory speech cues. The current research is concerned with exploring perceptual interactions between the senses of hearing and touch and is motivated by recent results from anatomical and physiological studies demonstrating significant interactions between these two senses.

In anatomical research, recent studies indicate that areas of the central nervous system that have traditionally been thought to receive auditory-only inputs may also receive inputs from the somatosensory system. For example, in the brainstem, the trigeminal nerve sends somatosensory input to the cochlear nucleus of the guinea pig (Zhou and Shore, 2004), while in the thalamus, somatosensory projections are sent to non-primary areas of the auditory cortex of the macaque monkey (Hackett *et al.*, 2007). Projections within the cortex have been found from the secondary somatosensory cortex to the primary auditory cortex of the marmoset monkey (Cappe and Barone, 2005) as well as to non-primary auditory cortical areas of the macaque monkey (Smiley *et al.*, 2007). Additionally, recent physiological studies in humans (using non-invasive imaging) as well as in non-human primates (using electrophysiology) suggest that the auditory cortex is an active multisensory area, responding to somatosensory input alone as well as to combined auditory and tactile stimuli in a manner that is different from responses to auditory-only stimulation (Schroeder *et al.*, 2001; Foxe *et al.*, 2002; Fu *et al.*, 2003; Caetano and Jousmaki, 2006; Kayser *et al.*, 2005; Schurmann *et al.*, 2006; Lakatos *et al.*, 2007).

Although there is increasing anatomical and physiological evidence that tactile and auditory stimuli interact, there is less direct perceptual evidence for this interaction. Previous perceptual studies of auditory and tactile interactions can be organized into two broad categories: the influence of tactile stimulation on auditory perception and the influence of auditory stimulation on tactile perception. In the first category, experiments have shown that tactile stimuli can influence auditory localization (Caclin *et al.*, 2002) and auditory motion (Soto-Faraco *et al.*, 2004). Other perceptual studies have examined the effects of tactile stimulation on the perceived loudness or discriminability of auditory stimuli (Schurmann *et al.*, 2004; Schnupp *et al.*, 2005; Gillmeister and Eimer, 2007; Yarrow *et al.*, 2008). These studies employed a variety of experimental procedures (i.e., loudness matching, signal detectability, and signal discriminability) and, under certain

experimental conditions, have shown increased loudness or discriminability for paired auditory-tactile stimuli compared with the single-modality stimulus.

In the second category, auditory stimuli have been effective in influencing tactile perception, including such examples as changes in tactile threshold or tactile magnitude when paired with an auditory stimulus (Gescheider *et al.*, 1969; Gescheider *et al.*, 1974; Ro *et al.*, 2009). Other studies have shown that changing the high-frequency components of the auditory stimulus on a tactile task can affect the roughness judgment of the tactile stimulus (Jousmaki and Hari, 1998; Guest *et al.*, 2002) and that judgments of tactile numerosity can be affected by the presence of competing auditory signals (Bresciani *et al.*, 2005). In several of these studies (Soto-Faraco *et al.*, 2004; Bresciani *et al.*, 2005; Gillmeister and Eimer, 2007), temporal synchrony between the auditory and tactile stimuli was an important factor in eliciting interactive effects.

Further systematic and objective studies exploring the perceptual characteristics of the auditory and tactile systems are necessary for understanding the interactions between these sensory systems. In addition, perceptual studies will aid in interpreting the anatomical and neurophysiological studies which demonstrate significant interactions between the auditory and tactile sensory systems. The goal of the current research was to obtain objective measurements of auditory-tactile integration for near-threshold signals through psychophysical experiments conducted within the framework of signal-detection theory using d' (and percent correct) as a measure of detectability. Our hypothesis [derived from a general model proposed by Green (1958)] states that if the auditory and tactile systems do integrate into a common neural pathway, then the detectability of the two sensory stimuli presented simultaneously will be significantly greater than the detectability of the individual sensory stimuli. Specifically, if the stimuli are judged independently of one another, the resulting d' should equal the root-squared sum of the individual sensory d' values. If, on the other hand, the stimuli are integrated into a single percept before being processed, the resulting d' should equal the sum of the individual d' values.

The experiments reported here explore the perceptual integration between auditory pure tones and vibrotactile sinusoidal stimuli as a function of (1) phase and (2) stimulus-onset asynchrony (SOA). Manipulations of the relative phase of the tactile and auditory tonal stimuli were conducted as a means of exploring whether the interaction of the stimuli occurs at the level of the fine structure or envelope of the signals from the two separate sensory modalities. Manipulations of SOA between the tactile and auditory signals were conducted to explore the time course over which cross-modal interactions may occur. Measurements of d' (and percent correct) were obtained for auditory-alone, tactile-alone, and combined auditory-tactile presentations. The observed performance in the combined condition was then compared to predictions of multi-modal performance derived from observed measures of detectability within each of the two separate sensory modalities.

II. METHODS

A. Stimuli

The auditory stimulus employed in all experimental conditions was a 250-Hz pure tone presented in a background of pulsed 50 dB SPL (sound pressure level) Gaussian broadband noise (bandwidth of 0.1–11.0 kHz). The tactile stimulus employed in all experimental conditions was a sinusoidal vibration with a frequency of 250 Hz. The background noise was utilized to mask possible auditory cues arising from the tactile device and was present in all auditory (A), tactile (T), and combined auditory plus tactile (A+T) test conditions. The 250-Hz signals in both modalities were generated digitally (using MATLAB 7.1 software) to have a total duration of 500 ms that included 20-ms raised cosine-squared rise/fall times.

The digitized signals were played through a digital-analog sound card (Lynx Studio Lynx One) with a sampling frequency of 24 kHz and 24-bit resolution. The auditory signal was sent through channel 1 of the sound card to an attenuator (TDT PA4) and headphone buffer (TDT HB6) before being presented diotically through headphones (Sennheiser HD 580). The tactile signal was passed through channel 2 of the sound card to an attenuator (TDT PA4) and amplifier (Crown D-75) before being delivered to an electromagnetic vibrator (Alpha-M Corporation model A V-6). The subject's left middle fingertip made contact with the vibrator (0.9 cm diameter). A laser accelerometer was used to calibrate the tactile device.

B. Subjects

Eleven subjects ranging in age from 18 to 48 years (five females) participated in this study. Audiological testing was conducted on the first visit to the laboratory. Only those subjects who met the criterion of normal audiometric thresholds (20 dB hearing level or better at frequencies of 125, 250, 500, 1000, 2000, 4000, and 8000 Hz) were included in the studies. All subjects were paid an hourly wage for their participation in the experiments and signed an informed-consent document prior to entry into the study. Six subjects participated in experiment 1 (S_1, S_2, S_3, S_4, S_6, and S_7), four in experiment 2A (S_6, S_8, S_9, and S_{10}), four in experiment 2B (S_2, S_4, S_7, and S_8), and four in experiment 2C (S_8, S_{10}, S_{21}, and S_{24}). Six of the subjects participated in multiple experiments (S_2, S_4, S_6, S_7, S_8, and S_{10}).

An additional 11 subjects passed the audiometric criteria and began participation in the study but were terminated from the experiments on the basis of instability in their threshold measurements over the course of the 2-h test sessions. Further details of the criteria that were used for disqualifying a subject from continued participation in the study are provided below in Sec. II D.[1]

C. Experimental conditions

The experiments examined the perceptual integration of 250-Hz sinusoidal auditory and vibrotactile signals that were each presented near the threshold of detection. Threshold measurements were first obtained under each of the two

TABLE I. Description of experimental conditions studied in Experiments 1, 2A, 2B, and 2C. In all experiments, the frequency of the auditory and tactile stimulus was always 250 Hz with duration of 500 ms. SOA was defined as SOA=Onset $\text{Time}_{\text{tactile}}$−Onset $\text{Time}_{\text{auditory}}$.

Condition	Starting phase (deg)		SOA (ms)	Subjects	No. of repetitions
	Auditory stimulus	Tactile stimulus			
Experiment 1: Variable studied: phase					
1–1	0	0	0	S_1,S_2,S_3,S_4,S_6,S_7	6
1–2	0	90	0	S_1,S_2,S_3,S_4,S_6,S_7	6
1–3	0	180	0	S_1,S_2,S_3,S_4,S_6,S_7	6
1–4	0	270	0	S_1,S_2,S_3,S_4,S_6,S_7	6
Experiment 2: Variable studied: SOA					
Experiment 2A: Auditory stimulus precedes tactile stimulus					
2A-1	0	0	0	S_6,S_8,S_9,S_{10}	≥4
2A-2	0	0	500	S_6,S_8,S_9,S_{10}	4
2A-3	0	0	550	S_6,S_8,S_9,S_{10}	4
2A-4	0	0	600	S_6,S_8,S_9,S_{10}	4
2A-5	0	0	650	S_6,S_8,S_9,S_{10}	4
2A-6	0	0	700	S_6,S_8,S_9,S_{10}	4
2A-7	0	0	750	S_6,S_8,S_9,S_{10}	4
Experiment 2B: Tactile stimulus precedes auditory stimulus, no temporal overlap					
2B-1	0	0	0	S_2,S_4,S_7,S_8	≥4
2B-2	0	0	−500	S_2,S_4,S_7,S_8	4
2B-3	0	0	−550	S_2,S_4,S_7,S_8	4
2B-4	0	0	−600	S_2,S_4,S_7,S_8	4
2B-5	0	0	−650	S_2,S_4,S_7,S_8	4
2B-6	0	0	−700	S_2,S_4,S_7,S_8	4
2B-7	0	0	−750	S_2,S_4,S_7,S_8	4
Experiment 2C: Tactile stimulus precedes auditory stimulus, with temporal overlap condition					
2C-1	0	0	0	S_8,S_{10},S_{21},S_{24}	≥4
2C-2	0	0	−250	S_8,S_{10},S_{21},S_{24}	4
2C-3	0	0	−500	S_8,S_{10},S_{21},S_{24}	4
2C-4	0	0	−550	S_8,S_{10},S_{21},S_{24}	4
2C-5	0	0	−600	S_8,S_{10},S_{21},S_{24}	4
2C-6	0	0	−650	S_8,S_{10},S_{21},S_{24}	4
2C-7	0	0	−700	S_8,S_{10},S_{21},S_{24}	4
2C-8	0	0	−750	S_8,S_{10},S_{21},S_{24}	4

single-modality conditions (A and T separately). Then the detectability of the combined auditory plus tactile (A+T) signal was measured at levels established for threshold within each of the two individual modalities. The experimental conditions examined the effects of relative phase (Experiment 1) and SOA (Experiments 2A, 2B, and 2C) of the tactile signal relative to the auditory signal.

A summary of the conditions employed in the two experiments is provided in Table I. Throughout the experiments, the stimuli were 250-Hz sinusoids of 500-ms duration (including 20-ms rise/fall times). The stimulus parameters are described in terms of the starting phase of the auditory (column 2) and tactile (column 3) stimuli and SOA (column 4). Specifically, we define SOA to be: Onset $\text{Time}_{\text{tactile}}$ −Onset $\text{Time}_{\text{auditory}}$. Thus, the SOA is positive when the auditory stimulus precedes the tactile, 0 when the two stimuli have simultaneous onsets, and negative when the tactile stimulus precedes the auditory. Information concerning the subjects and the number of repetitions of each experimental condition is provided in the final two columns of Table I.

Baseline condition. A baseline condition employing 0-ms SOA and starting phase of 0° for both auditory and tactile stimuli was included in each of the experiments (conditions 1-1, 2A-1, 2B-1, and 2C-1 in Table I). Performance on this baseline condition was generally measured as the first A+T condition in each test session for each subject under each of the four experiments.

Experiment 1. Experiment 1 examined the effect of the starting phase of the tactile relative to the auditory stimuli and is described in Table I (conditions 1-1 through 1-4). The auditory starting phase was always 0°, while the tactile starting phase took on four different values: 0°, 90°, 180°, and 270°. In each of these four conditions, the auditory and tactile stimuli were temporally synchronous (0-ms SOA) and thus had identical onset and offset times. This experiment was conducted on six subjects; each completed six repeti-

tions of each condition in six or seven test sessions. The order of the four experimental conditions was randomized within each replication for each subject.

Experiment 2. Experiment 2 examined the effect of asynchronous presentation of the auditory and tactile stimuli and is described in Table I (conditions 2A, 2B, and 2C). The starting phase of the auditory and tactile sinusoids was 0° throughout all of the conditions of Experiment 2. The presentation order of the experimental conditions in Experiments 2A, 2B, and 2C was randomized across sessions for each of the subjects.

In Experiment 2A (Table I, conditions 2A-1 through 2A-7), the auditory stimulus preceded the tactile stimulus with six values of SOA in the range of 500–750 ms (i.e., there was never any temporal overlap between the two stimuli). Thresholds in the baseline condition (0-ms SOA) were also measured for a total of seven conditions. Four subjects completed four replications of each of the non-zero SOA conditions, while the 0-ms SOA condition was measured at the start of each session (resulting in more than four measurements of this condition for some subjects). The number of test sessions required to complete the experiment ranged from 4 to 9 across subjects.

In Experiments 2B and 2C (Table I, conditions 2B-1 through 2B-7 and 2C-1 through 2C-8), the tactile stimulus preceded the auditory stimulus. In Experiment 2B, six values of SOA were studied in the range of −500 to −750 ms (there was no temporal overlap between the two stimuli), in addition to the baseline (0-ms SOA) condition. Four subjects[2,3] completed four replications of each of the six non-zero SOA conditions, requiring four to nine test sessions. In Experiment 2C, in addition to the conditions described above for Experiment 2B, an SOA of −250 ms was included in order to examine the effect of partial temporal overlap between the two stimuli. Four subjects each completed four replications of the seven non-zero SOA conditions. In Experiment 2C, one subject from Experiment 2B (S_8) returned to complete four repetitions of condition 2C-2 (−250–ms SOA) and a partial subset of the remaining SOA values. Three additional subjects (S_{10}, S_{21}, and S_{24}) completed four replications of the eight experimental conditions in five to nine sessions.

D. Experimental procedures

For all experimental conditions, subjects were seated in a sound-treated booth and were presented 50–dB SPL broadband noise diotically via headphones. For testing in conditions that involved presentation of the tactile stimulus (T and A+T), the subject placed the left middle finger on a vibrator which was housed inside a wooden box for visual shielding and sound attenuation. A heating pad was placed inside the box in order to keep the box and tactile device at a constant temperature.

The following protocol was employed for testing within each experimental session: (i) Thresholds for each single-modality condition (A and T) were estimated adaptively (Levitt, 1971). (ii) Fixed-level testing was conducted for A and T separately to establish a signal level for single-modality performance in the range of 63%–77%-correct. (iii) Fixed-level performance was measured for the baseline A+T condition (0-ms SOA, 0° phase). (iv) Fixed-level performance was measured in the experimental A+T conditions. (v) Single-modality fixed-level testing was repeated as in (ii) except with an expanded acceptable performance range of 56%–84%-correct. (Data from the second set of single-modality conditions were not otherwise used.) The number of experimental A+T conditions that could be completed within a given test session was dependent on the time required to establish signal levels that met the single-modality performance criterion.

A test session typically lasted 2 h, during which performance was measured in fixed-level experiments for the A and T conditions and A+T conditions associated with a given experiment. For each subject, three training sessions identical to the experimental sessions were provided before data were recorded. If a subject participated in multiple experiments, the three training sessions were provided only prior to the first experiment (i.e., subjects S_2, S_4, S_6, S_7, S_8, and S_{10} underwent only three training sessions even though they participated in multiple experiments). Attention to the combined A+T stimulus was ensured by having subjects count the number of times they perceived a signal. Each experimental session lasted no more than 2 h on any given day, and subjects took frequent breaks throughout the session.

If the single-modality threshold re-tests at the end of a given session were less than 56%-correct or greater than 84%-correct (±2 standard deviations assuming an original score of 70%-correct), the data for that session were discarded. The number of sessions discarded per subject ranged from 0 to 3 in Experiments 1 and 2A, 0–2 in Experiment 2B, and 0 in Experiment 2C. Subjects were terminated from the experiment if their scores shifted by more than 2 standard deviations in three non-training sessions, resulting in the disqualification of 11 subjects from participation in the study. Typically, disqualification resulted from increased variability in tactile threshold measurements. On average, the difference between scores measured at the beginning and end of a test session was 10.8 percentage points in the disqualified subjects compared to −0.6 percentage points in the retained subjects (with the differences in absolute values being 16.2 and 8.7 percentage points, respectively). Differences between disqualified and retained subjects were not as great for auditory scores: The corresponding differences were 4.8 and 3.0 percentage points (with the differences in absolute values being 7.5 and 6.6 percentage points, respectively).

2-I, 2-AFC fixed-level tests. The adaptive threshold estimates under the single-modality A and T conditions were employed in a 2-I, 2-AFC fixed-level procedure with 75 trials per run. Stimulus levels were adjusted and runs were repeated until scores of 63%–77%-correct were obtained. These stimulus levels were then used in testing the combined A+T conditions with the fixed-level 2-I, 2-AFC procedure.

On each presentation, the tone (auditory, tactile, or auditory-tactile) was presented with equal *a priori* probability in one of the two intervals. The interval duration was 1.15 s for Experiment 1 and 1.25 s for Experiment 2. Each inter-

val was cued by visually highlighting a push-button on the computer screen located in front of the subject. Noise was presented diotically over headphones starting 500 ms before the first interval and played continuously throughout a trial (including the durations of the two intervals and the 500-ms duration between intervals) before being turned off 500 ms after the end of the second interval. Each trial had a fixed duration of 3.8 s (Experiment 1) or 4 s (Experiment 2), plus the time it took subjects to respond. The onset of the stimulus (A, T or combined A+T) was always coincident with the onset of the observation interval in which it appeared. Subjects responded between trials by selecting the interval in which they thought the stimulus was presented (using either a mouse or keyboard) and were provided with visual correct-answer feedback.

E. Data analysis

A two-by-two stimulus-response confusion matrix was constructed for each 75-trial experimental run and was used to determine percent-correct scores and signal-detection measures of sensitivity (d'). These measures were averaged across the repetitions of each experimental condition within a given subject. Statistical tests performed on the data included analyses of variance (ANOVAs) on the arcsine transformed percent-correct scores, with statistical significance level defined for probability (p-values) less than or equal to 0.01. For statistically significant effects, a *post hoc* Tukey–Kramer analysis was performed with alpha=0.05.

F. Models of integration

The results of the experiments were compared with three different models of integration: the Optimal single-channel model (OSCM), the Pythagorean sum model (PSM), and the algebraic sum model (ASM). The OSCM assumes that the observers' responses are based on the better of the tactile or auditory input channels. The predicted D'_{OSCM} [4] for the combined A+T condition is the greater of the tactile (d'_{T}) or auditory (d'_{A}), $D'_{\mathrm{OSCM}}=\max(d'_{\mathrm{T}},d'_{\mathrm{A}})$. The PSM assumes that integration occurs across channels (e.g., as in audio-visual integration, Braida, 1991) and that the d' in the combined auditory-tactile condition is the Pythagorean sum of the d-primes for the separate channels, $D'_{\mathrm{PSM}}=\sqrt{d'^2_{\mathrm{A}}+d'^2_{\mathrm{T}}}$. The ASM, on the other hand, assumes that integration occurs within a given channel and that the combined d' is the linear sum of the d-primes for the separate channels, $D'_{\mathrm{ASM}}=d'_{\mathrm{A}}+d'_{\mathrm{T}}$. For example, if the auditory d'_{A} was 1.0 (69%-correct) and the tactile d'_{T} was 0.8 (66%-correct), the OSCM would predict a D'_{OSCM} of 1.0 (69%-correct), the PSM would predict a D'_{PSM} of 1.28 (74%-correct), and the ASM would predict a D'_{ASM} of 1.8 (82%-correct). The OSCM prediction is never greater than the PSM prediction, which in turn is never greater than the prediction of the ASM.

Chi-squared goodness-of-fit calculations were employed to compare observed with predicted values from each of the three models. The predictions of the models were evaluated as follows: First, d-prime values were determined for each auditory (d'_{A}) and tactile (d'_{T}) experiment on the basis of 75 total trials. Second, predicted d-prime values were computed for the three models according to the formulas given above. Third, predicted percent correct scores were computed for each of the models in the following manner: Percent Correct$=100\phi(D'_{\mathrm{A+T}}/2)$, where ϕ is the cumulative of the Gaussian distribution function and $D'_{\mathrm{A+T}}$ is the predicted D'. Fourth, the observed A+T confusion matrix was analyzed to estimate $d'_{\mathrm{A+T}}$ and the "no bias" estimate of percent correct score was computed as $\%_{\mathrm{A+T}}=100\phi(d'_{\mathrm{A+T}}/2)$. This relatively small adjustment (1.6 percentage points on average, 13 points maximum) was necessary because the predictions of the models assumed that the observer is not biased. Predictions (D'_{OSCM}, D'_{PSM}, and D'_{ASM} or $\%_{\mathrm{OSCM}}$, $\%_{\mathrm{PSM}}$, and $\%_{\mathrm{ASM}}$) were compared with observations ($d'_{\mathrm{A+T}}$ or $\%_{\mathrm{A+T}}$). The proportion of the observations that agreed with predictions was judged by having a chi-squared value less than 3.841 (the 95% criterion) between predicted and observed scores (corrected as discussed above) using a contingency table analysis (Neville and Kennedy, 1964). This analysis allows for errors in both the observed score and the predicted score.

III. RESULTS

A. Signal levels employed in single-modality conditions

Single-modality auditory and tactile thresholds were obtained both at the beginning and at the ending of each individual test session. The data reported here, however, are based solely on the initial measurements. Analyses that used the average of the beginning and ending single-modality measurements were not significantly different from these. Thus, we used the post-experiment measurements merely as a tool for determining threshold stability.

Levels for auditory-alone conditions. The mean signal levels in dB SPL established for performance in the range of 63%–77%-correct for a 250-Hz tone in 50-dB SPL broadband noise are shown in the upper panel of Fig. 1. Mean levels of the tone are plotted for each individual subject in each of the four experiments. Each data point depicted in the plot is based on an average of at least 4 and as many as 11 measurements in the fixed-level 2-I, 2-AFC procedure (each of which yielded performance in the range of 63%–77%-correct). Ten of the 11 subjects had average auditory masked thresholds within a 2.1-dB range of 22.3–24.4 dB SPL. The remaining subject (S_{10}) had a value of 27.8 dB SPL, measured consistently across multiple sessions. Within a given subject, tonal levels were highly stable for measurements made within a given experiment and across experiments. Values of ± 2 standard error of the mean (SEM) (accounting for 96% of the measurements) ranged from 0.095 to 1.1 dB across subjects and experiments.

These results are consistent with those obtained in previous studies of tonal detection in broadband noise. Critical ratios were calculated for the tone-in-noise levels shown in Fig. 1 by subtracting the spectrum level of the noise at 250 Hz (which was 7.4 dB/Hz) from the presentation levels of the 250-Hz tone. Across subjects and experiments, mean critical ratios ranged from 14.9 to 20.4 dB and are consistent with the critical ratio value of 16.5 dB at 250 Hz reported by

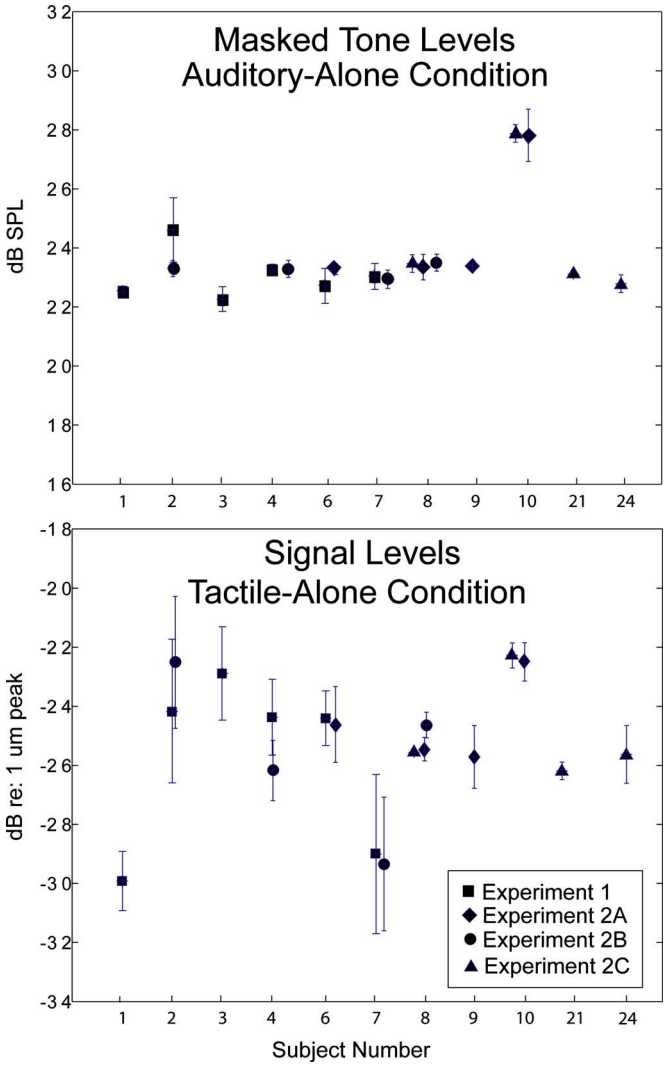


FIG. 1. (Color online) Single-modality signal levels employed for individual subjects tested in each of the four experiments. Auditory levels are for detection of a 250-Hz pure tone in 50-dB SPL broadband noise. Tactile levels are for detection of a 250-Hz sinusoidal vibration presented to the fingertip. Different symbols represent results obtained in different experiments. Some subjects participated in more than one experiment. Error bars are 2 SEM.

Hawkins and Stevens (1950). Thus, these results indicate that subjects were listening to the auditory tones in noise at levels that were close to masked threshold.

Levels for tactile-alone conditions. The mean signal levels established for performance in the range of 63%–77%-correct for a 250-Hz sinusoidal vibration to the left middle fingertip are shown in the lower panel of Fig. 1. All threshold measurements were obtained in the presence of a diotic 50–dB SPL broadband noise presented over headphones. Signal levels are plotted in decibel *re* 1 μm peak displacement for individual subjects who participated in each of the four experiments. Each mean level is based on 4–11 measurements across individual subjects and experiments. Average signal levels employed in the tactile-alone conditions ranged from −30 to −22 dB *re* 1μm peak. Within-subject values of ±2 SEM (accounting for 96% of the measurements) ranged from 0 to 2.4 dB across subjects and experiments. The Appendix discusses the unlikely possibility that the tactile stimulus was detected auditorally via bone conduction.

The signal levels employed for the tactile-alone conditions are generally consistent with previous results in the literature for vibrotactile thresholds at 250 Hz obtained using vibrators with contactor areas similar to that of the device employed in the present study (roughly 80 mm^2). Investigators using contactor areas in the range of 28–150 mm^2 have reported mean thresholds in the range of −21 to −32 dB *re* 1μm peak (Verrillo *et al.*, 1983; Lamore *et al.*, 1986; Rabinowitz *et al.*, 1987).

B. Baseline experiment

Results from the baseline experiment are shown for individual subjects in Experiments 1, 2A, 2B, and 2C in the four panels of Fig. 2. The mean percent correct scores with error bars depicting ±2 SEM are plotted for the three conditions of A-alone, T-alone, and A+T (SOA=0 ms, phase =0°; see Table I, experimental conditions 1-1, 2A-1, 2B-1, and 2C-1) for each subject within each experiment. Averages across subjects are provided as the rightmost data bars within each panel. Across the four experiments, there is a substantial increase in the percent correct score when the auditory and tactile stimuli are presented simultaneously compared with the A- and T-alone conditions. Averaged over subjects, the results indicate that scores for the A-alone and T-alone condition were similar (ranging from 67.8% to 74.9%-correct across experiments) and lower than the average scores in the A+T conditions (which ranged from 75.2% to 88.8%-correct across the four experiments). Variability was generally low, with values of ±2 SEM ranging from 0.6 to 15.1 percentage points across subjects and experiments with all but one subject less than 7 percentage points.

A two-way ANOVA was performed on the results of the baseline experiment to examine the main effects of Condition (A, T, A+T) and Subject (11 different subjects across experiments). These results indicate a significant main effect for Condition [$F(2,257)=91.44$, $p<0.01$] but not Subject [$F(10,257)=1.00$, $p=0.035$], and a significant effect for their interaction [$F(20,257)=2.8$, $p<0.01$]. A *post hoc* analysis of the main effect of Condition showed that scores on the A+T condition were significantly greater than on the A-alone and T-alone conditions and that the A-alone and T-alone conditions were not significantly different from one another. A *post hoc* analysis of the Condition by Subject effect indicated that all subjects were similar on the A-alone and T-alone conditions but different on the A+T condition. Specifically, of the 11 subjects tested, 8 had a significantly higher A+T score compared with the A-alone and T-alone scores; two subjects showed no significant increase in score (S_8 and S_{24}); and one subject (S_{21}) had significantly greater A+T scores compared to either A-alone or T-alone, but not to both.

C. Experiment 1: Effects of relative auditory-tactile phase

The results of Experiment 1 are shown in Fig. 3. Percent-correct scores averaged across six subjects and six repetitions per condition are shown for each of the six ex-

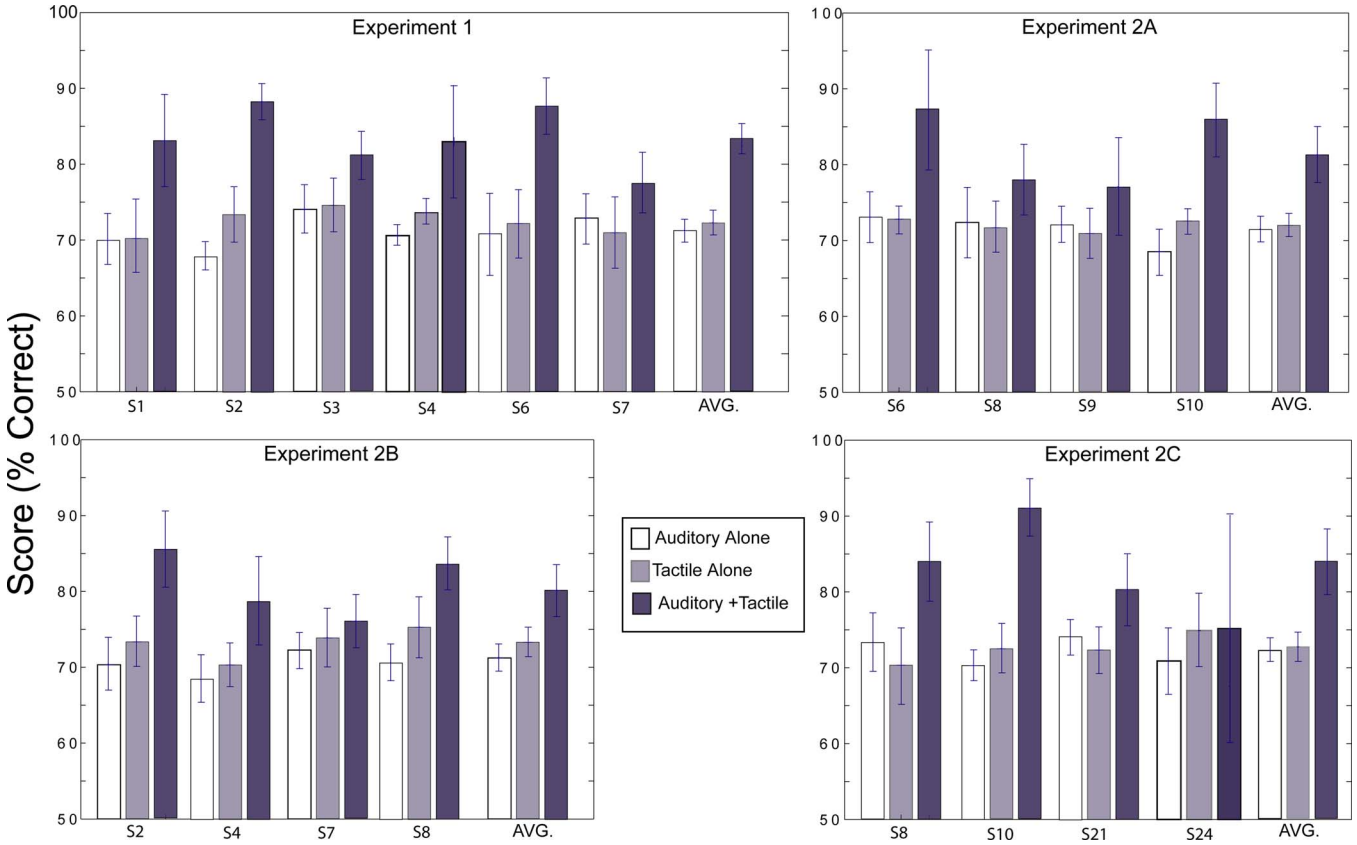


FIG. 2. (Color online) Summary of results for baseline condition in Experiments 1, 2A, 2B, and 2C. Percent correct scores for the individual subjects in each experiment are averaged across multiple repetitions per condition; number of repetitions varies by subject and is equal to or greater than 4 per subject. AVG is an average across subjects and repetitions in each experiment. White bars represent A-alone conditions, gray bars represent T-alone conditions, and black bars represent the A+T baseline condition with SOA=0 ms and phase=0°. Error bars are 2 SEM.

perimental conditions: A-alone, T-alone, and combined A+T with four different values of the starting phase of the tactile stimulus relative to that of the auditory stimulus (0°, 90°, 180°, and 270°). Average scores were 71.2%-correct for A-alone, 72.2%-correct for T-alone, and ranged from 83.2%- to 84.6%-correct across the four combined A+T conditions.

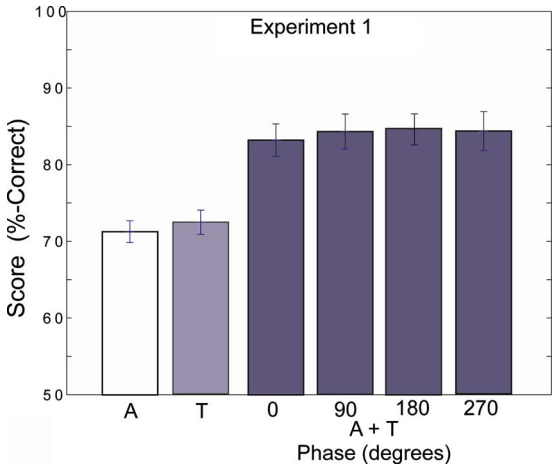


FIG. 3. (Color online) Summary of results for experiment 1. Percent correct scores are averaged across six subjects with six sessions per condition. Scores are shown for A-alone (white bar), T-alone (light gray bar), and combined A+T condition (dark gray bars) as a function of starting phase (in degrees) of the tactile stimulus relative to the auditory stimulus. Error bars are 2 SEM.

The Appendix discusses the unlikely possibility that this variation was caused by a bone-conducted interaction between the tactile and auditory stimuli. Variability in terms of ±2 SEM ranged from 2.0 to 2.5 percentage points across the four phase conditions.

A two-way ANOVA was performed with main factors of Condition (A, T, A+T: tactile phase) and Subject. The results of the ANOVA indicate a significant main effect for factors of Condition $[F(5,192)=44.93, p<0.01]$ and Subject $[F(5,192)=4.01, p<0.01]$ but not for their interaction $[F(25,192)=1.61, p=0.04]$. The *post hoc* analysis on Condition indicated that scores on the A-alone and T-alone conditions were not significantly different from one another, that scores for the four A+T combined conditions were not significantly different from one another, and that the scores for each of the four A+T conditions were significantly greater than the A- and T-alone scores. The *post hoc* analysis on subject indicated that the A+T scores for S_6 were significantly greater than those of S_1 and S_7, and that the scores for S_2 were significantly greater than those of S_7.

D. Experiment 2: Effects of SOA

Experiment 2 explored the effect of SOA between the auditory and tactile stimuli in three different experiments. Experiment 2A tested conditions in which the auditory stimulus preceded the tactile stimulus, and Experiments 2B and 2C tested conditions in which the tactile stimulus pre-

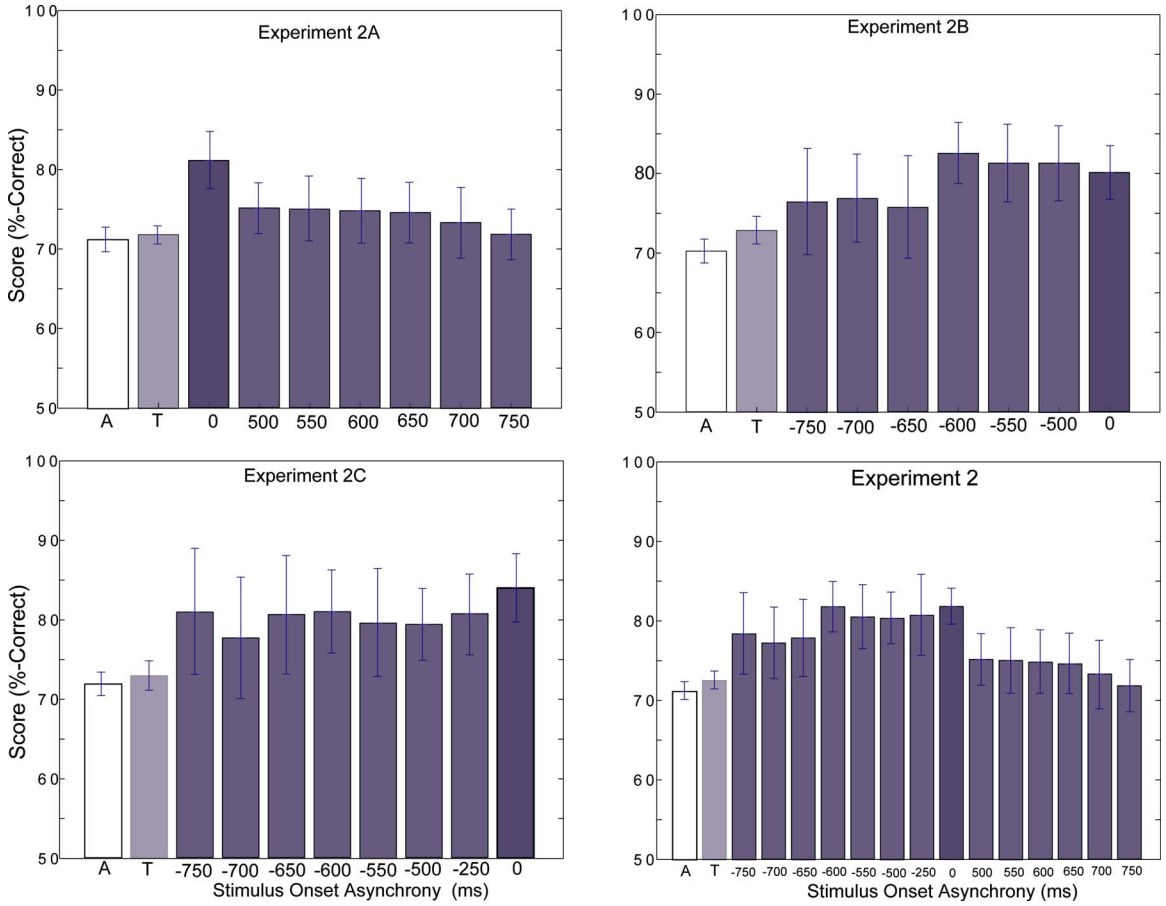


FIG. 4. (Color online) Summary of results for Experiment 2. In all panels, scores are shown for A-alone (white bars), T-alone (light gray bars), and combined A+T condition (dark gray bars). In the upper left panel (Experiment 2A: auditory precedes tactile), percent correct scores are averaged across four subjects with four sessions per condition (SOA=0 ms has more than four repetitions). In the upper right panel (Experiment 2B: tactile precedes auditory), percent correct scores are averaged across four subjects with four sessions per condition (SOA=0 ms has more than four repetitions). In the lower left panel (Experiment 2C: tactile precedes auditory, with temporal overlap), percent correct scores are averaged across four subjects with four sessions per condition (SOA=0 ms has more than four repetitions). The lower right panel (Experiment 2) provides a composite summary of percent correct scores averaged across all subjects and repetitions from Experiments 2A, 2B, and 2C. In all panels, error bars are 2 SEM.

ceded the auditory stimulus. Percent-correct scores averaged across four subjects and four repetitions of each non-zero SOA condition in each of these experiments are shown in Fig. 4. Error bars represent ±2 SEM.

In Experiment 2A (Fig. 4, upper left panel), scores for the A-alone and T-alone conditions averaged 71.1%- and 71.8%-correct, respectively. For the combined A+T conditions, average scores of the non-zero SOA conditions ranged from 71.8%-correct (SOA=500 ms) to 75.1%-correct (SOA=750 ms). Variability, in terms of ±2 SEM, ranged from 3.2 percentage points (SOA=500 ms) to 4.3 percentage points (SOA=700 ms). A two-way ANOVA was conducted using main factors of Condition (A, T, and the seven combined A+T conditions with different values of SOA) and Subject. The results of the ANOVA indicate that both main factors (Condition: $[F(8,156)=6.16, p<0.01]$; subject: $[F(3,156)=19.32, p<0.01]$), as well as their interaction $[F(24,156)=2.3, p<0.01]$, were significant. The *post hoc* analysis revealed that only one A+T combined condition, that of SOA=0 ms (i.e., the baseline condition), produced a score that was significantly greater than the A-alone or T-alone score. The scores for the remaining SOA conditions were not significantly greater than the scores in the A-alone or T-alone conditions. The *post hoc* analysis of the subject effect indicated that the scores for S_{10} were significantly different from those of the other three subjects. For the interaction effect, S_{10} showed significantly greater A+T scores at all SOA's except 750 ms compared with A- and T-alone, while none of the other subjects showed a significant difference between non-zero SOA and A-alone and T-alone scores.

In Experiment 2B (Fig. 4, upper right panel), scores for the A-alone and T-alone conditions averaged 70.5%- and 73.3%-correct, respectively. For the combined A+T conditions, averaged scores of the non-zero SOA conditions ranged from 75.7%-correct (SOA=−650 ms) to 82.5%-correct (SOA=−600 ms). Variability in terms of ±2 SEM ranged from 3.8 percentage points (SOA=−600 ms) to 6.7 points (SOA=−750 ms). The results of a two-way ANOVA indicated that the two main effects of condition and subject were both significant (Condition: $[F(8,139)=6.6, p<0.01]$; Subject: $[F(3,139)=14.76, p<0.01]$), but not their interaction $[F(24,139)=1.77, p=0.02]$. A *post hoc* analysis indicated that scores on the combined A+T conditions with SOA values of 0, −500, −550, and −600 ms were significantly greater than scores on the A-alone and T-alone conditions. Scores on the combined A+T conditions with SOA

TABLE II. Chi-squared tests: predicted vs observed. This table enumerates the number of observations that have passed/failed the chi-squared goodness-of-fit test for each of the three models (i.e., optimal single channel, Pythagorean sum, and algebraic sum).

			Optimal single channel			Pythagorean sum			Algebraic sum		
Experiment	Condition	Total	Pass	Fail	Under-predict, fail	Pass	Fail	Under-predict, fail	Pass	Fail	Under-predict, fail
					Baseline						
1	Phase=0°	40	26	14	14	33	7	7	27	13	3
2A	SOA=0 ms	26	17	9	7	20	6	3	14	12	2
2B & 2C	SOA=0 ms	37	20	17	15	29	8	6	24	13	3
Totals		103	63 (61%)	40	36 (35%)	82 (80%)	21	16 (16%)	65 (63%)	38	8 (8%)
					Phase						
1	0°	40	26	14	14	33	7	7	27	13	3
1	90°	36	24	12	12	29	7	7	31	5	2
1	180°	37	25	12	12	33	4	4	29	8	1
1	270°	35	21	14	14	30	5	5	32	3	2
Totals		148	96 (65%)	52	52 (35%)	125 (84%)	24	23 (16%)	119 (80%)	29	7 (5%)
					SOA						
2A	500 ms	19	18	1	1	18	1	1	10	9	1
2A	550 ms	18	15	3	3	15	3	0	9	9	0
2A	600 ms	19	15	4	1	15	4	0	10	9	0
2A	650 ms	18	17	1	1	15	3	1	7	11	1
2A	700 ms	18	14	4	2	11	7	0	8	10	0
2A	750 ms	18	16	2	1	15	3	0	6	12	0
Totals		110	95 (86%)	15	9 (8%)	89 (81%)	21	2 (2%)	50 (45%)	60	2 (2%)
					SOA						
2B & 2C	−250 ms	17	13	4	4	11	6	3	11	6	2
2B & 2C	−500 ms	33	20	13	12	21	12	9	22	11	2
2B & 2C	−550 ms	31	21	10	9	19	12	7	19	12	3
2B & 2C	−600 ms	36	23	13	12	25	11	7	23	13	4
2B & 2C	−650 ms	28	16	12	8	17	11	6	18	10	1
2B & 2C	−700 ms	29	14	15	11	18	11	4	15	14	1
2B & 2C	−750 ms	28	15	13	9	16	12	6	16	12	2
Totals		202	122 (60%)	80	65 (32%)	127 (63%)	75	42 (21%)	124 (61%)	78	15 (7%)

values of −650, −700, and −750 ms, on the other hand, were not significantly different from A-alone and T-alone scores. A *post hoc* analysis of the subject effect indicated that three of the four subjects demonstrated the main trends for condition described above.

The results of Experiment 2C (Fig. 4, lower left panel) were similar to those found in Experiment 2B. Average scores for the A-alone and T-alone conditions were 71.9%- and 72.7%-correct, respectively. Average scores on the combined A+T conditions ranged from 77%-correct (SOA=−700 ms) to 81%-correct (SOA=−600 and −750 ms). Variability in terms of ±2 SEM ranged from 4.5 percentage points (SOA=−500 ms) to 8 points (SOA=−750 ms). A two-way ANOVA with main factors of condition and subject indicated significant effects for both (Condition: $[F(9,102)=10.6,\ p<0.01]$; Subject: $[F(2,102)=91.57,\ p<0.01]$), as well as for their interaction $[F(18,102)=4.69,\ p<0.01]$. A *post hoc* analysis of the condition effect indicated that the scores in the combined A+T conditions for every value of SOA were significantly higher than scores on the A-alone and T-alone conditions. A *post hoc* analysis of the subject effect showed that scores from all subjects tested were significantly different from one another. The response pattern for S_{10} as a function of condition differed from that of the other three subjects.

E. Comparisons to model predictions

Chi-squared goodness-of-fit tests were performed in order to examine which model, the OSCM, the PSM or the ASM, best fits the measured percent correct scores (Sec. II F). The proportion of observations in agreement with predictions, i.e., having a chi-squared value less than 3.841, is summarized in Table II and also shown in Fig. 5.

The baseline condition (synchronous presentation, 0° tactile-auditory phase; Fig. 5, top row) was included in all testing sessions and involved 103 comparisons. Of these, 63 (61%) of the predictions agreed with the OSCM, 82 (80%) with the PSM, and 65 (63%) with the ASM. All three models failed a simple binomial test for symmetry of error.

The results of the four phases of Experiment 1 had similar proportions in agreement with the predictions of the PSM and ASM, indicating again that relative auditory-tactile phase had no effect on integration. The middle three panels of Fig. 5 show the predicted vs observed for all four phases grouped together. Out of a total of 148 observations, 96

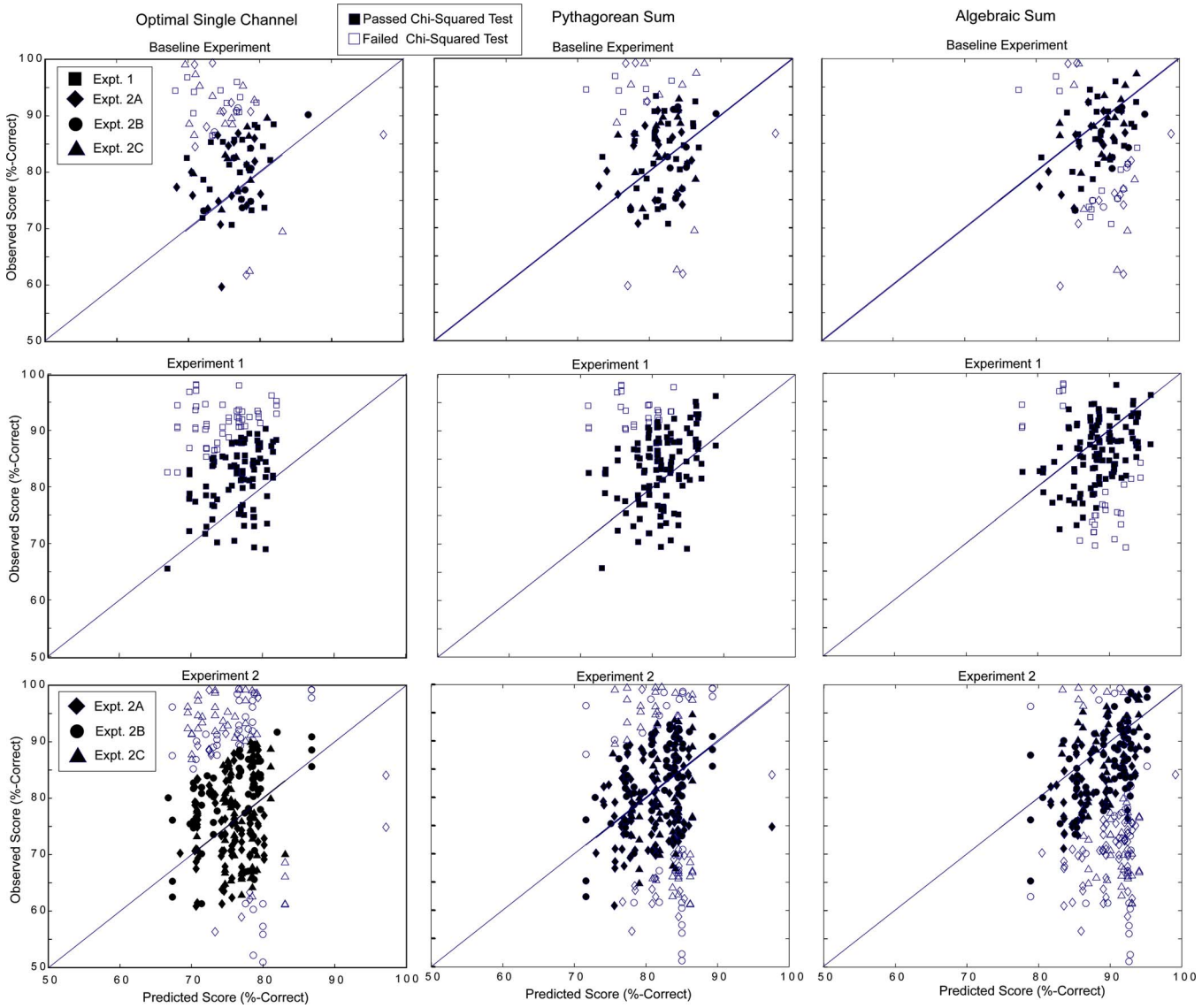


FIG. 5. (Color online) Predicted vs Observed values for the three models of integration: OSCM (far left column), PSM (middle column), and ASM (right column). The first row shows all values from the baseline experiment (SOA=0 ms, phase=0°); data from each experiment are designated by a different shape (see legend). The second row shows values from all phases in experiment 1 (relative phase). The third row shows all non-zero SOA values from experiment 2; each sub-experiment delineated by shape (see legend). Open symbols indicate that the observed value failed the chi-squared test, and filled symbols indicate that the observed value passed the chi-squared test.

(65%) agreed with the OSCM; 125 (85%) agreed with the PSM; and 119 (80%) agreed with the ASM. It can be seen that most of the data points that do not satisfy the chi-squared test are higher than the predictions of the OSCM and PSM (middle left and center panels, respectively) and lower than those of the ASM (middle right panel). The OSCM failed the symmetry test for all four phases, the PSM passed only 0 and 90°, and the ASM passed only 90°.

Discussion of the results of Experiment 2 (SOA, Fig. 5, bottom row) will be restricted to non-zero SOA because the case of zero SOA was considered above (baseline). The bottom three panels in Fig. 5 compare observed and predicted scores in Experiment 2, segregated by sub-experiment (i.e., diamond symbols represent Experiment 2A, circles are experiment 2B, and triangles are Experiment 2C). The OSCM (lower left panel) tends to under-predict the observed scores, the PSM (lower center panel) tends to over- and under-predict to a roughly equal degree, while the ASM (lower right panel) tends to over-predict scores. Table II enumerates the results of Experiment 2A separately and groups the results of Experiments 2B and C together.

For Experiment 2A, the symmetry test was performed for each model and on all of the non-zero SOA values. The OSCM passed all six non-zero SOA values; the PSM passed all non-zero SOA values except 750 ms; and the ASM failed all non-zero SOA values. The results of a chi-squared test showed that the observed and predicted scores agreed 95 out of 110 times (86%) for the OSCM, and 89 (81%) for the PSM, while only 50 (45%) agreed with the ASM. Of the cases that did not pass the chi-squared test, the OSCM produced roughly an equal number of under- (9) and over-predictions (6), while nearly all errors were over-predictions for the PSM and ASM models.

In Experiments 2B and 2C, the OSCM passed the sym-

metry test for SOA values=−750, −700, and −650 ms; the PSM passed the test for all non-zero SOA values; and the ASM passed the test for only SOA=−500 ms. The results of a chi-squared test showed that out of 202 observations (across all non-zero SOA values), 122 (60%) agreed with predictions of the OSCM, 127 (63%) with the PSM, and 124 (61%) with the ASM. However, there was a change in proportion of observations in agreement with model predictions as a function of SOA. In the case of the OSCM, for SOA values −600 or less, observations agreed with predictions 64%–76% of the time, while for SOA values greater than −600 ms this fell to less than 58% of the time. In the case of the PSM, observations agreed with predictions for all SOA values except −750 ms (all between 61% and 69%) with the lowest agreement with predictions for SOA of −750 ms (57%). In terms of the ASM, SOA values of −250, −500, −550, −600, and −650 agreed with predictions 62%–67% of the time, while SOA values of −700 and −750 ms agreed 52%–57% of the time.

These results could be due to within- or across-subject factors. Confining attention to within-subject factors, it appears that the PSM predicted the results of 4 of 11 observers in the baseline condition and two of the four observers in experiment 1. The OSCM and PSM each made correct predictions for one observer in Experiment 2A and for two observers each in Experiments 2B and 2C (−500 to −600 and −650 to −750 ms SOA). The ASM made no correct predictions for any subjects in Experiments 2A and 2B and made correct predictions for one subject in Experiment 2C (−500 to −600 ms SOA range).

Across subjects, the PSM predicted 80% of the results in the baseline condition, while in Experiment 1 the PSM and ASM predicted 85% and 80% of the results, respectively. For experiment 2A, the OSCM predicted 86% of the results, the PSM 81%, and the ASM 45%. The results for Experiments 2B and 2C did not differentiate among models, each model predicting roughly 60% of the results. When applied to results from groups of observers, none of the models considered gave an accurate statistical description of all the data (i.e., greater than 95% of measurements agreeing with the predictions of a particular model). Failures to satisfy the predictions of the models are of two types: over- and under-prediction. Over-predictions relative to the OSCM accounted for only roughly 5% of the failures for the baseline condition, Experiments 1, and 2A, and only 7% for Experiments 2B and 2C. Under-predictions relative to the ASM were 2% and 5% for experiments 1 and 2A, and 8% and 7% for the baseline condition and Experiments 2B and 2C, respectively. The cause of the over-prediction failures may be the observer's use of the sub-optimal channel or simple inattention. The cause of the under-prediction failures may be simple inattention in the single-channel presentation conditions.

IV. DISCUSSION

A. Phase and temporal asynchrony effects

Our finding of phase insensitivity leads to several important interpretations regarding the facilitative effects found in the A+T conditions. First, the lack of a phase effect on the combined-modality scores strongly suggests that the auditory background noise present in all testing was sufficient to mask any possible acoustic artifacts arising from the sinusoidal vibrations produced at the tactile device. If this had not been the case, then the relative phase of the two signals would have resulted in addition and cancellation effects, which would improve or decrease their detection. A second possibility that is ruled out by the present results is that of fine-structure operations at the neural level.[5] Instead, the similar A+T scores, independent of the relative phase of the auditory and tactile stimuli, suggest that the integration may operate on the envelopes of these stimuli rather than their fine structure. The response pattern measured in the current experiment is consistent with an envelope interaction effect: i.e., an overall increase in response but no change that is correlated with changing relative auditory-tactile phase.

The asymmetry in response patterns for the auditory-leading conditions compared to the tactile-leading conditions found in Experiment 2 (see Fig. 4, lower right panel) is consistent with differences in time constants between the auditory and tactile systems. The auditory-first condition suggests an integration window of no more than 50 ms, while the tactile-first condition suggests a window of up to 150–200 ms.

These implications of a short auditory time constant are consistent with results obtained in studies of auditory forward masking (e.g., Robinson and Pollack, 1973; Vogten, 1978; Kidd and Feth, 1981; Jesteadt *et al.*, 1982; Moore and Glasberg, 1983; Moore *et al.*, 1988; Plack and Oxenham, 1998), which indicate time constants less than 50 ms. The results reported in the current study suggest that the preceding auditory stimulus was not effective in interacting with the tactile stimulus at any SOA. In the single-modality case of auditory forward masking, however, there is significant interaction between the probe and masker at small time delays. The relatively long (500 ms) signal durations of both the auditory ("masker") and the tactile ("probe") stimuli may be partially responsible for the shorter auditory time constant observed here. Auditory studies typically employ brief (tens of milliseconds) probes and strong effects have been demonstrated for an increase in the amount of forward masking with an increase in masker duration (Fastl, 1977; Kidd and Feth, 1981).

Our finding of a relatively long time constant for tactile stimulation is consistent with results obtained in studies of tactile-on-tactile forward masking (e.g., Hamer *et al.*, 1983; Gescheider *et al.*, 1989; Gescheider *et al.*, 1994; Gescheider and Migel, 1995). Using tactile maskers with durations on the order of hundreds of milliseconds and tactile probes with durations on the order of tens of milliseconds, previous investigators have reported significant amounts of threshold shift for time delays between masker offset and probe onset on the order of 150–200 ms. Such results suggest that the tactile system maintains a persistent neural response even after cessation of the stimulus (see Craig and Evans, 1987). Our results are consistent with the sensory effect of the tactile stimulus persisting for at least 150–200 ms following its offset and that this effect is capable of interacting with the subsequent auditory stimulus to facilitate detection. For tac-

tile offset times longer than 200 ms, the facilitatory effect declined and performance on the A+T condition was similar to that in the unimodal conditions.

B. Comparisons with previous multisensory work

The facilitatory effects obtained for simultaneous presentation of A+T signals in our baseline experiment, as well as the effects of temporal asynchrony of the auditory and tactile stimuli, are generally consistent with previous reports in the literature. Facilitative interactions for synchronously presented auditory and tactile stimuli were reported by Schnupp *et al.* (2005) using objective techniques to measure the discriminability of visual (V), auditory (A), and tactile (T) stimuli in VA, VT, and AT combinations. Auditory stimuli were 100–ms bursts of broadband noise presented at a background reference (sound) level of 51 dB SPL. Tactile stimuli were 100–ms bursts of 150–Hz sinusoidal vibrations presented at background reference (force) levels of 16.2–48.5 N. The stimulus on a given trial was a simultaneous pair of either VA, VT, or AT bursts that ranged from 0% to 14% (V and A) or from 0% to 35% (T) in 2% or 5% increments of intensity relative to the background reference level. Observers were instructed to respond whether the background level or an incremented level was presented. Data were analyzed in terms of analogs of both the PSM and ASM. While 2 of 5 AT data sets could be adequately accounted for by the ASM (Schnupp, 2009, personal communication with L.D. Braida) and 5 of 17 data sets overall, all 17 could be accounted for by the PSM.

Ro *et al.* (2009) measured the effect of presenting a relatively intense (59 dB) 500-Hz, 200-ms tone on the detection of a near-threshold 0.3 ms square-wave electrocutaneous stimulus that felt like a faint tap. They found that the presentation of the auditory stimulus increased *d*-prime from 2.4 to 2.8. This result was interpreted as evidence that "a task-irrelevant sound can enhance somatosensory perception."

Facilitative interactions have also been observed using subjective techniques such as loudness matching (Schurmann *et al.*, 2004; Yarrow *et al.*, 2008) and loudness magnitude estimation (Gillmeister and Eimer, 2007). In the two loudness-matching studies, the average intensity required to produce equal loudness of an auditory reference tone was 12%–13% (roughly 0.5 dB) lower under the combined auditory-tactile condition compared with the auditory-alone condition, thus suggesting a facilitative interaction between the auditory and tactile stimuli. Gillmeister and Eimer (2007) found that magnitude estimates of an auditory tone presented in a background of white noise were increased by simultaneous presentation of a tactile stimulus for near-threshold auditory tones, but no loudness increase was observed either for higher intensity tones or for non-simultaneous presentation of the tactile and auditory stimuli. It should be noted, however, that based on the results of other experiments, Yarrow *et al.* (2008) attribute the increase in loudness to a bias effect. They conclude that the tactile stimulus "does not affect auditory judgments in the same manner as a real tone."

Other previous studies of auditory-tactile integration have measured effects of temporal asynchrony between the two stimuli and have also demonstrated dependence on the order of stimulus presentation: Gescheider and Niblette (1967) for inter-sensory masking and temporal-order judgments and Bresciani *et al.* (2005) for judgments of auditory numerosity. Consistent with the results of the current study, higher levels of interaction between the two senses were obtained for conditions in which the tactile stimulus is delivered before the auditory stimulus. One exception to this pattern is found in the results of Gillmeister and Eimer (2007). While demonstrating effects of temporal synchrony on the detectability of an auditory tone in the presence of a vibratory pulse, they found no effects of stimulus order. Their detectability results, however, are consistent with the results of their loudness-estimation study.

While the experimental conditions used in these studies differ from one another, they all suggest that temporal synchrony is an important factor in showing facilitative auditory-tactile interaction. The current study has shown in greater detail the asymmetry in the temporal window involved in auditory-tactile detection, such that when the tactile stimulus precedes the auditory by up to 200 ms, a facilitative interaction significantly greater than the unimodal levels is measured. This level of response is not seen when the auditory stimulus precedes the tactile, however, as bimodal responses at all asynchronous time periods are not different from unimodal levels.

C. Implications of model results

The amount of integration measured in this study was quantified by comparing performance with the predictions of three models of the integration process: the OSCM, the PSM, and the ASM. It should be noted that these models are not mutually exclusive in the sense that observers need not base their decisions exclusively on one model in all experiments. If the auditory stimulus is presented before the tactile, it is unlikely that the ASM would apply, while it might apply when there is temporal overlap. Also, the predictions of more than one model may fit the data equally well. For example, in the hypothetical case considered in Sec. II F, based on 75 trials the score of 75%-correct would be within 2 standard deviations of the predictions of all three models. Many of the two-frequency results of Marrill (1956) can be accounted for by two of these three models. It is only possible to distinguish among the three models based on more than one experimental result, i.e., several results from one observer or the results of multiple observers. When performance exceeds the predictions of the OSCM, this implies at least partial integration of cues, and when performance exceeds predictions of the PSM, this implies at least partial within-channel integration.

In this study, the results show that measurements are more often successfully modeled by the PSM approach than by the OSCM or ASM approaches and are consistent with those found previously in auditory-alone studies (Green, 1958), tactile-alone studies (Bensmaia *et al.*, 2005), and in multisensory studies (audio-visual and audio-visual-tactile: Braida, 1991; visual-tactile: Ernst and Banks, 2002; discrimination of pairs of visual-auditory, visual-tactile, and audio-

tactile stimuli: Schnupp *et al.*, 2005). Although most of these studies did not attempt to model the observations with an ASM, Schnupp (2009), personal communication with L.D. Braida, found that two of five audio-tactile discrimination data sets could be fit by an ASM (all five were fit by a PSM), and we found in experiment 1 that nearly the same number of experiments were accounted for by the ASM as by the PSM.

Thus, we found, in accord with Schnupp *et al.* (2005), that overall the PSM best accounts for the improvement in detectability when auditory and tactile stimuli are combined. There are significant differences, however: The OSCM provides a slightly better account when auditory stimuli precede tactile stimuli, and the ASM provides nearly as good an account of the (non-)effects of varying relative auditory-tactile phase. One problem with this interpretation is that the different models make predictions of detectability that are always ordered: OSCM $\leq$ PSM $\leq$ ASM. Thus, for example, if an observer behaves in accord with the ASM but makes a few responses due to inattention, the PSM will tend to be favored. While we discarded data sets for which there were indications that unimodal observer detection had decreased during the course of a single session, it is likely that some reduction in bimodal detection may have occurred as well. Because Schnupp *et al.* (2005) collected data over two or three sessions, it is also possible that criterion shifts may have reduced apparent performance, thus favoring the PSM over the ASM.

It is also possible that the PSM provides a better description of the data than the ASM when qualitatively different stimuli are detected or discriminated. The traditional explanation for the two-frequency detection results of Marrill (1956) and Green (1958) is that the PSM provides a good account of the detection of pairs of tones whose frequencies lie in distinct critical bands while the ASM is appropriate for tones whose frequencies lie in the same critical band. Wilson *et al.* (2008), who tested the detection of auditory and tactile tones of varying frequency, found that performance generally declined as the frequency difference increased. It is possible that Schnupp *et al.* (2005) found that a PSM-like model applied to discrimination of auditory noise and a tactile tone for this reason.

Stein and Meredith (1993) suggested that additive and super-additive responses are a way of measuring facilitative multisensory responses. The different models suggest different mechanisms for integration, with the Pythagorean sum modeling two independent pathways integrating the different stimuli after each has been processed by its own sensory system and the algebraic sum modeling stimuli that are integrated before being processed, leading to a greater level of integration overall. It is possible that both the results of Schnupp *et al.* (2005) and our results, which show that subjects can utilize both Pythagorean and algebraic approaches to integration, suggest that the auditory and tactile sensory systems are capable of integrating in both manners, and both mechanisms are being employed during our experiment.

D. Relationship to neuroanatomy

One potential anatomical pathway for Pythagorean integration may be the ascending somatosensory inputs to the somatosensory cortex, which then project to the auditory cortex. Thus, two independent pathways are operating on input from each of the modalities, and the multisensory stimuli are processed only after the single-modality operations have taken place. A different anatomical pathway that may account for algebraic integration comes from the ascending somatosensory inputs that target early auditory centers (i.e., in the brainstem and thalamus) and thereby affect changes in auditory-tactile integration before the combined signal reaches the auditory cortex. The fact that we see observed responses that are greater than the prediction of PSM suggests that the auditory and somatosensory systems are working together in one multisensory area to process the stimuli.

V. CONCLUDING REMARKS

Our study has shown that certain combinations of auditory and tactile signals result in a significant increase in detectability above the levels when the stimuli are presented in isolation. This is not due to changes in response bias (e.g., Yarrow *et al.*, 2008), as indicated by a detection theory analysis. Specifically, we have shown significant increases in detectability that are independent of relative auditory-tactile phase when the auditory and tactile stimuli are presented simultaneously, suggesting that the envelopes, and not the fine structure, of the two signals interact in a facilitative manner. Additionally, we have also shown asymmetric changes in detectability when the two signals are presented with temporal asynchrony: When the auditory signal is presented first, detectability is not significantly greater than in A-alone or T-alone conditions, but when the tactile signal is presented first, detectability is significantly greater for almost all values of SOA employed. These differences are consistent with the neural mechanics of auditory-on-auditory masking and tactile-on-tactile masking.

Our results were compared with three models of integration. While it is not always possible to differentiate among the models on the basis of a single experimental outcome, the models sort themselves out if one combines results across sessions and/or observers. If one assumes that all observers use a single model in all experiments, then the PSM gives a better fit to the data than the OSCM or the ASM.

Further research is being conducted to examine the effects of other stimulus parameters (including frequency and intensity) on the perceptual aspects of auditory-tactile integration.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institutes of Health (Grant Nos. 5T32-DC000038, R01-DC000117, and R01-DC00126) and by a Hertz Foundation Fellowship (E.C.W.). The authors would like to thank D.M. Green for helpful discussions, J.W.H. Schnupp for sharing data, and Richard Freyman, Fan-Gang Zeng, and the anonymous reviewer for helpful comments on earlier versions of this manuscript.

APPENDIX: ESTIMATIONS OF BONE-CONDUCTED SOUND LEVELS ARISING FROM VIBROTACTILE STIMULATION AT 250 HZ

We consider two possibilities and show that they are unlikely to be responsible for our results: (1) In baseline conditions the vibratory stimulus is detected through the auditory sense. (2) In Experiment 1, the phase dependent combination of vibratory and acoustic stimuli is responsible for the phase dependence of our results. Note that masking noise was used in an attempt to ensure that the task is performed solely through the sense of touch without spurious auditory cues.

Consider first the possibility that bone-conducted sound from vibratory stimulation was responsible for detection of the tactile stimulus. In the measurements of Dirks *et al.* (1976), bone-conduction thresholds for normal listeners in force and acceleration units indicate that the 250-Hz bone-conduction threshold, when measured with a vibrator placed on the mastoid, is 10 dB *re* 1 cm/s^2 (acceleration units). The maximum displacement of our 250-Hz signal (roughly 5 dB SL) corresponds to a peak displacement of −20 dB *re* 1 μm peak and an acceleration of roughly 5 dB *re* 1 cm/s^2, roughly 5 dB less than the bone-conduction threshold for mastoid stimulation. It is fairly safe to assume that stimulation of the middle finger results in a highly attenuated bone-conducted signal compared to stimulation of the mastoid. The bone-conducted threshold at the forehead is 12 dB higher than at the mastoid. The impedance mismatches created by tissue and bone junctions from the fingertip to the skull would lead to even higher thresholds, perhaps by 13 dB, than for the forehead. Thus, the highest signal reaching the ear through bone-conducted sound at 250 Hz would be −20 dB *re* 1 cm/s^2. The bone-conducted threshold at 250Hz is 10 dB cm/s^2; thus, our bone-conducted stimulus would be roughly −30 dB SL, that is, roughly 30 dB below the air-conducted threshold of 18 dB SPL at 250 Hz (Houtsma, 2004) or equivalent to an acoustic stimulus of −12 dB SPL. Such bone-conducted sound would be undetectable.

Assuming a critical ratio of 17.5 dB at 250 Hz and a noise spectral level of 7.4 dB/Hz, the level of the acoustic tone is roughly 25 dB SPL, and (as noted above) the equivalent vibratory stimulus is −12 dB SPL, 37 dB below the level of the acoustic tone. This would cause the 25 dB SPL tone to vary at most from 24.9 to 25.1 dB SPL as the phase is changed. To understand the effect of this phase change, we make use of some unpublished data on the detection of auditory stimuli of different amplitudes: 25 and 27 dB, which correspond to detection rates in the 50 dB SPL noise of 70.9%- and 79.9%-correct, respectively, or about 4.5 percentage points per decibel. Thus, the combination of the bone- and the air-conducted sound would cause the detection rate to change from 70.3% to 71.5%. This is contrary to the results of Experiment 1, which indicate that in the A+T condition in scores varied between 83.2%- and 84.6%-correct with standard errors of less than 1.3 percentage points. This indicates that the effect of combining the vibratory and acoustic stimuli cannot be accounted for by bone conduction alone.

[1]Data collected from an additional two subjects (S_5 in experiment 1 and S_7 in experiment 2C) were discarded on the basis of abnormally low values of thresholds for the tactile stimuli that were inconsistent with those of the other subjects and with results in the literature.

[2]Three subjects (S_2, S_4, and S_7) were also tested in two additional conditions in experiment 1 (phase=0°, SOA=+600 ms, and SOA=−600 ms) in addition to the four phase conditions. These subjects later participated in experiment 2B, and SOA values of ±600 ms were not repeated.

[3]Due to experimenter error, performance on the combined A+T (SOA =0 ms) condition was not measured in several of the experiment 2B test sessions for three subjects (S_2, S_4, and S_7), although performance on A-alone and T-alone conditions was always established at the beginning of each session.

[4]We denote *d*-primes that can be estimated directly from the data using lower case (d'), *d*-primes that are predicted by models in upper case letters (D').

[5]Although both types of interactions might occur simultaneously and cancel, we regard this possibility as unlikely.

Bensmaia, S. J., Hollins, M., and Yau, J. (**2005**). "Vibrotactile intensity and frequency information in the Pacinian system," Percept. Psychophys. **67**, 828–841.

Bhattacharya, J., Shams, L., and Shimojo, S. (**2002**). "Sound-induced illusory flash perception: Role of gamma band responses," NeuroReport **13**, 1727–1730.

Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments," Q. J. Exp. Psychol. **43A**, 647–677.

Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., and Kheddar, A. (**2005**). "Feeling what you hear: Auditory signals can modulate tactile tap perception," Exp. Brain Res. **162**, 172–180.

Caclin, A., Soto-Faraco, S., Kingstone, A., and Spence, C. (**2002**). "Tactile 'capture' of audition," Percept. Psychophys. **64**, 616–630.

Caetano, G., and Jousmaki, V. (**2006**). "Evidence of vibrotactile input to human auditory cortex," Neuroimage **29**, 15–28.

Cappe, C., and Barone, P. (**2005**). "Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey," Eur. J. Neurosci. **22**, 2886–2902.

Craig, J. C., and Evans, P. M. (**1987**). "Vibrotactile masking and the persistence of tactual features," Percept. Psychophys. **42**, 309–317.

Dirks, D. D., Kamm, C., and Gilman, S. (**1976**). "Bone-conduction thresholds for normal listeners in force and acceleration units," J. Speech Hear. Res. **19**, 181–186.

Ernst, M. O., and Banks, M. S. (**2002**). "Humans integrate visual and haptic information in a statistically optimal fashion," Nature (London) **415**, 429–433.

Fastl, H. (**1977**). "Subjective duration and temporal masking patterns of broadband noise impulses," J. Acoust. Soc. Am. **61**, 162–168.

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., Ritter, W., and Murray, M. M. (**2002**). "Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study," J. Neurophysiol. **8**, 540–543.

Fu, K. M. G., Johnston, T. A., Shah, A. S., Arnold, L., Smiley, J., Hackett, T. A., Garraghty, P. E., and Schroeder, C. E. (**2003**). "Auditory cortical neurons respond to somatosensory stimulation," J. Neurosci. **23**, 7510–7515.

Gescheider, G. A., Barton, W. G., Bruce, M. R., Goldberg, J. H., and Greenspan, M. J. (**1969**). "Effects of simultaneous auditory stimulation on the detection of tactile stimuli," J. Exp. Psychol. **81**, 120–125.

Gescheider, G. A., Bolanowski, S. J., and Verrillo, R. T. (**1989**). "Vibrotactile masking: Effects of stimulus onset asynchrony and stimulus frequency," J. Acoust. Soc. Am. **85**, 2059–2064.

Gescheider, G. A., Hoffman, K. E., Harrison, M. A., and Travis, M. L. (**1994**). "The effects of masking on vibrotactile temporal summation in the detection of sinusoidal and noise signals," J. Acoust. Soc. Am. **95**, 1006–1016.

Gescheider, G. A., Kane, M. J., and Sager, L. C. (**1974**). "The effect of auditory stimulation on responses to tactile stimuli," Bull. Psychon. Soc. **3**, 204–206.

Gescheider, G. A., and Migel, N. (**1995**). "Some temporal parameters in vibrotactile forward masking," J. Acoust. Soc. Am. **98**, 3195–3199.

Gescheider, G. A., and Niblette, R. K. (**1967**). "Cross-modality masking for touch and hearing," J. Exp. Psychol. **74**, 313–320.

Gillmeister, H., and Eimer, M. (**2007**). "Tactile enhancement of auditory detection and perceived loudness," Brain Res. **1160**, 58–68.

Green, D. M. (**1958**). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. **30**, 904–911.
Guest, S., Catmur, C., Lloyd, D., and Spence, C. (**2002**). "Audiotactile interactions in roughness perception," Exp. Brain Res. **146**, 161–171.
Hackett, T. A., de la Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C. E. (**2007**). "Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane," J. Comp. Neurol. **502**, 924–952.
Hamer, R. D., Verrillo, R. T., and Zwislocki, J. J. (**1983**). "Vibrotactile masking of pacinian and non-pacinian channels," J. Acoust. Soc. Am. **73**, 1293–1303.
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.
Houtsma, A. J. (**2004**). "Hawkins and Stevens revisited with insert earphones," J. Acoust. Soc. Am. **115**, 967–970.
Jesteadt, W., Bacon, S. P., and Lehman, J. R. (**1982**). "Forward masking as a function of frequency, masker level, and signal delay," J. Acoust. Soc. Am. **71**, 950–962.
Jousmaki, V., and Hari, R. (**1998**). "Parchment-skin illusion: Sound-biased touch," Curr. Biol. **8**, R190–191.
Kayser, C., Petkov, C. I., Augath, M., and Logothetis, N. K. (**2005**). "Integration of touch and sound in auditory cortex," Neuron **48**, 373–384.
Kidd, G., Jr., and Feth, L. L. (**1981**). "Patterns of residual masking," Hear. Res. **5**, 49–67.
Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., and Schroeder, C. E. (**2007**). "Neuronal oscillations and multisensory interaction in primary auditory cortex," Neuron **53**, 279–292.
Lamore, P. J., Muijser, H., and Keemijik, C. J. (**1986**). "Envelope detection of amplitude-modulated high-frequency sinusoidal signals by skin mechanoreceptors," J. Acoust. Soc. Am. **79**, 1082–1085.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Marrill, T. (**1956**). "Detection theory and psychophysics." Technical Report No. 319, Research Laboratory of Electronics, MIT, Cambridge, MA.
McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay; implications for suppression in noise," J. Acoust. Soc. Am. **73**, 1249–1259.
Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Acoust. Soc. Am. **83**, 1102–1116.
Neville, A. M., and Kennedy, J. B. (**1964**). *Basic Statistical Methods for Engineers and Scientists* (International Textbook, Scranton, PA), p. 133.
Plack, C. J., and Oxenham, A. J. (**1998**). "Basilar-membrane nonlinearity and the growth of forward masking," J. Acoust. Soc. Am. **103**, 1598–1608.
Rabinowitz, W. M., Houtsma, A. J. M., Durlach, N. I., and Delhorne, L. A. (**1987**). "Multidimensional tactile displays: Identification of vibratory intensity, frequency and contactor area," J. Acoust. Soc. Am. **82**, 1243–1252.
Ro, T., Hsu, J., Yasar, N. E., Elmore, L. C., and Beauchamp, M. S. (**2009**). "Sound enhances touch perception," Exp. Brain Res. **195**, 135–143.
Robinson, C. E., and Pollack, I. (**1973**). "Interaction between forward and backward masking: A measure of the integrating period of the auditory system," J. Acoust. Soc. Am. **53**, 1313–1316.
Schnupp, J. W. H., Dawe, K. L., and Pollack, G. (**2005**). "The detection of multisensory stimuli in an orthogonal sensory space," Exp. Brain Res. **162**, 181–190.
Schroeder, C. E., Lindskey, R. W., Specht, C., Marcovici, A., Smiley, J. F., and Javitt, D. C. (**2001**). "Somatosensory input to auditory association cortex in the macaque monkey," J. Neurophysiol. **85**, 1322–1327.
Schurmann, M., Caetano, G., Hlushchuk, Y., Jousmaki, V., and Hari, R. (**2006**). "Touch activates human auditory cortex," Neuroimage **30**, 1325–1331.
Schurmann, M., Caetano, G., Jousmaki, V., and Hari, R. (**2004**). "Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels," J. Acoust. Soc. Am. **115**, 830–832.
Sekuler, R., Sekuler, A. B., and Lau, R. (**1997**). "Sound alters visual motion perception," Nature (London) **385**, 308.
Smiley, J. F., Hackett, T. A., Ulbert, I., Karmas, G., Lakatos, P., Javitt, D. C., and Schroeder, C. E. (**2007**). "Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys," J. Comp. Neurol. **502**, 894–923.
Soto-Faraco, S., Spence, C., and Kingstone, A. (**2004**). "Congruency effects between auditory and tactile motion: Extending the phenomenon of cross-modal dynamic capture," Cogn. Affect. Behav. Neurosci. **4**, 208–217.
Stein, B. E., London, N., Wilkinson, L. K., and Price, D. D. (**1996**). "Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis," J. Cogn. Neurosci. **8**, 497–506.
Stein, B. E., and Meredith, M. A. (**1993**). *The Merging of the Senses* (MIT, Cambridge, MA).
Verrillo, R. T., Gescheider, G. A., Calman, B. G., and Van Doren, C. L. (**1983**). "Vibrotactile masking: Effect of one- and two-site stimulation," Percept. Psychophys. **33**, 379–387.
Vogten, L. L. M. (**1978**). "Simultaneous pure-tone masking: The dependence of masking asymmetries on intensity," J. Acoust. Soc. Am. **63**, 1509–1519.
Wilson, E. C., Reed, C. M., and Braida, L. D. (**2008**). "Perceptual interactions between vibrotactile and auditory stimuli: Effects of frequency," in International Multisensory Research Forum Conference, Hamburg, Germany, July.
Woods, T. M., and Recanzone, G. H. (**2004**). "Cross-modal interactions evidenced by the ventriloquism effect in humans and monkeys," in *Handbook of Multisensory Processes*, edited by G. Calvert, C. Spence, and B. E. Stein (MIT, Cambridge, MA), pp. 35–48.
Yarrow, K., Haggard, P., and Rothwell, J. C. (**2008**). "Vibrotactile-auditory interactions are post-perceptual," Perception **37**, 1114–1130.
Zhou, J., and Shore, S. (**2004**). "Projections from the trigeminal nuclear complex to the cochlear nuclei: A retrograde and anterograde tracing study in the guinea pig," J. Neurosci. Res. **78**, 901–907.

Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition

Huw R. Cooper and Brian Roberts[a)]
Psychology, School of Life and Health Sciences, Aston University, Birmingham B4 7ET, United Kingdom

(Received 7 April 2009; revised 7 July 2009; accepted 8 July 2009)

The evidence that cochlear implant listeners routinely experience stream segregation is limited and equivocal. Streaming in these listeners was explored using tone sequences matched to the center frequencies of the implant's 22 electrodes. Experiment 1 measured temporal discrimination for short (ABA triplet) and longer (12 AB cycles) sequences (tone/silence durations = 60/40 ms). Tone A stimulated electrode 11; tone B stimulated one of 14 electrodes. On each trial, one sequence remained isochronous, and tone B was delayed in the other; listeners had to identify the anisochronous interval. The delay was introduced in the second half of the longer sequences. Prior build-up of streaming should cause thresholds to rise more steeply with increasing electrode separation, but no interaction with sequence length was found. Experiment 2 required listeners to identify which of two target sequences was present when interleaved with distractors (tone/silence durations = 120/80 ms). Accuracy was high for isolated targets, but most listeners performed near chance when loudness-matched distractors were added, even when remote from the target. Only a substantial reduction in distractor level improved performance, and this effect did not interact with target-distractor separation. These results indicate that implantees often do not achieve stream segregation, even in relatively unchallenging tasks.
 [DOI: 10.1121/1.3203210]

PACS number(s): 43.66.Mk, 43.66.Ts, 43.66.Fe, 43.66.Hg [BCM] Pages: 1975–1987

I. INTRODUCTION

Grouping together only those acoustic elements that arise from a common source is an important function of the auditory system, and the extent to which this is possible has a major impact on one's ability to recognize auditory objects in complex listening environments. Normal-hearing (NH) and cochlear implant (CI) listeners alike must solve this scene analysis problem (Bregman, 1990), but CI listeners are much less well equipped for this challenge owing to the impoverished sensory information available via their implants (e.g., Throckmorton and Collins, 2002; Fu and Nogaki, 2005). However, relatively few studies have investigated auditory grouping in CI listeners, and even fewer have used accuracy of performance as a measure of perceptual organization with implant listening. The experiments reported here used temporal discrimination and interleaved melody recognition to measure the extent to which CI listeners typically experience auditory stream segregation.

Bregman (1990) proposed two distinct processes in the construction of auditory representations: "primitive" scene analysis and schema-based selection. The first is driven by the incoming acoustic data and is often assumed to be involuntary and pre-attentive. The latter is assumed to be voluntary and to involve, through selective attention, the "activation of stored knowledge of familiar patterns or schemas in the acoustic environment and of a search for confirming stimulation in the auditory input" (Bregman, 1990, page 397). The distinction between these processes is reflected in the findings of van Noorden (1975). He examined the perceptual organization of sequences comprising alternating high- and low-frequency pure tones that differed in frequency separation and presentation rate (tone onset-to-onset time). van Noorden (1975) distinguished between the temporal coherence boundary (TCB) and the fission boundary (FB) by varying the instructions given to his participants, and hence their listening "set." He found that the sequence could not be heard as a single stream above the TCB presumably because of the obligatory operation of primitive stream segregation, whereas it could not be heard as two streams below the FB presumably because of the limits of attentional selection. The TCB was influenced by frequency separation and rate; larger separations were required to induce streaming at slower rates. In contrast, the FB showed little dependence on rate; sounds could be selected voluntarily from a larger set if their frequency separation was larger than a minimum critical value (a few semitones).

A. Temporal discrimination as a streaming measure in NH listeners

The perceptual properties of sound events are computed within streams but not across them. Therefore, one approach to measuring auditory stream segregation is to use a task for which performance will be affected adversely if streaming occurs. For example, performance is impaired in tasks that require listeners to count or report the order of sequences of pure tones if these sounds are heard to split into separate

[a)]Author to whom correspondence should be addressed. Electronic mail: b.roberts@aston.ac.uk

streams on the basis of frequency separation or rate (Warren *et al.*, 1969; Bregman and Campbell, 1971). Similarly, stream segregation increases thresholds for detecting changes in rhythm arising from temporal asymmetries between sequentially presented tones (e.g., van Noorden, 1975; Vliegen *et al.*, 1999; Cusack and Roberts, 2000; Roberts *et al.*, 2002). Given that listeners must attempt to hear the sequence as a single stream to perform well in these tasks, the results should indicate the limits of their ability to overcome primitive stream segregation.

One important caveat regarding this approach merits note. Even for two isolated pure tones, for which one would expect little or no build-up in the tendency for stream segregation (Bregman, 1978), thresholds for detecting a temporal gap between them rise for NH listeners as the frequency separation increases (e.g., Grose *et al.*, 2001). Thus, the effects of frequency differences between tones on performance in gap discrimination tasks do not always reflect increased stream segregation. Indeed, gap detection thresholds have been shown to rise in CI listeners as the physical separation of stimulated electrode pairs is increased; Hanekom and Shannon (1998) used this as a measure of electrode interaction, not of streaming.[1] Therefore, an increase in temporal-discrimination thresholds with greater frequency separation between two stimuli cannot in itself be taken as firm evidence of stream segregation.

To demonstrate convincingly that changes in threshold reflect changes in stream segregation, it is necessary to observe a greater dependence of temporal judgments on frequency separation in situations where stream segregation is expected to be greater, i.e., when there has been sufficient time for the tendency for segregation to build up. This was first demonstrated by van Noorden (1975), who measured the just perceptible displacement in time (ΔT) of the B tones from the midpoint between neighboring A tones in long sequences of alternating AB tones and in short ABA "triplets." There was a very clear dependence of ΔT on the frequency ratio between tones A and B for the long sequences, but there was far less dependence of ΔT on the AB ratio for the triplets.

B. Temporal discrimination as a streaming measure in implant listeners

To date, the only published example of this approach being applied to CI listeners is the study by Hong and Turner (2006). They used sequences of pure tones configured in the same way as those described by Roberts *et al.* (2002), where the target stimulus began with an isochronous portion to allow significant build-up of the strength of stream segregation before the anisochrony was introduced. Hence, the ability to detect the irregular rhythm was taken to indicate an absence of streaming. Stimuli were presented via a loudspeaker at 90 or 95 dB SPL (sound pressure level) to eight CI listeners, who used their normal speech processors and "maps" (speech processing strategies and stimulus levels), and were allowed to adjust their own volume or sensitivity settings if desired. Tone A was set at 200, 800, or 2000 Hz, while the frequency of tone B was set at a ratio $[(B-A)/A]$ of 0 (i.e., A=B), 0.01, 0.1, 0.5, 1.0, or 3.0. Using an adaptive two-interval two-alternative forced-choice task, they measured the threshold delay for tone B over this range of separations from each "base frequency" for tone A. The results were normalized relative to the threshold delay for the case where the ratio was 0. Hong and Turner (2006) found considerable variability in the slopes of the function relating threshold to frequency separation for their CI listeners. The AB frequency differences were converted into electrode separations based on the frequency allocations to each electrode for the individual CI listeners, and significant linear regressions were found between the normalized threshold delays and electrode separations for some (though not all) of the listeners.

In an attempt to determine whether the observed relationship between frequency (or electrode) separation and threshold delay truly reflected stream segregation, they repeated the task with three NH and three CI listeners at two ratios (0 and 0.5) using isolated ABA tone triplets ("short rhythm task") as well as the longer sequences ("streaming task"). For the three base frequencies, they found that the normalized thresholds for the ratio of 0.5 were usually greater for the longer sequences than for the triplets. In effect, they found an interaction between tone sequence length and frequency separation, which they concluded was evidence that the task was a genuine measure of stream segregation. They went on to report correlations between normalized thresholds in their streaming task and speech reception thresholds in noise.

In contrast with Hong and Turner (2006), Cooper and Roberts (2007) argued that most CI listeners typically experience little or no auditory stream segregation. Listeners in their task were required to report whether they heard one or two streams when listening to sequences of alternating high and low tones. A significant relationship was found between electrode separation and reported segregation, but sequence rate had essentially no effect on listeners' judgments. In contrast, rate changes usually have a major effect on streaming judgments by NH listeners (van Noorden, 1975). Furthermore, there was little evidence of the perceptual instability for intermediate electrode separations that would have been predicted if stream segregation were occurring (Anstis and Saida, 1985). A similar effect of electrode separation on reported segregation was observed by Chatterjee *et al.* (2006), but they did not explore the effect of changes in rate. Cooper and Roberts (2007) concluded that their results probably reflected simple channel discrimination rather than streaming judgments.

What might account for the discrepancy between the findings of these two studies? One possibility concerns the listening set of the participants in the study of Cooper and Roberts (2007), for which listeners were not instructed to try and hear the stimuli in any particular way. However, as noted by Cooper (2008), while it is true that the dependence of reported segregation on sequence rate may not be as strong under neutral instructions as when listeners are asked to try and hear a single stream, there is clear experimental evidence of this dependence under neutral instructions in NH listeners (e.g., Anstis and Saida, 1985, experiment 1). Another possibility concerns underlying differences between measures of

streaming based on subjective report and on temporal discrimination. Although these measures are often regarded as closely related (e.g., Roberts *et al.*, 2002), there is recent evidence of discrepancies between them [compare the results of Roberts *et al.* (2008) with those of Rogers and Bregman (1993, 1998)]. However, so far these discrepancies appear to be restricted to studies concerned specifically with the resetting of the build-up of stream segregation, which does not apply here.

The results reported by Hong and Turner (2006) should be treated with some caution because there are a number of caveats relating to their experimental methods. First, stimuli were presented via loudspeaker rather than by direct stimulation, and so the effects of each listener's speech processing strategy on the resulting patterns of electrical stimulation are not entirely clear. Indeed, unlike our listeners, a mixture of different implant systems and speech-coding strategies was used among their participants, and they were allowed to adjust the volume or sensitivity levels of their speech processors. Therefore, the precise characteristics of the electrical stimuli that were delivered are unknown. Second, frequency separation was the experimental parameter directly manipulated, and this was only converted into electrode separation afterward. Again, this inevitably leads to some uncertainty about the precise stimuli that were delivered to the electrodes of the listeners' implants. Finally, only three CI listeners (one user of the 22-channel Nucleus implant and two users of the 16-channel Clarion implant) participated in their critical second experiment, which compared longer sequences and triplets, and only two frequency separations were tested (ratios of 0 and 0.5). As noted by Cooper and Roberts (2007), results based on more precisely controlled stimuli for a wider range of electrode separations, and from more participants, would be required to provide convincing evidence of involuntary stream segregation in the majority of implant listeners. Therefore, experiment 1 reported here revisited the use of a temporal-discrimination task to explore stream segregation in CI listeners. The experiment was designed to avoid some of the limitations associated with Hong and Turner's (2006) methods.

C. Interleaved melody recognition and schema-based stream segregation

One example of a situation that would be expected to be influenced by schema-based streaming is when the listener has to select and recognize a familiar melody or pattern of tones from a sequence containing interleaved interfering sounds (distractors). Dowling (1973) described a task of this type, in which he played to NH listeners pairs of familiar melodies (e.g., "Happy Birthday") that were temporally interleaved, i.e., where the tones of one melody alternated with the tones of another. The listeners' task was to name either of the melodies. He found that their pitch ranges should not overlap for identification to be successful.

Cusack and Roberts (2000) modified Dowling's (1973) task such that an arbitrary six-tone melody was heard once in isolation, followed by two intervals in which the melody was interleaved with distractors. The melody was identical to the isolated version in one interval and was modified in the other interval; the task was to identify the interval containing the exact match. Differences between targets and distractors in pitch range and in timbre (pure tones vs narrow-band noises) both improved performance considerably. Bey and McAdams (2002, 2003) have since used a similar/different task to compare the case where an isolated arbitrary melody (comparison) is heard before or after the interleaved stimulus. Listeners were better able to judge whether the interleaved melody was identical to the comparison melody when the comparison was presented beforehand, indicating that prior knowledge of the melody to be extracted from the mixture enabled a contribution from a schema-based process for segregation. Thus, tasks requiring the recognition of interleaved melodies have been effective in experimental studies of auditory stream segregation in NH listeners. Such tasks are likely to involve both primitive and schema-driven segregation processes but are probably dominated by the latter.

We are aware of only one study that has explored the ability of CI listeners to select a subset of acoustic elements from a longer sequence, and only a short summary of this study has been published (Chatterjee and Galvin, 2002). These authors used repeating patterns of loudness-matched stimuli that were composed of two or three different tones (tone duration = 50 ms; inter-tone interval = 50 ms), and the tonotopic distance (electrode pair separation) between these tones was varied. Before each test sequence, listeners heard a "preview" sequence with a rhythm corresponding to a subset of the tones and were asked to judge whether or not they could hear this rhythm within the test sequence. The proportion of positive responses was taken as evidence that the subset of elements could be heard out as a separate perceptual stream. Chatterjee and Galvin (2002) found that this proportion was very dependent on tonotopic distance, as would be expected for NH listeners. Although their method was described as a yes/no task, the preview was always contained within the longer sequence, and so the task was subjective in nature. Experiment 2 reported here used an interleaved melody task to evaluate schema-based segregation in CI listeners using an accuracy of performance measure, to our knowledge for the first time.

II. EXPERIMENT 1

A. Overview

To investigate stream segregation in CI listeners using a temporal-discrimination task, experiment 1 employed the same task and stimulus sequence configurations used by Roberts *et al.* (2002) and also used ABA tone triplets. In contrast with Hong and Turner (2006), the stimuli were presented via direct input to the listeners' speech processors. The frequencies of the tones used were such that they fell in the center of the passband for the intended electrodes (see Table I). The speech processor was programed so that it could deliver stimulation on only one channel at a time using the advanced combination encoders (ACE) strategy and selecting only one spectral peak. The precise pattern of electrical stimulation was verified by routing the stimuli through an experimental speech processor and by analyzing the output using a "dummy" implant within the manufacturer's com-

TABLE I. Frequency characteristics of the tones used in experiments 1 and 2 and their relation to the implant's 22 channels. Lower- and higher-numbered electrodes correspond to more basal and more apical places, respectively. Rows shown in bold indicate the restricted set of electrodes used in experiment 1. All electrode numbers were used in experiment 2.

Channel/electrode number	Lower frequency boundary (Hz)	Upper frequency boundary (Hz)	Channel center frequency and frequency of the pure tone stimuli (Hz)
1 (most basal)	**6938**	**7938**	**7438**
2	6063	6938	6500
3	**5313**	**6063**	**5688**
4	4688	5313	5000
5	**4063**	**4688**	**4375**
6	3563	4063	3813
7	**3063**	**3563**	**3313**
8	2688	3063	2875
9	**2313**	**2688**	**2500**
10	**2063**	**2313**	**2188**
11	**1813**	**2063**	**1938**
12	**1563**	**1813**	**1688**
13	**1313**	**1563**	**1438**
14	**1188**	**1313**	**1250**
15	**1063**	**1188**	**1125**
16	938	1063	1000
17	**813**	**938**	**875**
18	688	813	750
19	**563**	**688**	**625**
20	438	563	500
21	**313**	**438**	**375**
22 (most apical)	188	313	250

puter interface. This allowed a frame-by-frame listing of the output of the implant's transmitter coil and the generation of an "electrodogram," which illustrates visually the output on each electrode over the selected time window. This procedure replaced checking with a test implant and oscilloscope.

All stimuli were presented at C level (maximum comfortable level) and were loudness balanced using a standard clinical method of "sweeping" stimulation across the electrode array; listeners had no control over the volume or sensitivity. Thus, the relation between electrode separation and temporal discrimination was measured directly across a wide range of separations, both for the longer tone sequences and the short triplets, in a group of CI listeners all using the same type of Nucleus implant. An interaction between sequence length and electrode separation, across a range of AB frequency separations, would provide more convincing evidence that stream segregation had indeed mediated the results; in this case, one would expect to see a steeper function relating threshold to electrode separation for the longer tone sequences than for the triplets.

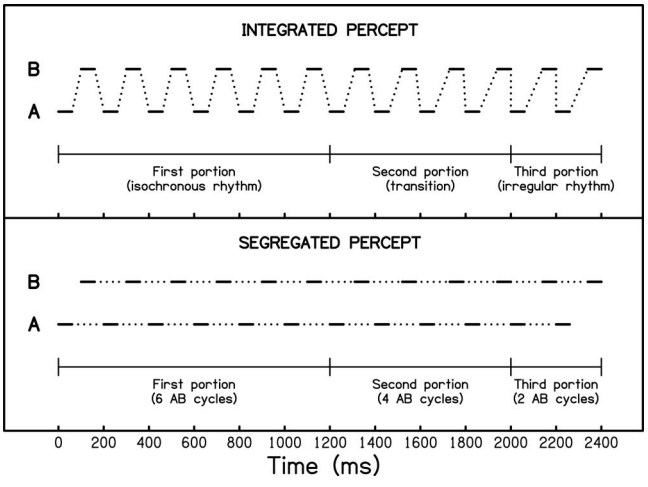


FIG. 1. Stimuli for experiment 1—schematic representation of the longer test sequences used. Each tone is depicted by a short solid line; tones heard as belonging to the same auditory stream are linked by dotted lines. The upper and lower panels represent a sequence heard as integrated (one stream) or segregated (two streams), respectively. Reproduced with permission from Roberts *et al.* (2002), p. 2078. Copyright 2002 by the Acoustical Society of America.

B. Method

1. Stimuli and conditions

There were two conditions, one using longer sequences of alternating tones A and B and the other using short ABA triplets. The longer sequences had an identical configuration to those used by Roberts *et al.* (2002) and by Hong and Turner (2006); each sequence lasted 2.4 s and consisted of 12 AB cycles of alternating pure tones. These sequences were long enough to allow considerable build-up in the tendency for stream segregation in NH listeners (see, e.g., Roberts *et al.*, 2008). The duration of each tone was 60 ms, including 10-ms onset and offset ramps, and the standard inter-tone interval was 40 ms, corresponding to an onset-to-onset time of 100 ms between consecutive tones. The ABA triplets were constructed in the same way; each triplet lasted 300 ms. The frequency characteristics of the sequences were identical in the two conditions. Table I illustrates the relationship be-

TABLE II. Demographic and other details for implant listeners in experiments 1 and 2.

Listener	Age	Gender	Expt.	Open-set speech recognition score (% correct for BKB sentences in quiet)	Type of electrode array	Pulse rate in normal daily use (pulses/s)	Pulse width in normal daily use (μs)
L1	42	F	2	80	Perimodiolar	1200	25
L2	48	F	1	76	Straight	250	25
L3	61	F	2	91	Straight	1200	25
L5	58	M	1	48	Straight	250	25
L6	25	M	1, 2	98	Straight	250	25
L9	46	M	1	99	Straight	1800	25
L11	45	M	1, 2	92	Straight	900	25
L12	69	F	1	93	Perimodiolar	1200	25
L13	51	F	2	100	Straight	1200	25
L14	25	F	2	90	Straight	250	25

tween the frequencies of the tones used and the electrodes stimulated. Tone A was always at 1938 Hz and hence stimulated electrode 11 (e11). Tone B was at a frequency set to stimulate either one of six electrodes more basal than e11 (e1, e3, e5, e7, e9, or e10), or one of seven electrodes more apical than e11 (e12, e13, e14, e15, e17, e19, or e21), or e11 itself. As the task was time consuming, even-numbered electrodes spatially distant from e11 were omitted, while every electrode was included for the region of most interest (centered on e11). The electrode used for tone B was selected quasi-randomly from the test set between trials but was constant within each sequence and did not vary within a trial.

In the longer-sequence condition, each tone B was presented at the exact midpoint in time between the preceding and following tones A for the standard (isochronous) sequences. The structure of the test (anisochronous) sequences is illustrated in Fig. 1. These sequences were configured in the same way as the standard case for the first six AB cycles, but tone B was delayed progressively in equal steps over the next four cycles, and the cumulated delay was maintained for the final two cycles. Hence, the rhythm was regular for the first half of a test sequence but changed to irregular in the second half. Cumulated delays of 10, 20, 30, or 40 ms were used in quasi-random order for the test sequences across trials; 40 ms was the maximum delay possible without temporal overlap between tones A and B. In the triplet condition, the test sequences were created by applying a delay of 10, 20, 30, or 40 ms to the middle tone B of the standard case. Thus, these test triplets were identical to the first three tones comprising the final two cycles in their longer counterparts. For all listeners, a constant pulse rate of 900 pps and a pulse width of 25 μs were used; stimulation was monopolar (MP1+2), and pulses were interleaved across stimulated electrodes.

2. Listeners

Listeners were six adult, post-lingually deafened, experienced users of the Nucleus CI24 implant system; all had normal electrode insertions. Table II shows their demographic and other details.

3. Procedure

A two-interval, two-alternative, forced-choice procedure was used. In both conditions, all possible combinations of electrode pairings and onset delay for tone B were used in quasi-random order. Thus, there were four possible delays (10, 20, 30, or 40 ms) $\times$ 14 electrode pairings = 56 combinations. Stimulus presentation was initiated by the listener using a key press, and there was a 1-s silence between each interval. In each trial block, the test sequence occurred at random, once in the first interval and once in the second, for each combination of delay and electrode pair, giving a total of 112 presentations per block. A total of ten trial blocks was completed, giving 20 presentations of each combination.

All stimuli were delivered via a computer-controlled sound card (16-bit resolution, 20-kHz sampling frequency) to the external input socket of the speech processor via an electrically isolated adaptor cable supplied by the manufacturer for use with the Nucleus implant system. The method of constant stimuli was used instead of an adaptive procedure, as it was considered important that the CI listeners regularly experienced trials in which they were able to perform reasonably well. Pilot work had shown that when the task was consistently difficult, with frequent incorrect responses, listeners found it hard to maintain concentration and remain "on task."

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing. Each listener was seated in front of a computer screen and keyboard. They were instructed to listen to both sequences of tones and to respond by pressing key 1 or 2 on the keyboard to indicate which sequence, first or second, was irregular in rhythm (i.e., to identify the anisochronous interval). Feedback was provided on the computer screen as to whether the response was correct or incorrect. All listeners received training and practice on both sequence lengths prior to the experimental runs. Stimulus pairs expected to be most easily discriminable were used, i.e., when tone B had the same frequency as tone A and the delay on tone B was either 0 or 40 ms. Pilot work indicated that all CI listeners performed well above chance in this case. Training was continued to ensure that each listener understood the task and what to listen for, responded confidently and appropriately, and

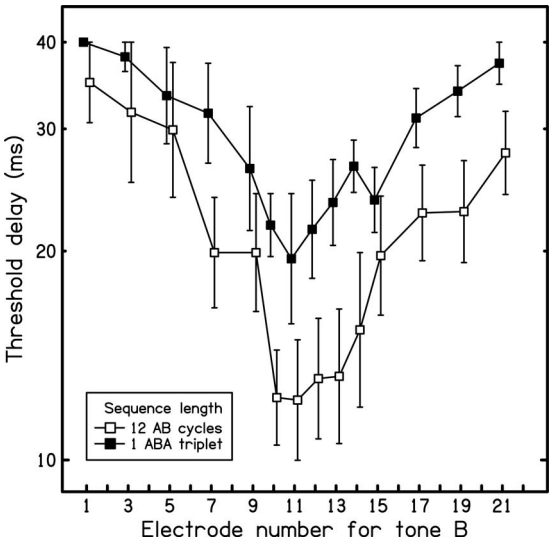


FIG. 2. Results for experiment 1—geometric mean threshold delays for six implant listeners in the triplets and the longer-sequence conditions (filled and open squares, respectively). For the calculation of means, thresholds above the maximum of 40 ms or below the minimum of 10 ms have been allocated values of 40 and 10 ms, respectively. Inter-subject standard errors for each mean are shown by vertical lines. Means and error bars for the two sequence lengths have been slightly displaced for clarity.

achieved good or near-perfect performance. Care was taken to ensure that any task learning was essentially complete before the experimental runs began.

C. Results

For each listener, the percentage of correct responses (out of 20) was calculated for each combination of delay and electrode separation. For each electrode separation, a threshold delay was derived from these scores using a logistic function to fit the data and to estimate the delay equivalent to an accuracy of 75% correct. This approach provided a good fit to the data in over 90% of cases. In cases of floor or ceiling effects on performance, thresholds of 40 or 10 ms were assumed, as appropriate.

The geometric mean thresholds for the six implant listeners are shown in Fig. 2. Overall, a strong effect of electrode number for tone B on threshold can be seen, with higher thresholds associated with greater electrode separation from e11, and this effect is broadly symmetrical. There is also clear evidence of an overall tendency toward higher thresholds for the triplets than for the longer sequences for all electrode positions. A within-subjects analysis of variance (ANOVA) performed using log-transformed threshold estimates revealed a highly significant main effect of electrode number for tone B [$F(13,65)=15.056$, $p<0.001$]. There was also a significant main effect of sequence length [$F(1,5)=7.652$, $p=0.040$], but there was not a significant interaction between electrode number and sequence length [$F(13,65)=1.104$, $p=0.372$]. To reduce the possibility that ceiling effects had distorted the outcomes of this analysis, the ANOVA was repeated, excluding the data for electrodes 1, 3, 19, and 21, which are toward each end of the electrode array. This analysis confirmed a significant effect of electrode number [$F(9,45)=11.333$, $p<0.001$] and of sequence length [$F(1,5)=9.025$, $p=0.030$]. Again, there was no significant interaction between sequence length and electrode number [$F(9,45)=1.109$, $p=0.376$]

To explore further the relationship between threshold delay and electrode separation (between tones A and B) separately for the basal and apical directions (relative to e11), the ANOVA was repeated for each half of the electrode array. The data for e11 were included in both analyses. In the apical direction, there was a significant main effect of electrode separation [$F(6,30)=20.764$, $p<0.001$] and of sequence length [$F(1,5)=12.233$, $p=0.017$], but no significant interaction between sequence length and electrode separation [$F(6,30)=0.761$, $p=0.606$]. Indeed, the two curves are essentially parallel. In the basal direction, there was again a significant main effect of electrode separation [$F(6,30)=22.782$, $p<0.001$]. Although there was no significant main effect of sequence length [$F(1,5)=3.682$, $p=0.113$], a similar trend was apparent to that observed in the apical direction. Again, there was no significant interaction between sequence length and electrode separation [$F(6,30)=1.306$, $p=0.285$]. The hint of a narrowing distance between the two curves for the greatest electrode separations in the basal direction probably reflects ceiling effects on performance in the triplet condition. Excluding the more extreme electrode separations (e1 and e3 in the basal case; e19 and e21 in the apical case) did not change the outcome of these analyses.

Results for the six individual listeners are shown in Fig. 3. For the purposes of illustration, where threshold delays were greater than the maximum delay used (40 ms), a symbol has been plotted at the 40-ms point on each graph with an upward-pointing arrow. Similarly, where thresholds were less than the shortest delay used (10 ms), a symbol has been inserted at the 10-ms point with a downward-pointing arrow. Although variation is apparent in individual listeners' results, some common patterns are observable. The strong dependence of threshold delay on electrode separation is obvious for both the triplets and the longer sequences, with greater separations of tone B from e11 in both apical (higher electrode numbers) and basal directions (lower electrode numbers) generally associated with higher threshold delays. Also, for most listeners, thresholds for the triplets were generally worse than those for the longer sequences.

There were, however, some marked differences in performance between individuals. For example, L2 showed thresholds above 40 ms for the longer sequences on all the electrode separations across the array, except when tone B stimulated one of three central positions (e10, e11, and e12), and a similar pattern was apparent for the triplets. In contrast, L11 had thresholds of 10 ms or better for the longer sequences on all the electrodes stimulated by tone B from e3 through e14, with only slightly higher thresholds for the most basal and for the four most apical electrodes tested. For the triplets, however, L11 generally showed much higher thresholds with a clear dependence on electrode separation (higher thresholds for greater separations from e11).

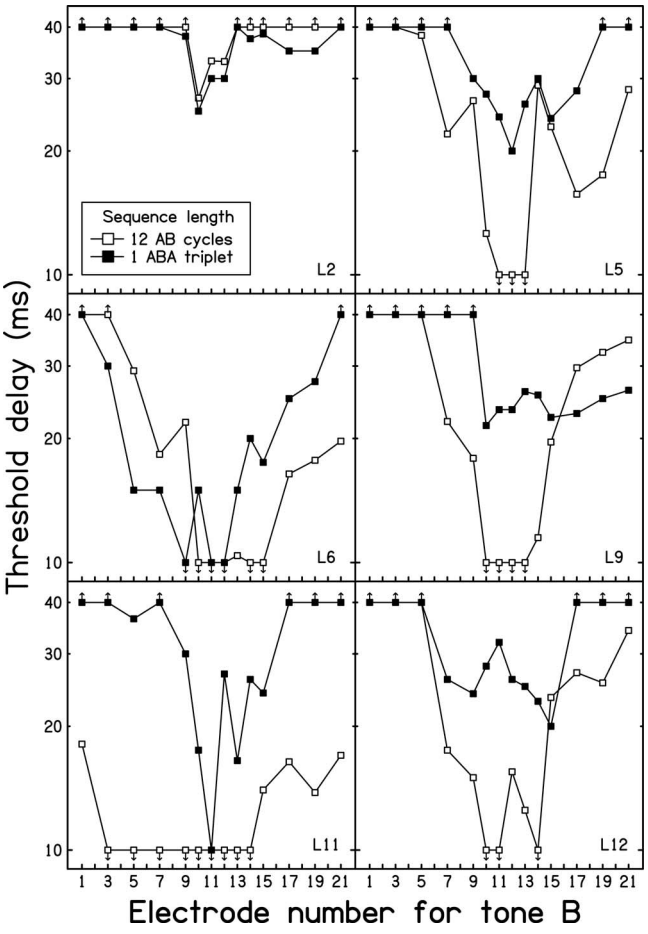


FIG. 3. Results for experiment 1—thresholds for six individual listeners in the triplets and the longer-sequence conditions (filled and open squares, respectively). An upward-pointing arrow on a symbol plotted at 40 ms indicates a threshold delay greater than the maximum tested. A downward-pointing arrow on a symbol plotted at 10 ms indicates a threshold delay lower than the minimum tested.

D. Discussion

The main findings can be summarized as follows. First, increased electrode separation between tones A and B is associated in all listeners with worse performance (higher threshold delays for tone B) for both longer sequences and triplets. Second, for most listeners, overall performance was generally worse for the triplets than for the longer sequences. Third, there was no interaction between sequence length and electrode separation.

The first of these findings is consistent with reports of temporal judgments for short sound sequences of different frequencies in NH listeners (e.g., van Noorden, 1975; Grose *et al.*, 2001) or which stimulate different electrodes in CI listeners (e.g., Hanekom and Shannon, 1998). The results demonstrate that CI listeners found judgments about the relative timing of sounds more difficult when they stimulated electrodes that were more widely separated spatially. The wider separation would have resulted in greater perceptual differences between the sounds, particularly in pitch or timbral brightness. The second finding, that performance was generally worse with the triplets than with the 2.4-s long sequences, tallies with anecdotal reports from our listeners that they found the triplet condition significantly more taxing and is fully consistent with the findings of Hong and Turner (2006). While better performance overall in the triplet condition might have been expected as a result of a lack of build-up of stream segregation, any such benefit was clearly more than offset by the small number of tones available for listeners to detect the relative timing of tones A and B.

The third finding, that there was no interaction between electrode separation and sequence length, is very different from the results reported by Hong and Turner (2006). They measured thresholds when the ratios of the frequencies of tones A and B were 0 (1:1) and 0.5 (1:1.5). Thresholds for the latter case were reported as normalized values after dividing by the corresponding "baseline" performance in the former case; these normalized thresholds were typically much greater for the longer-sequence condition than for the triplet condition. This suggests a much stronger dependence of threshold on electrode separation for the longer sequences, presumably reflecting the contribution of stream segregation. The original thresholds from Hong and Turner's (2006) data for the ratio of 0.5 can be reconstructed from the normalized values and the corresponding thresholds for the ratio 0. For the three individuals that they tested, these values indicate that the difference in thresholds between the triplets and longer sequences was usually much smaller for the ratio of 0.5 than for the ratio of 0, and it was sometimes reversed (i.e., crossovers were observed).

With respect to e11 (centered on 1938 Hz), frequencies giving a ratio of 0.5 in the present experiment would be 2907 Hz (basal direction) and 1292 Hz (apical direction). Table I indicates that these values would be produced by positioning tone B on e8 or e14, respectively. Given that the present experiment included comparisons of thresholds for the two sequence lengths across a much larger range of frequency separations than ratios of 0 and 0.5, one would expect to see crossovers in both the basal and apical directions for most, if not all, individual listeners. Typically, our listeners did not show any crossovers in their curves of threshold delay vs electrode separation (aside from L6 for electrodes 1–9 and L9 for electrodes 13–21). While it is not possible to prove a null hypothesis, it is worth noting that our ANOVA had sufficient power to show significant effects for electrode separation and sequence length. Indeed, as noted earlier, our mean results show broadly parallel curves for the triplets and longer-sequence conditions in both the basal and apical directions. Overall, these results clearly differ from those of Hong and Turner (2006).

Another issue that merits comment is our use of a log time scale to analyze the threshold estimates, following the approach used by Roberts *et al.* (2002, 2008) and by Hong and Turner (2006). If we had used a linear time scale instead, it is possible that a significant interaction between electrode separation and sequence length would have emerged. However, this would have been in the opposite direction to that predicted (i.e., steeper slopes for the triplets than for the longer-sequence condition). It should also be acknowledged that we cannot rule out entirely the possibility that CI listeners may require far longer sequences than do NH listeners (i.e., much greater than 2.4 s) for substantial build-up to occur. Note, however, that a much slower rate of build-up is

likely to reflect a greatly reduced tendency for stream segregation, most probably arising from the reduced effective separation between stimulation at different frequencies for CI listeners compared with NH listeners.

In summary, the results of experiment 1 did not provide convincing evidence that CI listeners commonly experience involuntary stream segregation. If the worse performance associated with greater electrode separations were providing an indirect measure of increased involuntary stream segregation, then a stronger dependence of threshold on electrode separation would have been expected for the 2.4-s long sequences than for the triplets. The latter were too brief to have allowed any appreciable build-up in the tendency for stream segregation to have occurred, and yet slopes for the function relating electrode separation to threshold delay were seen that were similar to those obtained with the longer sequences.

The results for experiment 1 support the conclusions of Cooper and Roberts (2007) rather than those of Hong and Turner (2006). To explore the ability of CI listeners to achieve voluntary stream segregation, for which selective attention is an important factor, experiment 2 used a task requiring the selection of a target pattern from a background of distractors.

III. EXPERIMENT 2

A. Introduction

An interleaved melody task (Dowling, 1973; Hartmann and Johnson, 1991; Cusack and Roberts, 2000) offers a potentially useful approach to investigating stream segregation in CI listeners (with the caveats outlined below), but thus far no published results are available from experiments of this type. The ability of CI listeners to recognize melodies has been evaluated as part of a wider interest in their appreciation of music. For example, Kong *et al.* (2004) used a closed-set melody identification task with six CI and six NH listeners. They pre-selected a set of 12 familiar songs and presented them on a musical synthesizer in two conditions, with or without rhythmic information (the latter case contained only pitch information—the notes were of equal duration with gaps between notes of equal duration). The NH listeners achieved near-perfect performance in melody recognition in both the rhythm and no-rhythm conditions. In contrast, the CI listeners' performance was around chance in the no-rhythm condition and above chance but significantly poorer than that for the NH listeners in the rhythm condition.

The design of any experiment based on interleaved melody recognition in CI listeners must take into account the findings of Kong *et al.* (2004); even good implant users who demonstrate high levels of open-set speech discrimination in quiet are likely to be poor at melody recognition unless rhythmic cues are available. Thus, in the absence of rhythmic cues, target melodies used in such tasks with CI listeners should be very simple, comprising few elements. Pitch differences between successive notes should be as salient as possible, bearing in mind the generally poor pitch perception that CI listeners can achieve. Indeed, the percepts conveyed by place of electrical stimulation via an implant may only be defined loosely as pitch; timbral brightness is arguably a better descriptor (Moore and Carlyon, 2005). If discrimination between two melodies is required, the difference between them should be clear and obvious to the participants.

CI listeners can hear pitch differences when different electrodes are stimulated, so they might be expected to be able to select and recognize a simple melody from a background of distracting tones, provided that the distractors sound sufficiently different from the tones contained in the melody. This outcome would provide evidence that CI listeners can use schema-based selection to hear out a subset of acoustic elements from a sequence as a separate stream, even in the absence of primitive stream segregation. Directing interleaved distractor tones to electrodes spatially separated along the array from the target tones should provide pitch (or brightness) differences that can be used to segregate the melody. If CI listeners can use such cues, their melody identification should be better when the distractors stimulate a different and distinct part of the electrode array than when they overlap with the part of the array stimulated by the target tones. Also, performance should be better when a loudness difference cue is additionally available. Note that any improvement seen when distractors are reduced in level cannot be accounted for by changes in energetic masking because distractor and target tones are presented at different times from one another.

In summary, two main predictions were tested in this experiment. First, in a task requiring discrimination between two simple tone sequences (melodies) interleaved with distractor tones, performance should be better when the distractors stimulate a spatially separate part of the electrode array from the melodies, compared with when the electrode ranges for the melodies and distractors overlap. Second, better performance should be achieved when the loudness of the distractors is reduced relative to that of the melody tones, compared with when they are at the same loudness.

B. Method

This experiment was a simplified version of the interleaved melody identification task used by Cusack and Roberts (2000). Two simple "melodies" were created, each a sequence of five pure tones, that were easy to distinguish from one another and to memorize for most CI listeners. Other details of the experimental set-up, including the configuration of the speech processor, frequency allocations, the mode of stimulation, and pulse characteristics, were identical to those used in experiment 1. The effect on recognition of interleaving the notes of the melody (target tones) with distractor tones was investigated for different target-distractor electrode separations and for different distractor levels, relative to the targets.

1. Stimuli and conditions

Each melody was centered on the middle of the electrode array; e9, e11, and e13 were stimulated. A spacing of two electrodes was used between successive sounds in an effort to provide a clear pitch (or brightness) contrast between them. The sequence of stimulation for melody 1 was

e9, e11, e13, e11, and e9. This should lead to a falling pitch followed by a rise, as higher-numbered electrodes are located toward the apical end of the electrode array and therefore should produce a lower pitch. For melody 2, the order was e13, e11, e9, e11, and e13, thus giving a rising pitch followed by a fall. No rhythm cues were available because all tones were equal in duration and were presented at the same rate.

The target tones comprising the melodies were interleaved with distractor tones, selected from one of five sets corresponding to different electrode ranges. Distractor sets whose ranges of electrode stimulation were spatially remote from the central portion of the array (which encompassed the target tones) were termed "distant," and sets whose ranges were adjacent to but did not overlap with the targets were termed "neighboring." The five sets and associated ranges were as follows: (i) distant-basal (DB), corresponding to e1–e5; (ii) neighboring-basal (NB), corresponding to e4–e8; (iii) overlapping (OV), corresponding to e9–e13 (same range as for the target tones); (iv) neighboring-apical (NA), corresponding to e14–e18; and (v) distant-apical (DA), corresponding to e18–e22. On each trial, six distractors from the appropriate range were selected randomly, in steps of one-electrode spacing, and were interleaved with the five targets such that the whole sequence (11 tones in total) began and ended with a distractor. This ensured that listeners could not perform the task correctly simply by listening for the first or last note.

As well as varying the region along the electrode array that was stimulated by the distractor tones, their level was systematically varied relative to that of the target tones. Without careful loudness-balancing of all the stimuli relative to one another, it is not possible to be sure of their precise relative loudness. Hence, the loudness of the distractors was expressed in terms of percentage of dynamic range. The threshold (T) and maximum comfortable loudness (C) levels for each electrode were measured using standard clinical methods before running the experiment. Stimulus levels are described in terms of stimulus units, which correspond to log current level. The target tones comprising the melody were always presented at C level. The stimulus level for the distractor tones was calculated from the dynamic range. For example, if the measured T and C levels were 100 and 200 stimulus units, respectively, the dynamic range would be 100 stimulus units. In this case, the levels of stimulation on each electrode would be 100% DR (100% of dynamic range) = 200 stimulus units, i.e., C level; 75% DR = 175 stimulus units; 50% DR = 150 stimulus units; and 25% DR = 125 stimulus units.

Individual stimuli comprising both the target melody and distractor sequence were pure tones that were 120-ms long, including 10-ms rise and fall times. The tone onset-to-onset time in an interleaved sequence was 200 ms; i.e., there was a silent interval of 80 ms between consecutive tones. The total duration of an interleaved sequence was 2.2 s. As before, the pure tones used were at frequencies equal to the center frequencies of the channel allocations defined in the experimental speech processor (see Table I). Figure 4 shows representative examples of electrodograms depicting sequences comprising the target tones of a melody interleaved with distractors.

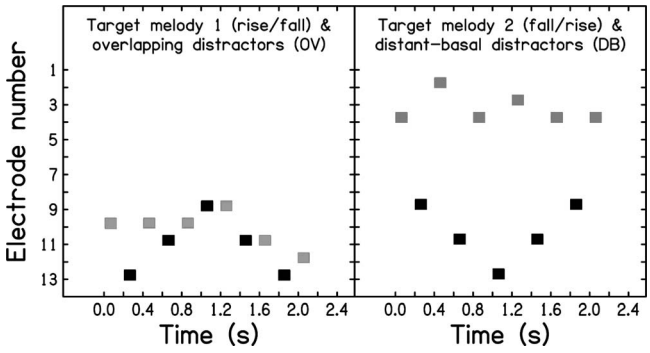


FIG. 4. Stimuli for experiment 2—electrodograms showing examples of interleaved sequences. The left panel shows target melody 1 (rise/fall in pitch) interleaved with distractor tones in the OV condition, i.e., taken at random from within the same electrode range as the target tones (e9–e13). The right panel shows target melody 2 (fall/rise in pitch) interleaved with distractor tones in the DB condition, i.e., taken at random from the range e1 to e5. The filled rectangles illustrate when electrical stimulation was present on each electrode. Black boxes show the target stimulation; gray boxes show the distractors. The ordinate shows electrode number, and the abscissa shows time elapsed in seconds. Only the range of electrode numbers that encompasses those stimulated is illustrated.

2. Listeners

Listeners were six experienced adult CI users of the Nucleus CI24 device, all with normal electrode insertions and pitch ranking. Demographic and other details are shown in Table II. All listeners were screened to ensure that they were able to discriminate reliably between the two target melodies when presented in isolation (i.e., without distractors). A criterion was set of obtaining 100% correct on at least two successive experimental blocks (ten repetitions per block) following practice runs. Several potential participants screened in this way were rejected as they were unable to perform above chance for the isolated melodies, even after training.

3. Procedure

The experiment used a single-interval, two-alternative, forced-choice design. In the main experiment, the stimulus presented on each trial contained one or other of the target melodies (1 or 2), selected at random and interleaved with distractors. Following practice runs without any distractors present, the task was made progressively more difficult by introducing the distractors and increasing their level in each successive experimental block, in the sequence: 25%, 50%, 75%, and finally 100% DR. In the 100%-DR case, the distractors were presented at approximately equal loudness to the target tones. Each block contained 100 presentations in a quasi-random order (5 distractor sets × 2 target melodies × 10 repetitions), giving a total of 400 trials in the main experiment (100 presentations × 4 distractor levels). Every repetition of a block had an equal number of presentations of each melody, and a new randomization was used for the set of stimuli tested.

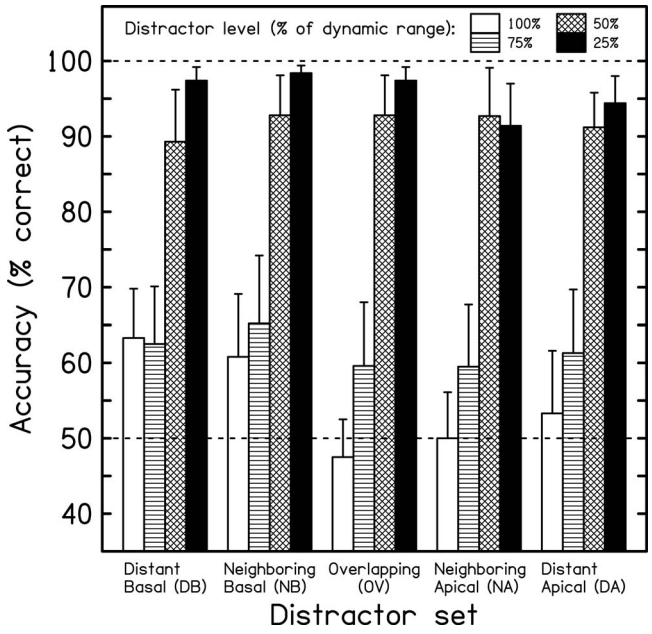


FIG. 5. Results for experiment 2—mean accuracy for six implant listeners. Clustered columns show results for each of the five distractor sets. The results for the four different distractor levels are shown by the differently shaded columns (see inset). Inter-subject standard errors for each mean are shown by vertical lines. The horizontal dashed lines indicate chance (50%) and perfect performance.

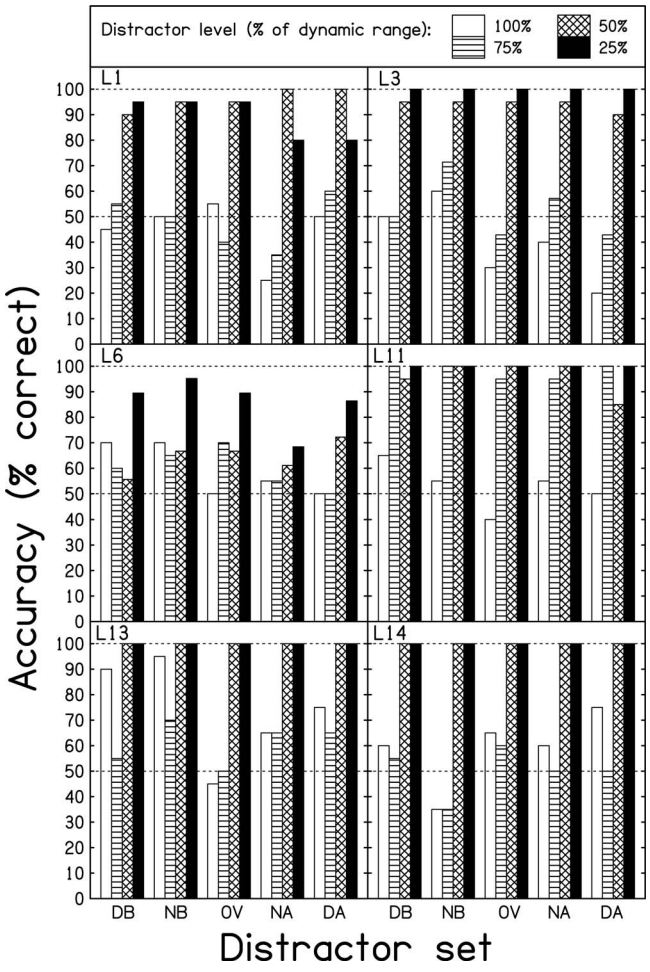


FIG. 6. Results for experiment 2—accuracy scores for six individual listeners. Clustered columns show results for each distractor set; abbreviations for these sets are as indicated in Fig. 5. The results for the four different distractor levels are shown by the differently shaded columns (see inset). The horizontal dashed lines indicate chance (50%) and perfect performance.

The purpose of the experiment was explained to the listeners; they were instructed both verbally and in writing. Listeners first received training in recognition of each melody in the absence of any distractors and then practice on the interleaved melody task prior to the main experimental runs. Care was taken to ensure that all CI listeners understood the task, that they were able to respond appropriately, and that any task learning was essentially complete before the experimental runs began. Each listener was seated in front of a computer screen and keyboard. A visual representation of each melody was provided on screen as a reminder of what to listen for. This illustrated the expected pitch changes for each melody. Listeners were instructed to respond after each trial by pressing either "1" or "2" on the keyboard to indicate which of the two target melodies they had heard. Feedback was provided on screen following each response, displaying "correct" or "incorrect" as appropriate.

C. Results

Responses in each condition were analyzed in terms of percentage correct responses; there was no evidence of bias toward reporting one melody in preference to the other. Mean results for all six CI listeners are shown in Fig. 5. There is a clear and progressive improvement in performance with decreasing distractor level. When averaged across distractor set, mean performance for distractor levels at 100%, 75%, 50%, and 25% DR was 55%, 62%, 92%, and 96% correct, respectively. Note that overall performance was close to chance for the 100%- and 75%-DR conditions but very good for the 50%- and 25%-DR conditions.

Contrary to expectation, which distractor set was used appeared to have little effect on performance. In particular, there was no consistent evidence of poorer performance in the OV condition. A within-subjects ANOVA showed a highly significant effect of distractor level [$F(3,15)=16.56$, $p<0.001$], but there was no effect of distractor set [$F(4,20)=1.29$, $p=0.307$], and there was no interaction between distractor set and level [$F(12,60)=1.15$, $p=0.338$]. As a precaution, to reduce heterogeneity of variance associated with percent correct scores, the ANOVA was repeated using arcsine-transformed data [$Y'=2\arcsin(\sqrt{Y})$; see Keppel and Wickens (2004), p. 155]. The results confirmed the outcome of the original analysis. Pairwise comparisons between the means for different distractor levels were calculated using the restricted least-significant-difference test (Snedecor and Cochran, 1967). Four out of the six possible comparisons showed a significant difference (100% vs 50% DR: $p=0.004$; 100% vs 25% DR: $p=0.001$; 75% vs 50% DR: $p=0.023$; 75% vs 25% DR: $p=0.005$). The difference between the 100%- and 75%-DR cases was not significant ($p=0.472$), and the difference between the 50%- and 25%-DR cases was marginal ($p=0.053$).

Individual results for the six CI listeners are shown in Fig. 6. Overall performance on this task in the 100%-DR condition was generally close to chance, with no systematic

difference between the different distractor sets. The only clear exception was L13; she showed the best scores for the two basally located distractor sets, DB and NB, both well above chance (90% and 95% correct, respectively), and the worst score for the OV set. Surprisingly, she performed better overall in the 100%-DR than in the 75%-DR condition. This may reflect some form of cue learning, as the 100%-DR condition was the last block run. Again, all but one of the CI listeners (in this case, L11) showed performance close to chance in the 75%-DR condition; whereas L11 showed strikingly good performance, with 100% correct scores for three of the five distractor sets and 95% correct for the other two (including the OV set). In general, the best performance was seen for the two lowest levels of distractor, i.e., the 50%- and 25%-DR conditions.

D. Discussion

With some exceptions, individual performance on this task was broadly consistent. Aside from one individual (L13) for some of the distractor sets, CI listeners were not able to identify reliably the target melodies when they were interleaved with distractors of equal loudness (i.e., 100% DR); they responded at or near chance, irrespective of the electrode range occupied by the distractors. When the distractors were attenuated to 75% DR, performance was at (or only slightly above) chance for five out of six listeners. However, further attenuation of the distractors improved performance considerably. At 50% DR, five out of six listeners were able to demonstrate near-perfect performance, and when the distractors were reduced to 25% DR, all listeners showed above-chance (and usually near-perfect) performance. Although they do not provide any details, Chatterjee *et al.* (2006) claimed on the basis of their pilot work that loudness differences can be used to segregate sequences of stimuli.

The fact that the mean results for each distractor set did not differ significantly from each other shows that these CI listeners were not generally able (with the possible exception of L13) to benefit from any pitch (or brightness) differences between the target tones and distractors. In principle, the inclusion of intermediate distractor levels (between 75% and 50%) might have allowed us to distinguish more effectively between the impacts of different distractor sets. However, even if this were true, a much smaller effect of distractor set is implied than was anticipated. In contrast, three NH listeners who completed a comparable task showed a strong dependence of performance on distractor set. These listeners were presented with pure-tone targets and distractors via headphones (Sennheiser HD535) at equal levels (65 dB SPL) in a sound-attenuating room. The frequencies of the tones corresponded to those used for the CI listeners (see Table I), although it should be noted that the pitches evoked will have differed between the NH and CI listeners owing to place mis-matching that arises from the restricted insertion depth characteristic of cochlear implantation. So long as the distractors were drawn from a different frequency range to the target tones, all three NH listeners scored well above chance (two with near-perfect performance), even when the distractors were equal in level to the targets.

Thus, the performance of our CI listeners on this task differs substantially from that obtained with NH listeners; of the two predictions given earlier, only the second has been upheld. Based on these results, it would seem that CI listeners are generally not able to utilize differences in pitch or timbral brightness between target sounds and distractors, resulting from stimulation of different sections of the electrode array, to segregate them and attend to the targets. This outcome appears to be inconsistent with Chatterjee and Galvin's (2002) claim that electrode separation influences the ability of listeners to hear a preview rhythm embedded in a larger sequence, as that implies a form of pattern recognition similar to that needed to identify successfully an interleaved melody. The reason for this discrepancy is unclear in the absence of detailed information about their study, but it may relate to their use of subjective ratings, in contrast with our use of an accuracy measure. Nonetheless, our CI listeners were able to ignore the distractors and successfully attend to the target melodies, provided that the distractors were reduced in level on each electrode to 50% DR or below. This would be equivalent to attending to sounds at the maximum comfortable level while ignoring distracting sounds that are clearly audible but softer. Of course, this finding does not necessarily imply that CI listeners would be able to use the level difference to attend to targets presented at 50% DR when accompanied by distractors presented at C level, as the relative audibility of the targets is also likely to be an important factor affecting performance.

IV. GENERAL DISCUSSION

The results of experiment 1 suggest that our CI listeners generally did not experience involuntary stream segregation when making judgments of the relative timing of tones of different frequency. This task was intended to provide an objective measure of streaming; segregation into two streams should lead to worse performance, despite the efforts of the listeners. The results did show that increased electrode separation between successive tones led to worse performance on judgments about their relative timing, presumably as a result of perceptual differences between them (i.e., differences in pitch or brightness). However, this effect was found for short sequences (tone triplets) to about the same degree as for the longer sequences, and it is therefore not possible to conclude that the effect of electrode separation on temporal discrimination was mediated by stream segregation.

This outcome contrasts with the findings of Hong and Turner (2006). However, it is consistent with those of Cooper and Roberts (2007), who found that judgments by CI listeners of whether a tonal sequence is heard as one stream or two do not conform to the pattern seen in NH listeners. Specifically, the effect of electrode separation on streaming judgments did not depend on sequence rate and did not show evidence of spontaneous flipping between alternative interpretations (Anstis and Saida, 1985). Chatterjee *et al.* (2006) did not test the effect of sequence rate in their study of auditory streaming in CI listeners. Cooper and Roberts (2007) suggested that the perceptual space within which successive sounds can differ from one another is much more limited for

CI listeners than for NH listeners and that this in turn impairs the ability of CI listeners to achieve stream segregation (Moore and Gockel, 2002). Cooper and Roberts (2007) concluded that the judgments of their CI listeners were more akin to a measure of channel discrimination (i.e., a measure of how many pitches were heard) than to a genuine measure of stream segregation.

When there are no effective cues for involuntary stream segregation, CI listeners might, in principle, rely instead on schema-based selection to hear out a subset of acoustic elements from a sequence as a separate stream. Such a reliance on effortful attentional mechanisms would inevitably limit the ability of CI listeners to cope with complex listening environments, particularly given the fairly impoverished nature of the sensory information they receive from their implants. It is, therefore, noteworthy that stimulation of electrodes in differing portions of the electrode array generally did not enable CI listeners in experiment 2 to ignore the distractors and attend to the target melody. At the very least, one might have expected the benefits of differences in electrode range and in level to have combined to enable schema-based selection of the target melody from the distractors, but no evidence of such an interaction was found.

The apparent absence of an effect of distractor set on interleaved melody recognition suggests that that the perceptual differences between sensations evoked by stimulating different places along the electrode array may not be great enough even to support schema-based selection. This may seem somewhat surprising, given that our CI listeners were generally aware of pitch differences between stimulation on different electrodes and were able to discriminate with almost perfect accuracy between our two simple melodies, in the absence of any distractors. However, in NH listeners, the frequency separation at the FB is usually about eight times larger than the frequency difference limen (FDL) below 2 kHz (Rose and Moore, 2005), indicating that the ability to discriminate between two sounds is not in itself sufficient to enable stream segregation by selecting voluntarily a subset of tones from a larger sequence. Nonetheless, it must be acknowledged that the relationship between stream segregation and frequency discrimination is a complex one. In particular, a much wider range of FB/FDL ratios is typical of hearing-impaired listeners, and even in NH listeners this ratio declines as the frequency is raised beyond 2 kHz (Rose and Moore, 2005).

In summary, differences in pitch (or timbre) resulting from activation of different electrodes—even when they are widely separated along the array—do not seem to be an effective cue for most CI listeners, either for involuntary or schema-based segregation. Stimulation of different electrodes does produce perceptual differences, but it appears that these are generally insufficient to facilitate the perceptual grouping of sounds. Differences in level can be used to select target tones interleaved with distractors, but even these must be large to obtain above-chance performance in most CI listeners. On the basis of these results, and of related studies of simultaneous grouping in CI listeners (Carlyon *et al.*, 2007; Cooper, 2008), one might speculate that the ability of many CI listeners to perform well in speech recognition tasks has very little to do with auditory grouping. The processing strategies used in implants are optimized for speech recognition. For example, the Nucleus devices used by all our listeners employ the ACE strategy (*n* of *m* channels), which leads to the activation of only a subset of electrodes in the array, specifically those associated with frequency bands containing the largest spectral peaks. This approach is a robust way of representing the critical formant frequencies for successful speech perception except under very adverse signal-to-noise ratios.

ACKNOWLEDGMENTS

This research was supported by the "Hear and Now" Trust Fund (University Hospital Birmingham, U.K.). We are grateful to Brian Moore and to Rob Morse for their comments on an earlier version of this manuscript. Preliminary presentations of this research were given at the following meetings: (a) British Society of Audiology, Short Papers on Experimental Studies of Hearing and Deafness (University College London, September 2004); (b) Conference on Implantable Auditory Prostheses (Asilomar, Pacific Grove, CA, July–August 2005); (c) British Cochlear Implant Group (BCIG) Meeting (Dublin, Ireland, April 2007).

[1]The dependence of gap detection thresholds on channel separation is reduced for most CI listeners by extensive training (van Wieringen and Wouters, 1999), but it is not abolished.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–271.

Bey, C., and McAdams, S. (**2002**). "Schema-based processing in auditory scene analysis," Percept. Psychophys. **64**, 844–854.

Bey, C., and McAdams, S. (**2003**). "Postrecognition of interleaved melodies as an indirect measure of auditory stream formation," J. Exp. Psychol. Hum. Percept. Perform. **29**, 267–279.

Bregman, A. S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Campbell, J. (**1971**). "Primary auditory stream segregation and perception of order in rapid sequences of tones," J. Exp. Psychol. **89**, 244–249.

Carlyon, R. P., Long, C. J., Deeks, J., and McKay, C. M. (**2007**). "Concurrent sound segregation in electric and acoustic hearing," J. Assoc. Res. Otolaryngol. **8**, 119–133.

Chatterjee, M., and Galvin, J. J. (**2002**). "Auditory streaming in cochlear implant listeners," J. Acoust. Soc. Am. **111**, 2429.

Chatterjee, M., Sarampalis, A., and Oba, S. I. (**2006**). "Auditory stream segregation with cochlear implants: A preliminary report," Hear. Res. **222**, 100–107.

Cooper, H. R. (**2008**). "Auditory grouping in cochlear implant listeners," Ph.D. thesis, Aston University, Birmingham, UK.

Cooper, H. R., and Roberts, B. (**2007**). "Auditory stream segregation of tone sequences in cochlear implant listeners," Hear. Res. **225**, 11–24.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Dowling, W. J. (**1973**). "The perception of interleaved melodies," Cognit. Psychol. **5**, 322–337.

Fu, Q.-J., and Nogaki, G. (**2005**). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," J. Assoc. Res. Otolaryngol. **6**, 19–27.

Grose, J. H., Hall, J. W., Buss, E., and Hatch, D. (**2001**). "Gap detection for similar and dissimilar gap markers," J. Acoust. Soc. Am. **109**, 1587–1595.

Hanekom, J. J., and Shannon, R. V. (**1998**). "Gap detection as a measure of electrode interaction in cochlear implants," J. Acoust. Soc. Am. **104**, 2372–2384.

Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channeling," Music Percept. **9**, 155–183.
Hong, R. S., and Turner, C. W. (**2006**). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," J. Acoust. Soc. Am. **120**, 360–374.
Keppel, G., and Wickens, T. D. (**2004**). *Design and Analysis: A Researcher's Handbook*, 4th ed. (Prentice-Hall, Englewood Cliffs, NJ).
Kong, Y.-Y., Cruz, R., Jones, J. A., and Zeng, F.-G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
Moore, B. C. J., and Carlyon, R. P. (**2005**). "Perception of pitch by people with cochlear hearing loss and by cochlear implant users," in *Pitch, Neural Coding and Perception*, Handbook of Auditory Research Vol. **24**, edited by C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper (Springer, Berlin).
Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acust. Acust. **88**, 320–333.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112**, 2074–2085.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2008**). "Effects of the build-up and resetting of auditory stream segregation on temporal discrimination," J. Exp. Psychol. Hum. Percept. Perform. **34**, 992–1006.
Rogers, W. L., and Bregman, A. S. (**1993**). "An experimental evaluation of three theories of auditory stream segregation," Percept. Psychophys. **53**, 179–189.
Rogers, W. L., and Bregman, A. S. (**1998**). "Cumulation of the tendency to segregate auditory streams: Resetting by changes in location and loudness," Percept. Psychophys. **60**, 1216–1227.
Rose, M. M., and Moore, B. C. J. (**2005**). "The relationship between stream segregation and frequency discrimination in normally hearing and hearing-impaired subjects," Hear. Res. **204**, 16–28.
Snedecor, G. W., and Cochran, W. G. **(1967)**, *Statistical Methods*, 6th ed. (Iowa University Press, Ames, IA).
Throckmorton, C. S., and Collins, L. M. (**2002**). "The effect of channel interactions on speech recognition in cochlear implant subjects: Predictions from an acoustic model," J. Acoust. Soc. Am. **112**, 285–296.
van Noorden, L. P. A. S. (**1975**). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology, The Netherlands.
van Wieringen, A., and Wouters, J. (**1999**). "Gap detection in single- and multiple-channel stimuli by LAURA cochlear implantees," J. Acoust. Soc. Am. **106**, 1925–1939.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.
Warren, R. M., Obusek, C. J., Farmer, R. M., and Warren, R. P. (**1969**). "Auditory sequence: Confusion of patterns other than speech and music," Science **164**, 586–587.

Simultaneous measures of electropalatography and intraoral pressure in selected voiceless lingual consonants and consonant sequences of German

Susanne Fuchs[a)]
Center for General Linguistics (ZAS), Schuetzenstrasse 18, 10117 Berlin, Germany

Laura L. Koenig
Haskins Laboratories, New Haven, Connecticut 06511 and Long Island University, Brooklyn, New York 11201

(Received 26 September 2008; revised 23 June 2009; accepted 24 June 2009)

This work assessed relationships among intraoral pressure (IOP), electropalatographic (EPG) measures, and consonant sequence duration, in the following obstruents, clusters, and affricates of German: /t/, /ʃ/, /ʃt/, and /t͡ʃ/. The data showed significant correlations between IOP and percentage of articulatory contact (PC) for all speakers, whereas duration and place of articulation (measured by the EPG center of gravity) contributed less to IOP changes. Speakers differed in the strength of this relationship, possibly reflecting differences in vocal tract morphology or degree of laryngeal abduction. Single-point EPG and IOP measures in fricatives showed consistent correspondences across consonantal contexts, but the relationships for the stops were more complex and reflected positional effects. Temporal compression was observed for both members of the cluster, but only the fricative portion of the affricate. Conversely, coarticulation was observed for both the stop and fricative portion of the affricate, but only for the stop portion of the cluster, possibly reflecting biomechanical constraints. No clear differences were observed in coarticulatory resistance for stops and fricatives. These data contribute to a limited literature on articulatory-aerodynamic relationships in voiceless consonants and consonant sequences, and will provide a baseline for considering longer combinations of obstruents. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3180694]

PACS number(s): 43.70.Aj [DAB] Pages: 1988–2001

I. INTRODUCTION

The general goal of this work is to obtain a better understanding of how intraoral pressure (IOP) varies in obstruents and obstruent sequences in relation to articulatory actions. A fuller description of IOP and articulatory variation in consonant sequences can provide insight into laryngeal and supralaryngeal control mechanisms in the rapidly-changing aerodynamic conditions of running speech, and ultimately inform aerodynamic modeling of obstruents. The specific purpose of this paper is to quantify the relationships between IOP and lingua-palatal contact, obtained via electropalatography (EPG), in a subset of voiceless stops, fricatives, affricates, and clusters of German. The results will serve as a foundation for future studies on longer obstruent sequences.

High values of IOP are a crucial feature of obstruents, particularly voiceless ones. These sounds are typically among the most affected, for example, by velopharyngeal insufficiency or by loss of laryngeal valving in laryngectomy (e.g., Edels, 1983; Rosenfield *et al.*, 1991). A variety of mechanisms can contribute to high levels of IOP: Glottal abduction, the presence of supralaryngeal closure or constriction (assuming a closed velopharyngeal port), and the duration of closures and constrictions. This work seeks to further understand these mechanisms, focusing on the relationship between IOP, supralaryngeal articulation, and duration.

Much past research has investigated acoustic and articulatory characteristics of consonants, clusters, and affricates, but studies of IOP have mostly considered single stop consonants, with less attention to other obstruents and obstruent sequences. Further, experimental sample sizes have often been limited by rather invasive methodological (viz., transnasal) procedures. Clarifying how consonantal aerodynamics reflect oral articulation in a wider variety of phonetic contexts and across speakers is important for several reasons. First, the acoustic features of voiceless consonants require specific aerodynamic conditions to be fulfilled (Howe and McGowan, 2005; Krane, 2005; Mooshammer *et al.*, 2006; Shadle, 1990), especially in the case of fricatives. Further, pressure variations in the vocal tract affect phonatory behavior (Koenig and Lucero, 2008; Müller and Brown, 1980; Stevens, 1990; Westbury 1979, 1983), for fricatives as well as stops. Finally, evidence from both normal and clinical populations indicates that speakers actively control certain aspects of vocal tract aerodynamics, and some articulatory actions serve aerodynamic requirements (Huber *et al.*, 2004; Müller and Brown, 1980; Prosek and House, 1975; Warren *et al.*, 1992; Westbury, 1983). Yet little past work has explicitly compared IOP with simultaneously-collected supralaryngeal data. Characterizing how vocal tract aerodynamic patterns relate to articulatory activity in consonant sequences is particularly relevant for a language like German, which allows long sequences of obstruents, both within and across syllable boundaries. This work focuses specifically on articulatory contact, which affects supraglottal resistance, and, in

[a)]Author to whom correspondence should be addressed. Electronic mail: fuchs@zas.gwz-berlin.de

turn, upper vocal tract pressure patterns. The combination of EPG and IOP used here allows an assessment of the degree to which changes in supraglottal conditions affect IOP in running speech.

The current study analyzed EPG patterns and IOP in utterance-internal, syllable-initial /t/, /ʃ/, /t͡ʃ/, and /ʃt/ for eight speakers of German to address the following questions: (1) Most simply, to what degree does IOP vary as a function of changes in articulatory contact? That is, how sensitive is the IOP signal to supraglottal articulation as measured via EPG? (2) How do affricates and fricative+stop clusters compare to singleton stops and fricatives in aerodynamic and articulatory contact patterns? (3) How do the affricate and cluster compare in their coarticulatory patterns (viz., changes in place of articulation as measured by EPG), and how much does IOP reflect such articulatory variation?

The following sections review the articulation and aerodynamics of voiceless consonants; coarticulatory patterns of obstruents, especially comparing singleton stops and fricatives with their realization in clusters and affricates; and the phonological representations of clusters and affricates.

A. Articulation and aerodynamics of voiceless consonants, affricates, and clusters

In voiceless obstruents, laryngeal actions must be coordinated with supralaryngeal constrictions. These actions and their coordination vary depending on the acoustic and aerodynamic requirements of the sound in question. Direct laryngeal data on single stops and fricatives indicate that peak glottal opening occurs near the time of oral release for aspirated stops (as occur in German), but earlier, during the turbulent noise region, for fricatives (Hoole *et al.*, 2003; Löfqvist, 1992; Ridouane *et al.*, 2006; Yoshioka *et al.*, 1981). Voiceless affricates and tautosyllabic fricative+stop clusters also tend to show peak abduction in the fricative region (Hoole *et al.*, 2003; Kagaya, 1974; Ridouane *et al.*, 2006; Yoshioka *et al.*, 1981). Finally, the degree of glottal abduction appears to be more extensive in fricatives than stops (Hirose *et al.*, 1978; Lindqvist, 1972; Lisker *et al.*, 1969; Löfqvist and Yoshioka, 1984; Ridouane *et al.*, 2006). Greater abduction extents in fricatives and abduction timed to occur within the fricative regions of obstruent sequences both suggest that speakers are implicitly sensitive to the aerodynamic requirements of fricatives, whereby sufficient airflow is needed to generate turbulent noise (e.g., Scully *et al.*, 1992). In other words, these studies provide evidence of articulatory control of aerodynamic conditions.

Much previous work on speech aerodynamics has focused on IOP in single stop consonants differing in voicing status (Lisker, 1970; Malécot, 1966; Miller and Daniloff, 1977; Müller and Brown, 1980; Svirsky *et al.*, 1997; Warren and Hall, 1973; Westbury, 1983). From studies that investigated differences among *voiceless* consonants, one comparison is particularly germane to the current work: that between stops and fricatives. The presence of a complete closure in stop consonants would lead one to expect higher IOP buildup than in fricatives, but greater vocal-fold abduction in fricatives than stops could counteract this effect. Past investigations of voiceless stops and fricatives have usually reported higher values in the stops (Arkebauer *et al.*, 1967, Koenig *et al.*, 1995; Prosek and House, 1975; Subtelny *et al.*, 1966), but there are some exceptions: Malécot (1968) found few differences, and Malécot (1969) found higher pressures in fricatives. Varying results across studies could result from a number of methodological factors, including (a) whether and how data were averaged across speakers and consonants, (b) instrumental methods, (c) speaker characteristics, and/or (d) differences in the speech materials, syllable position, elicitation conditions, speaking style, or speech rate (see Malécot, 1968, 1969). One possible effect of speech rate is that longer closure or constriction durations allow pressure to build to higher levels (at least until it reaches the maximum value of subglottal pressure; see Miller and Daniloff, 1977). The issue of duration is also relevant in considering obstruent sequences. Subtelny *et al.* (1966) found slightly higher peak pressures in affricates than in simple stops. This could be accounted for by the combination of a longer obstruent interval and extensive laryngeal abduction late in this interval, during the fricative region. In this case, it appears that the laryngeal actions affect IOP to a greater degree than the drop in oral resistance associated with the release of the stop into the fricative.

One might also hypothesize that place of articulation would affect IOP: Specifically, a more posterior place of articulation, or smaller back cavity, could contribute to a faster rise of IOP up to the ceiling level of subglottal pressure. Past authors have considered this possibility mainly for stop consonants, again usually focusing on stop voicing (e.g., Ohala, 1983). It is not clear how much such effects may hold for voiceless consonants, however; the rapid increase in IOP associated with laryngeal abduction may simply outweigh supraglottal place effects. In fact, whereas Subtelny *et al.* (1966) observed somewhat higher pressures in /d/ than /b/, for /p/ and /t/ higher pressures were seen in the bilabials. The situation is also considerably more complicated for fricatives than for stops, since the cross-sectional area of the constriction may covary with place of articulation. For example, the larger constriction size of /ʃ/ should work against any differences that may arise as a function of placement or posterior cavity size relative to more anterior sounds. The current work assesses place of articulation [using the EPG center of gravity (COG) index; see Sec. II C 2] as a possible contributor to IOP variation, but past literature does not lead to clear predictions on this point for the sounds considered here.

B. Context effects in stops, fricatives, affricates, and clusters

Acoustic studies have frequently observed a "compression effect," whereby stops and fricatives tend to be shorter in affricates and clusters than in singleton productions (Byrd, 1993; Crystal and House, 1988; Haggard, 1973; Hawkins, 1979; Klatt, 1974, 1976). Reduced segment durations suggest that the components of a cluster or affricate overlap, i.e., coarticulate. An expansive literature exists on the factors that promote or constrain gestural overlap; this review will focus

on work directly relevant to the current study, namely, comparisons between stops and fricatives and between lingual consonants varying in place of articulation.

Several authors have suggested that fricatives, particularly sibilants such as /s/ and /ʃ/, show limited context-related variability (Nguyen *et al.*, 1994; Recasens and Espinosa, 2007; Recasens *et al.*, 1997; Tabain, 2000). This could again reflect aerodynamic requirements: In a study of consonant sequences, Byrd (1996) proposed that speakers restrict coarticulatory overlap of fricatives with stops because such overlap could inhibit the airflow necessary for fricative production. Other authors have emphasized the articulators involved in forming the constriction. Recasens and colleagues (Recasens 1984, 1985; Recasens *et al.*, 1993, 1997) drew on EPG and acoustic data for alveolar, postalveolar/alveopalatal,[1] and palatal sounds in Italian and Catalan to argue that alveolar sounds permit extensive coarticulation as a function of phonetic context, whereas palatal and alveopalatal sounds restrict coarticulation because they place more constraints on tongue body position. A similar prediction for alveolar vs velar places is suggested by Butcher and Weiher's (1976) EPG study of /t/ and /k/ in German, as well as Bladon and Al-Bamerni's (1976) acoustic study of light (apical) vs dark (velar) /l/ in English. The general notion that restrictions on the tongue body may limit coarticulation dates from Öhman (1966), who proposed that the distinctively palatalized consonants of Russian led speakers of that language to constrain vowel-to-vowel coarticulation (measured by F2) as a general production strategy.

In the case of the cluster /ʃt/ and the affricate /t͡ʃ/, these considerations lead to the prediction that the stop should assimilate in place to the fricative more than the other way around. EPG studies of English speakers (Fletcher, 1989; Liker *et al.*, 2007) and a single Hindi speaker (Dixit and Hoffman, 2004) have reported a posterior placement for /t/ in /t͡ʃ/, similar to that for /ʃ/, with minimal coarticulatory influence on the fricative portion of the affricate. The study of Recasens and Espinosa (2007) on Valencian and Majorcan dialects of Catalan found that, on average, stop regions in /t͡ʃ d͡ʒ/ had less anterior contact and more palatal contact than in /t͡s d͡z/, but the differences were more extreme in Majorcan, and anteriority differences across the two places did not reach statistical significance in Valencian. This example suggests that there may be language-specific differences in the degree of coarticulation within affricates.

C. Phonological representations of clusters and affricates

A final consideration for affricates and clusters is their linguistic representation, in particular, their status as a sequence of phonemes vs a phonetic sequence associated with a single phonemic unit. Although there are some debates about whether affricates are best treated as complex stops (e.g., specified for the feature "strident") or as a phoneme-internal combination of stop and fricative features (e.g., a "contour" segment), phonologists do agree that affricates hold a single segmental position (Clements, 1999; Jakobson *et al.*, 1951; Lombardi, 1990; Rubach, 1994; Sagey, 1986). Clusters are then differentiated from affricates in that they represent two phonemes in sequence. Thus the present work will treat the distinction between affricates and fricative + stop clusters as the difference between a single phonological unit vs a sequence. For simplicity, the notations /t/ and /ʃ/ will be used hereafter to refer to the stop and fricative regions of the affricate as well as the individual stop and fricative phonemes, produced as singletons or in a cluster.

D. Predictions

In light of the aerodynamic, articulatory, and phonological considerations reviewed above, several specific predictions were identified for the current work. These are grouped into three categories to correspond to the three analysis subsets below (Secs. II D 1–II D 3).

(1) (a) Differences in oral aperture should lead to higher IOP values in stops than fricatives. A mitigating consideration is that, as noted above, laryngeal apertures may be larger for fricatives than stops. (b) Pressure should reach higher levels in longer consonantal sequences, at least so long as IOP has not reached its ceiling level of the subglottal pressure (see Miller and Daniloff, 1977; Subtelny *et al.*, 1966).
(2) (a) Stop and fricative durations should be shorter in affricates and clusters than in singleton consonants. (b) Coarticulatory effects should yield articulatory and aerodynamic differences in the stop and fricative portions of /t͡ʃ/ and /ʃt/ as compared to single stops and fricatives.
(3) In the affricate, articulatory constraints on the tongue body and aeroacoustic requirements for fricatives should yield greater accommodation of /t/ to /ʃ/ than vice versa. The same prediction should also hold true for the cluster, although one might expect less coarticulation in the cluster given its phonological representation as a sequence of two phonemes.

II. METHODS

A. Speakers and speech materials

Eight native speakers of Standard German, ranging in age from 27–42 years, were recorded. Three were females (F1–F3) and five were males (M1–M5). All of them had previously participated in EPG experiments, so they were accustomed to speaking with the artificial palates. Speakers also wore the palates for about 15 min before the recording sessions to allow for adaptation.

The full corpus was designed to include a wide variety of lingual obstruents and obstruent sequences of German. As noted above, German is useful in this regard since it allows lengthy obstruent combinations within and across syllable boundaries. Real German words were chosen that included the consonants/clusters/affricates of interest in minimal or near-minimal pairs for both syllable onset and coda positions. These words were then placed into compounds and carrier phrases where the target consonants or consonant strings were adjacent to vowels or one or more consonants so as to vary the length of the consonantal sequence. The resulting utterances were phonotactically legal in German, but

TABLE I. List of utterances.

Word	IPA of target VCV	English gloss
Tasche	ə#t^{h}ɑ	Pocket/bag
Schaf	ə#ʃɑ	Sheep
Stachel	ə#ʃtɑ	Spike
Tschad	ə#t͡ʃɑ	Chad

mostly semantically nonsensical. The present work addresses the small set of target words listed in Table I. Speakers read the utterances from a printed, randomized list in which each target utterance appeared ten times.

The words analyzed here were produced in the sentential frame "*Ich nasche* [ɪç nɑʃə] ___" ("I nibble ___"). The target words listed in Table I all occurred as the first member of a compound ending with *Stelle* ("place"); thus, the full utterance with the first word was "*Ich nasche Taschenstelle*." (A final *-n* is added to *Tasche* in the compound word.) The analysis focuses on the V#CV or V#CCV sequences from the final schwa of *nasche* through the first stressed syllable of the target word. The current data therefore consist of singleton /t/ and /ʃ/, the cluster /ʃt/, and the affricate /t͡ʃ/ in syllable onsets in the vocalic frame /ə#_ɑ/. As indicated by the transcriptions in Table I, the singleton /t/ in German is aspirated in this context, whereas /t/ in the fricative+stop cluster is unaspirated.

B. Instrumentation

Electropalatography (Reading system, EPG 3) was used to obtain articulatory contact information over time and in lateral as well as midsagittal planes. This method can differentiate between oral constrictions and closures and also reveals coarticulatory effects between adjacent stops and fricatives.

To record IOP, an experimental setup was designed whereby a piezoresistive pressure transducer (Endevco 8507C-2) measuring about 2.4 mm in diameter and 12 mm in length was affixed to the posterior end of the EPG palate via a flexible plastic tube (see Fig. 1). The sensor measures the difference between intraoral and atmospheric pressure (the latter obtained via a tube passed around the teeth). This arrangement offers several advantages: It permits simultaneous recording of EPG and IOP, it is not affected by saliva blocking the tube, and it is more comfortable for speakers than inserting a tube or catheter through the nose. As a result, it was possible to carry out recording sessions lasting 2 h or longer, and the recorded speech generally sounded quite natural. The speakers' past experience with EPG recordings also contributed to natural-sounding speech.

Three signals were simultaneously recorded: Acoustics, recorded to DAT at a sampling rate of 48 kHz; EPG, with a sampling rate of 100 Hz; and IOP, recorded into PCQUIRER with a sampling rate of 1859 Hz.[2] The acoustic signal was used to verify that the speaker produced the target utterance accurately and to obtain durational measures. All other analyses were performed in MATLAB (Mathworks, Natick, MA) from the EPG and IOP signals.

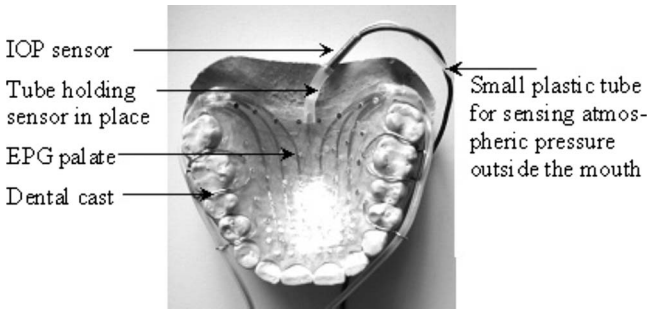


FIG. 1. Equipment setup, showing placement of IOP sensor at posterior end of the EPG palate.

C. Preliminary processing

1. Acoustics

Acoustic durations of stop, fricative, and burst or burst+aspiration regions were made for all consonants and sequences in PRAAT (version 4.4.20; see Boersma and Weenink, 2006) and used for subsequent extraction of EPG data as well as to assess the degree of durational compression in clusters and affricates. Measured examples of each sound or sequence are shown in Fig. 2. The regions were defined as follows: (a) The /t/ closure was measured from the offset of the second formant (F2) in the preceding schwa to the acoustic burst. The /t/ aspiration was measured from the burst onset to the offset of aspiration noise and beginning of voicing. (b) The /ʃ/ was measured from the onset to the offset of frication noise. (c) /ʃt/ was measured from the onset to offset of frication noise, from the offset of frication noise to the end of the stop closure, and then from the onset to the offset of the burst. (d) /t͡ʃ/ was measured from the F2 offset of the preceding vowel to the stop burst for /t/, and from the onset to the offset of frication noise for /ʃ/. As shown in Fig. 2, the burst and the frication noise were measured as a single unit since they are typically acoustically inseparable.

2. EPG

To quantify the EPG contact patterns, two parameters were extracted for each production within the acoustically-defined stop and fricative regions: The percentage of tongue-palate contacts (hereafter PC), out of a possible 62, and the COG, which represents a weighted index in the front-back dimension (Hardcastle *et al.*, 1991). Formally, the indices were defined as follows (where *R* corresponds to row):

$$\mathrm{PC}(\%) = \frac{\text{total number of electrodes contacted} \times 100}{62}, \quad (1)$$

$$\mathrm{COG} = \frac{(0.5 \times R8) + (1.5 \times R7) + (2.5 \times R6) + (3.5 \times R5) + (4.5 \times R4) + (5.5 \times R3) + (6.5 \times R2) + (7.5 \times R1)}{\text{total number of contacts}}. \quad (2)$$

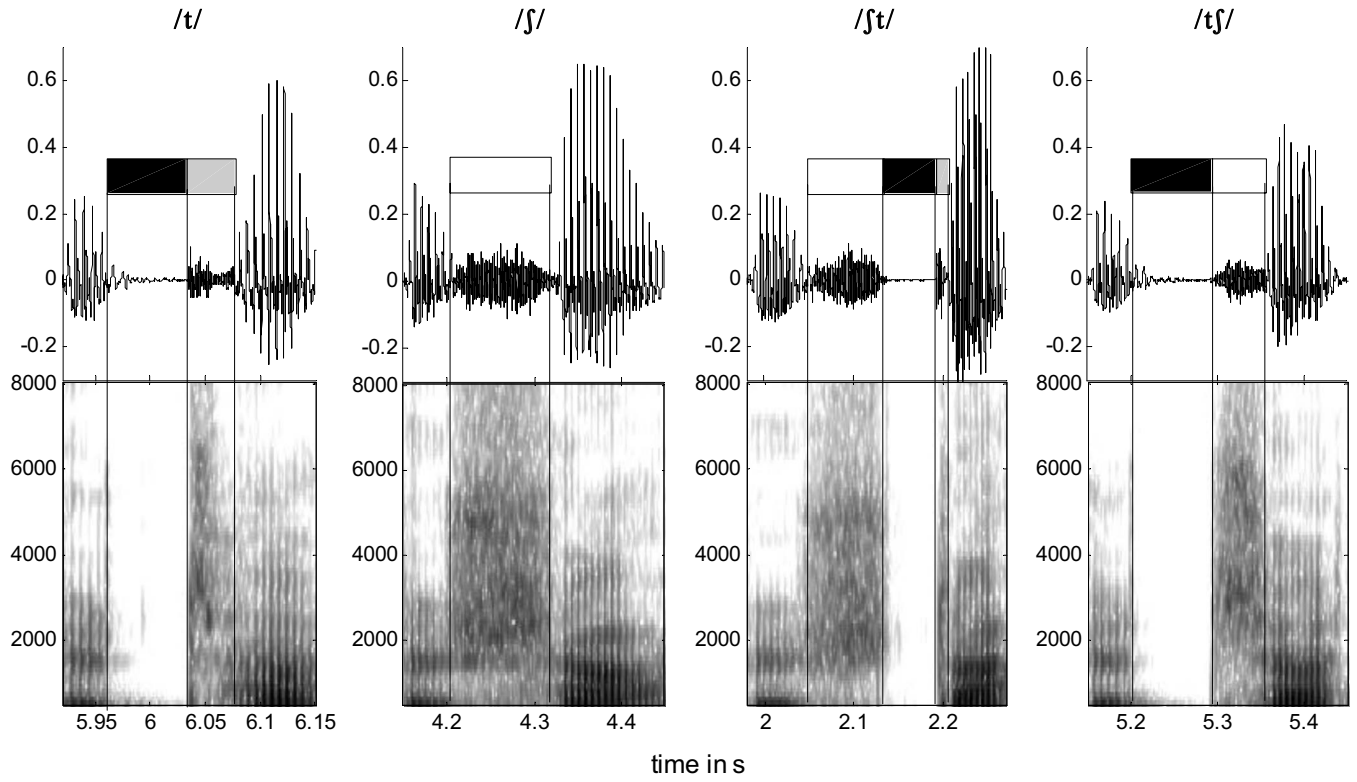


FIG. 2. Measurements of acoustic duration in /t/, /ʃ/, /ʃt/, and /t͡ʃ/ for speaker M5. Top panels: Acoustic waveforms. Bottom panels: Spectrograms. The vertical lines indicate the measured regions, following the color-coded bars shown above the waveform: Black=/t/ closure; white=/ʃ/ frication; gray=burst (in the cluster) or burst+aspiration (for single /t/ in syllable-initial position).

As indicated in Eq. (1), PC values are calculated across the entire EPG palate and are thus independent of the position of the contact (that is, place of articulation). COG, on the other hand, weights each EPG row by a coefficient that increases with anteriority and provides a measure of the place of articulation. Since the front rows are multiplied by greater weights, higher COG values correspond to a more anterior place of articulation.

To obtain a general sense of how the EPG indices varied over time, each index was ensemble averaged over all repetitions (usually 10) of each target word per speaker, and scaled to the average length of the sequence as determined from the acoustics. Specifically, the data were linearly interpolated to a common length of 100 samples; the average and standard deviation (SD) of the repetitions was obtained; and these data were expressed over time as the average duration. These average plots, ±1 SD, were then compared to a plot of the input tokens to verify that the average accurately represented the characteristics of most individual productions. The similarity between the individual tokens and the ensemble averages, as well as the lack of evident nonlinearities in the data, indicated that more complex averaging procedures (e.g., using functional data analysis) were not required. Examples of these average EPG trajectories are included as part of Fig. 3, described in the next section.

3. IOP

The IOP data were smoothed using a kaiser window, with 40 Hz passband and 100 Hz stopband edges, and a damping factor of 50 dB, using the filtfilt function in MATLAB to minimize time delays. These filtering specifications eliminated most of the oscillations associated with phonation, and yielded minimal distortion in regions of rapid pressure change (such as at stop releases). Consonantal regions from the smoothed IOP signal were then extracted to correspond with the consonantal regions as defined in the EPG signals. To correct for baseline drift in the pressure signal over the course of the recording session, a pressure minimum was obtained in the vowels preceding and following each target consonant or consonant string, and the minimum of these two values was subtracted off each extracted token. This effectively set the minimum pressure in each token to

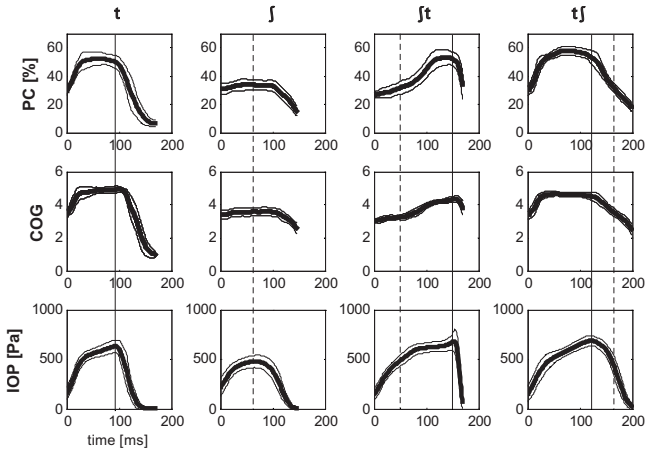


FIG. 3. Average EPG indices (COG=center of gravity, PC=percentage of contact) and IOP traces for all four utterances of one female speaker (F2). The *x*-axis is normalized time based on the average acoustic durations. The solid vertical lines correspond to the defined /t/ time point (either the burst or the peak IOP; see text for more details). The dotted vertical lines correspond to the fricative midpoint.

approximately zero. The IOP values presented here thus reflect the pressure change in the consonant(s) relative to this baseline. Finally, the data were ensemble averaged in a manner analogous to that used for the EPG trajectories.

Figure 3 shows an example of the average (±1 SD) EPG and IOP data for a typical speaker (F2). The lines indicate the measurement points used for statistical analysis (explained below in Sec. II D 2). This figure demonstrates some general characteristics of the data that are relevant for subsequent aspects of the methods as well as for interpreting the results. First, the two EPG measures are highly correlated over time, and the IOP trajectory generally follows the EPG patterns for the single stop, single fricative, and the cluster. Relationships between EPG and IOP are more complex for the affricate. Second, the EPG and IOP data typically do not show distinct stop and fricative regions for either the cluster or the affricate.

D. Analyses

To address the questions identified in Sec. I, three methods of analysis were carried out. Briefly, the first assessed the degree to which IOP changes across the entire consonant or consonant sequence could be predicted from changes in EPG measures and overall duration. The second investigated EPG and IOP differences between singletons vs clusters and affricates, considering the stop and fricative regions separately and assessing the relationships between IOP and EPG changes. The third investigated temporal compression and place coarticulation in the members of the cluster and affricate as compared to singleton stops and fricatives. The following methods sections (II D 1–II D 3) correspond to the three sections in the results (III A–III C).

1. Predicting variation in IOP

To obtain a global picture of how IOP varied as a function of articulatory contact and consonant duration, differences between the maximum and minimum PC, COG, and IOP values (henceforth PCdiff, COGdiff, and IOPdiff) were obtained over the whole consonantal sequence as defined from the acoustics. Together with the acoustic durations, these data were entered into a linear stepwise regression analysis using SPSS (version 15.0) with IOPdiff as the dependent variable and PCdiff, COGdiff, and duration as the predictor variables. This procedure determines how much variance is explained by different numerical models. Data were split by speaker, but not by word, because the rather small number of repetitions (9–10 per word per speaker) would have resulted in unacceptably low statistical power.

2. IOP and EPG patterns across consonantal contexts

To permit statistical comparisons of single /t/ and /ʃ/ with the stop and fricative regions of the clusters and affricates, IOP, PC, and COG values were obtained at single time points for each individual token (shown schematically for the average signals in Fig. 3). The time points were chosen to represent reliable features of stop and fricative production, taking into account the lack of clear fricative and stop regions in the EPG and IOP noted above for /ʃt/ and /t͡ʃ/. For the fricatives, measures were taken at the temporal midpoint of the fricative noise, whether the fricative occurred alone or in a cluster or an affricate. As shown in Fig. 3, the EPG indices were fairly stable at these timepoints. For the stop regions, slightly different criteria were used in the affricate as compared to the cluster and single stop. In the affricate, the IOP value was taken at the time of the acoustic burst. For the closures in /t/ and /ʃt/, the IOP and EPG values were taken at the time of the IOP peak, which consistently occurred immediately prior to the oral release of the stop, i.e., the burst. The peak IOP values rather than the acoustic burst were used in the latter cases because there was usually an abrupt pressure drop after the burst, so that a temporal error in burst location of just a few milliseconds could have yielded greatly reduced pressure values. In the affricate, however, pressure did not peak at the end of the closure, so the burst served as the most reliable indication of the high pressure value associated with the stop.

The single-point EPG and IOP values, along with the acoustic durations (described above in Sec. II C 1) were submitted to repeated-measures analyses of variance (ANOVAs) using the *R* (version 2.7.0) function aov with error terms for speaker and repetitions,[3] following Johnson (2008), and independent variables of consonant environment: The stop region in the singleton vs cluster vs affricate, and the fricative region in the singleton vs cluster vs affricate. Since eight independent ANOVAs were being run (/t/ and /ʃ/ values for duration, IOP, PC, and COG), a rather conservative α-criterion was used for establishing statistical significance: $0.05/8=0.00625$.

3. Clusters and affricates: Compression and coarticulation

The acoustic durations were assessed to determine the degree of compression of /t/ and /ʃ/ portions of the cluster and the affricate relative to the singleton productions. To evaluate the degree of place coarticulation in the stop and the fricative regions of the affricate and cluster, the means of the COG indices for each speaker's single /t/ and /ʃ/ were taken as reference values and set to 100%. The averages for the speaker's cluster and affricate were then expressed relative to this reference value. For example, in speaker F1, the mean COG value for single /ʃ/ was 4.11, whereas her /ʃ/ in /t͡ʃ/ had a mean COG value of 4.24. This yields a ratio of $4.24/4.11=103\%$, meaning that /ʃ/ was about 3% more anterior in the affricate than the single fricative for this speaker.

III. RESULTS

A. IOP variation as a function of changes in articulatory contact and consonant sequence duration

This analysis (see Sec. II D 1) used stepwise regression analysis to assess how well the IOP change over the entire consonantal interval (IOPdiff) could be predicted from changes in the percentage of contact, center of gravity (PCdiff, COGdiff), and consonant sequence duration (Dur). The results of the regression, given in Table II, indicate that PCdiff had the strongest relationship with IOPdiff for all speakers, explaining 37%–82% of the variance. For five

TABLE II. Results of the stepwise linear regression (showing solutions significant at $p<0.001$).

Speaker	Model	R^2	F
F1	PCdiff	0.44	29.0
	PCdiff and Dur	0.51	19.0
F2	PCdiff	0.48	32.8
	PCdiff and COGdiff	0.61	27.4
F3	PCdiff	0.82	163.9
M1	PCdiff	0.58	51.4
M2	PCdiff	0.59	53.4
	PCdiff and COGdiff	0.64	32.3
	PCdiff and COGdiff and Dur	0.74	33.1
	COGdiff and Dur	0.74	50.5
M3	PCdiff	0.37	21.9
	PCdiff and Dur	0.49	17.1
M4	PCdiff	0.49	35.1
	PCdiff and Dur	0.59	26.1
M5	PCdiff	0.62	58.8
	PCdiff and Dur	0.73	48.4
	PCdiff and COGdiff and Dur	0.78	39.9

speakers (F1, M2, M3, M4, and M5), including duration in the model explained an additional 7%-12% of the variance. Changes in place of articulation, as assessed by COGdiff, yielded a slightly better fit than PCdiff alone for speakers F2, M2, and M5. Across speakers, the best-fit solutions explained, on average, 64% of the variance.

The simple linear correlations between IOPdiff and PCdiff (the strongest predictor of IOP according to the stepwise regression) are shown for all speakers in Fig. 4. The relationship was positive (and highly significant) for all speakers, but they differed in the degree of scatter around the regression line, with r-values ranging from 0.609–0.903. Speakers also showed considerable variation in their regression slopes (with values ranging from 2.9–7.0), indicating that the magnitude of IOP change could be quite different across speakers given comparable changes in articulatory contact. Similar slope variation was observed for COGdiff (values 36.9–71.9) and duration (0.8–2.5); that is, slopes could be about two to three times higher in some speakers than in others.

Although the data were not split by consonant for statistical purposes, Fig. 4 permits a qualitative assessment of how well the correlation coefficients reflected within- vs cross-consonant effects. On the whole, the relationships between IOPdiff and PCdiff appear to arise from differences among the four consonant types more than token-to-token variation for an individual target sequence. [Note, for example, the vertical orientation of the repetitions of /ʃ/ (circles) and /ʃt/ (triangles) for speaker F2.] Thus, the positive slopes in Fig. 4 indicate that changes in IOP and lingua-palatal contact were lowest in the single fricative, somewhat higher in the cluster, and highest in the stop and affricate. Inspection of corresponding plots for IOPdiff vs COGdiff and duration suggested that these correlations also mostly reflected cross-context differences.

B. Differences in IOP and EPG across consonantal contexts

As described in Sec. II D 2, single-point /t/ and /ʃ/ regions in the EPG and IOP data were defined in the stop, fricative, cluster, and affricate to permit statistical comparisons across the consonantal contexts. The acoustic durations were also considered in assessing which factors influence IOP. The significance results of the repeated-measures ANOVAs for the four dependent measures (IOP, PC, COG, and Duration) are given in Table III. These correspond to the data shown in Figs. 6–8 and 11.

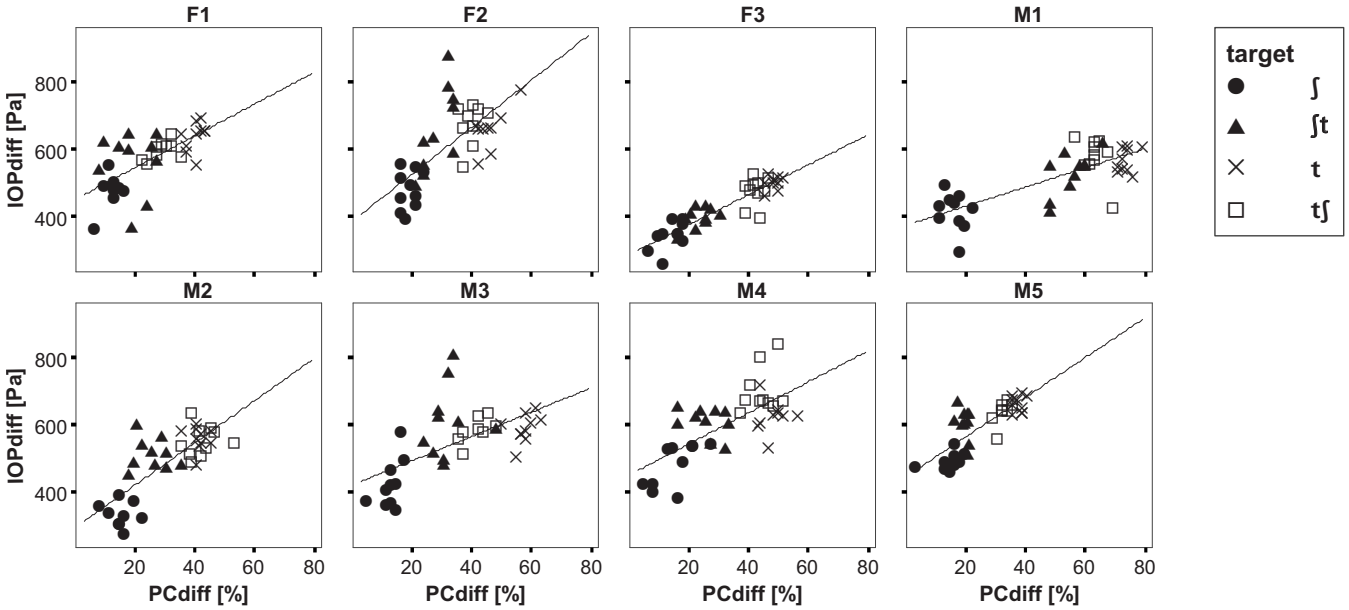


FIG. 4. Correlations for difference in percentage of articulatory contact (PCdiff) through the whole sequence interval against the change in intraoral pressure (IOPdiff) for each of the eight speakers, split by consonant type. Speakers F1–F3 are the females; M1–M5 are the males.

TABLE III. ANOVA results (p-values) for measures of IOP, PC, COG, and Duration (Dur), for /t/ and /ʃ/, alone and in clusters and affricates. Values in bold indicate that the effect of context (singleton, cluster, affricate) was significant at $p<0.00625$. Spkr=Speaker; Repn=Repetition; W/in=Within.

	Overall ANOVA: /t/	Pairwise comparisons			Overall ANOVA: /ʃ/	Pairwise comparisons		
		/t/-/ʃt/	/t/-/t͡ʃ/	/ʃt/-/t͡ʃ/		/ʃ/-/ʃt/	/ʃ/-/t͡ʃ/	/ʃt/-/t͡ʃ/
				IOP				
Spkr	0.210	0.825	0.275	0.114	0.072	0.084	0.031	0.237
Repn	0.598	0.936	0.045	0.420	0.707	0.715	0.484	0.902
W/in	0.007	**0.005**	0.687	0.019	**<0.001**	0.415	**0.001**	**<0.001**
				PC				
Spkr	0.154	0.733	0.148	0.051	0.027	0.293	**0.003**	0.016
Repn	0.518	0.711	0.078	0.124	0.414	0.277	0.935	0.402
W/in	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.042	**<0.001**	**<0.001**
				COG				
Spkr	0.398	0.82	0.176	0.303	0.615	0.879	0.273	0.334
Repn	**0.001**	0.019	0.087	**0.002**	0.979	0.841	0.990	0.108
W/in	**<0.001**	**<0.001**	**<0.001**	0.037	**<0.001**	**<0.001**	**<0.001**	**<0.001**
				Dur				
Spkr	0.634	0.855	0.376	0.450	0.419	0.614	0.209	0.752
Repn	0.010	**0.001**	0.173	**0.001**	**0.003**	0.048	**<0.001**	**0.014**
W/in	**<0.001**	**<0.001**	**0.007**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**

The acoustic durations of closure, frication, and burst or burst+aspiration for the four consonant strings are shown for the eight speakers in Fig. 5. These will be relevant in interpreting the IOP data. Of note is a general tendency for the affricate to be the longest of the four consonant sequences (the difference was statistically significant for speakers F2, F3, M2, M4; speaker M5 is the exception). The relative durations of stop and fricative regions within the affricate differ across speakers, however. For example, speaker F2 has a rather long stop portion in her affricate, whereas speaker M2 has a rather long fricative portion in his.

The average IOP data are presented in Fig. 6. The main effect of context for the fricatives was significant; for the stops the overall ANOVA did not meet significance using the conservative criterion ($\alpha=0.00625$), but one of the *post hoc* tests did (single /t/ vs /t/ in the cluster). The general pattern is that both stops and fricatives have the highest IOP values when they appear as the last member of an obstruent sequence (i.e., when they are durationally later, giving IOP more time to build). Thus, IOP is higher during the fricative of /t͡ʃ/ than in single /ʃ/ or in /ʃt/, and higher during the stop of /ʃt/ than during single /t/. Although one might expect IOP

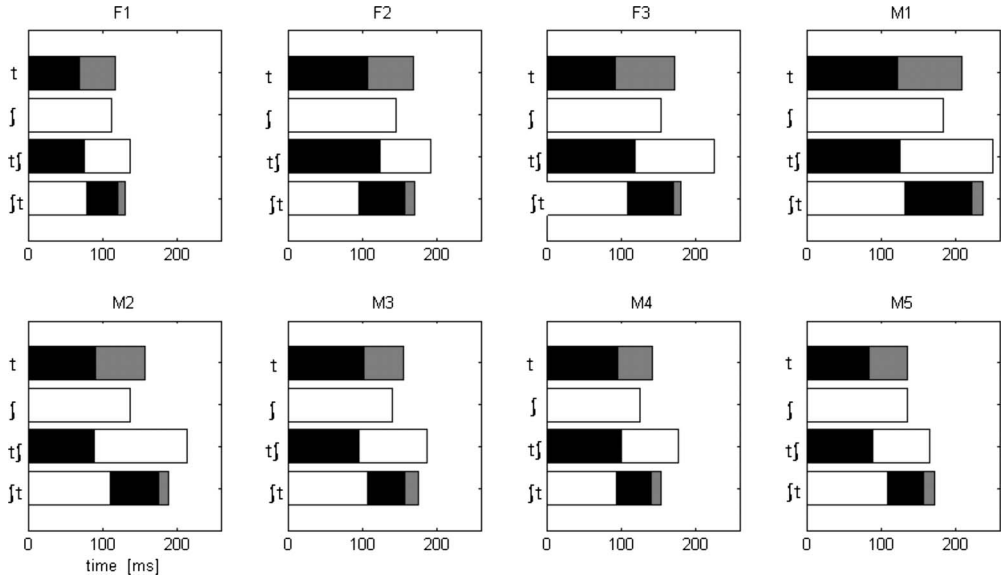


FIG. 5. Acoustic durations for individual speakers, averaged over all repetitions of a sequence. As in Fig. 2, black indicates /t/ closure; white represents /ʃ/; gray indicates burst (for /ʃt/) or burst+aspiration (for /t/). The x-axis is time in ms.

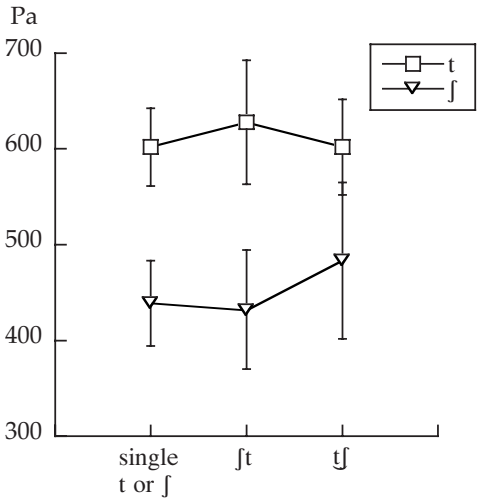


FIG. 6. Average IOP data for single /t/ and /ʃ/ compared with the stop and fricative regions of the cluster and affricate.

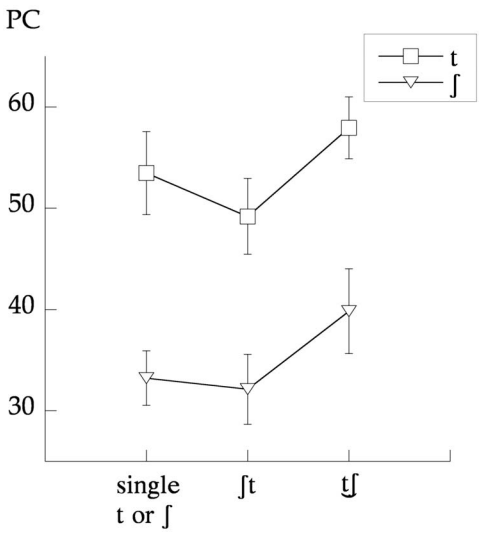


FIG. 7. Average PC data for single /t/ and /ʃ/ compared with the stop and fricative regions of the cluster and affricate.

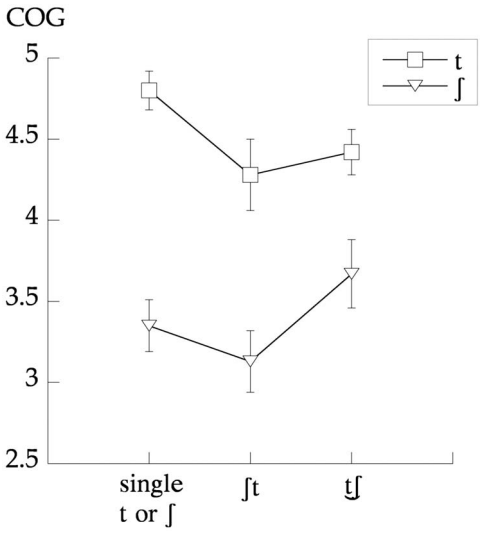


FIG. 8. Average COG data for single /t/ and /ʃ/ compared with the stop and fricative regions of the cluster and affricate.

to reach a ceiling value early in a voiceless stop, the current data show that IOP continues to rise slowly after its rapid initial increase, reaching slightly higher values in /t͡ʃ/ than /t/ (about 7% on average), as seen in Fig. 3 above.

Figures 7 and 8 present the results for the average EPG indices. To assist in interpreting these indices, Fig. 9 shows EPG contact frequencies across all productions of the four targets/target sequences in all speakers. The single time frame shown for each production is the same as that used to obtain the PC and COG indices.

As noted above (see Fig. 3), the percentage of contact and COG measures tend to show similar patterns. The higher PC and COG values for /t/ than /ʃ/ simply indicate that /t/ is articulated with complete closure at a more anterior place of articulation (refer to the EPG contact patterns in Fig. 9). Of more interest is how the stop and fricative regions in the clusters and affricates compare to the single stops and fricatives, and how the contact patterns relate to IOP.

The data for PC (Fig. 7) show that /t/ was articulated with the least contact in the cluster and the most in the affricate. The fricative also had the most contact in the affricate, but there was no significant difference between the fricative portions of the singleton and cluster. The EPG frequency plot (Fig. 9) indicates that the higher value of PC in the affricate mostly reflected greater lateral contact and/or a more anterior place of articulation. For COG (Fig. 8), the ANOVA showed lower values (a retracted place of articulation) for /t/ in the affricate and cluster as compared to single /t/, indicating coarticulation with /ʃ/. Conversely, /ʃ/ in the affricate had higher COG values than single /ʃ/, suggesting coarticulatory influence from /t/. Unexpectedly, /ʃ/ in the cluster was produced most posteriorly; i.e., it did not show more anterior placement under the influence of the upcoming /t/. The EPG frequency plot shows that this effect, although statistically significant, was rather small in magnitude. A more posterior placement in the cluster may reflect biomechanical constraints. Specifically, a more posterior fricative articulation could allow for greater tongue-tip flexibility for producing the apical stop.

A comparison of the EPG and IOP data in Figs. 6–8 reveals some general correspondences among the three measures for the fricatives. The posterior place of articulation (low COG values) for /ʃ/ in the cluster co-occurs with a reduced percentage of contact. As the tongue moves off the back end of the EPG palate, PC typically decreases. Less articulatory contact, in turn, corresponds to lower IOP values. The results for /t/ are more complex. In the cluster, the preceding fricative retracts the /t/ occlusion (lower COG), with corresponding reduction in PC as seen for the fricative. The high IOP value in this context can be attributed to the positional effect noted above (higher pressure in the second member of a sequence). In the affricate, the retracted place of articulation can again be explained as place assimilation to /ʃ/. The high PC value appears to reflect lateral tongue contact in preparation for /ʃ/. Although speakers varied somewhat in their IOP patterns for affricates, the peak IOPs in the affricate were generally as high as, or higher than, those in

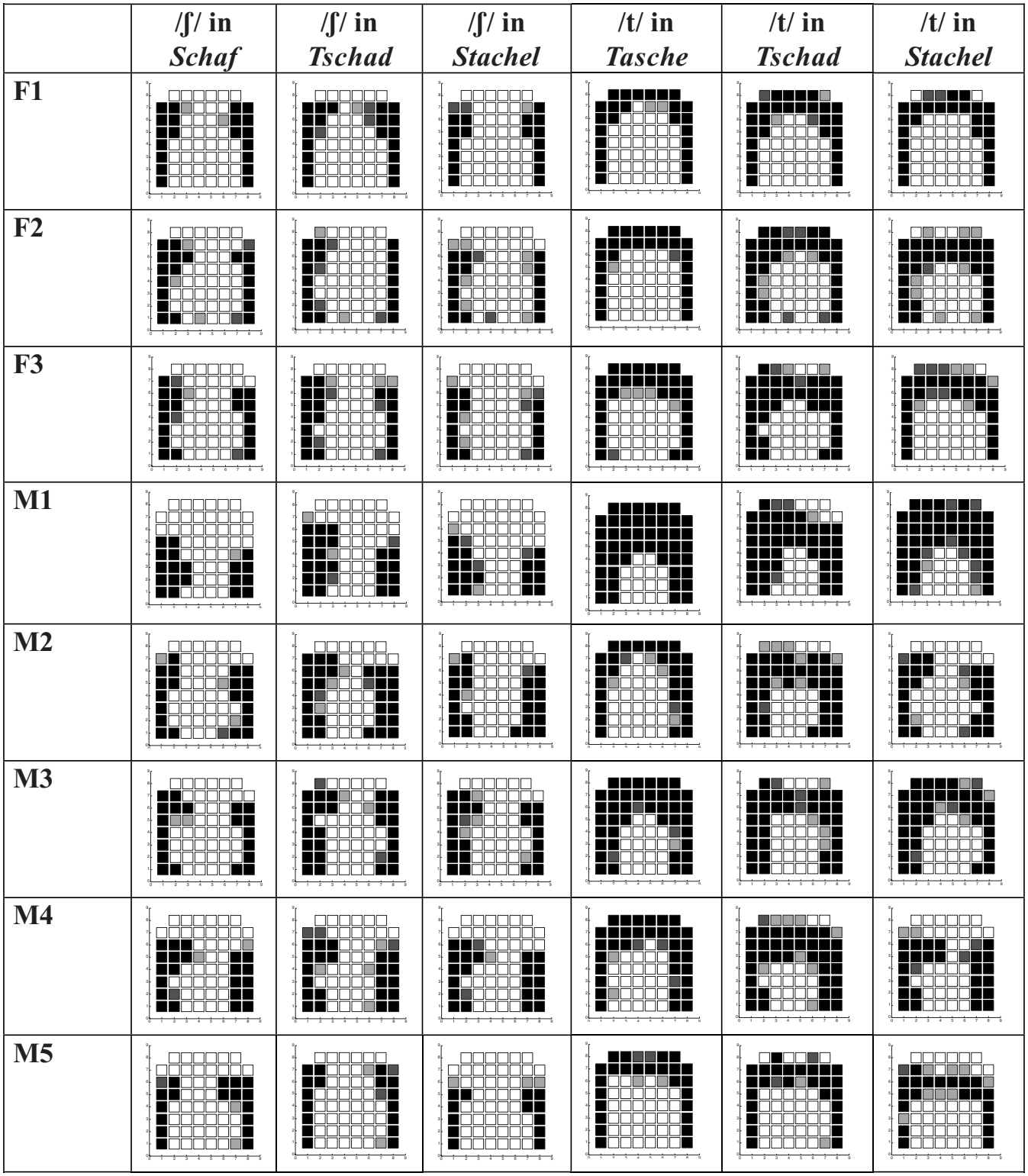


FIG. 9. EPG contact frequency in all repetitions of each sound for all speakers. White squares: Contact occurred in 0%–25% of productions. Light gray squares: Contact occurred in 26%–50% of productions. Dark gray squares: Contact occurred in 51%–75% of productions. Black squares: Contact occurred in 76%–100% of productions.

the single stop, possibly resulting from small supraglottal apertures throughout the course of the affricate combined with a longer duration of IOP buildup.

Some speaker differences were also evident in the relationships among EPG and IOP for these single-point measures. Examples of IOP and PC data from two speakers are shown in Fig. 10. For speaker F2 (left plot), PC was only slightly higher in the stop portion of the affricate than the cluster, whereas speaker M2 (right plot) had a much larger difference. Despite this difference in articulatory contact, both speakers had a similar, slight drop in IOP in the /t/ portion of the affricate compared to the cluster. Further, F2

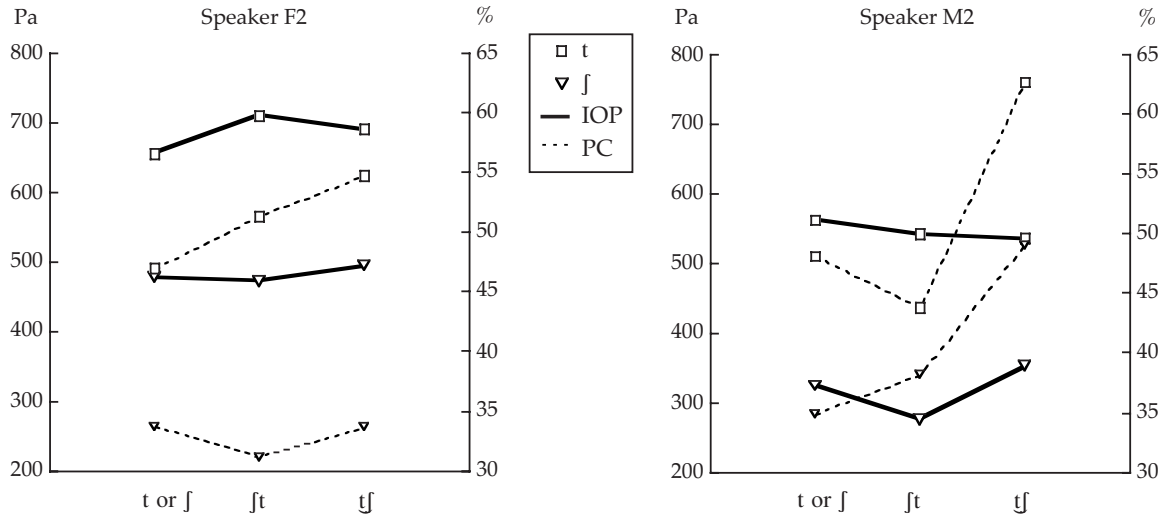


FIG. 10. IOP and PC data for single /t/ and /∫/ compared with the stop and fricative regions of the cluster and affricate, showing differences between two speakers.

had a slight drop in PC for the fricative portion of the cluster compared to single /∫/, with approximately equal IOP in the two contexts; in contrast, M2 had increased PC in the cluster but lower IOP values.

C. Differences between clusters and affricates in compression and coarticulation

Figure 11 shows the average durational data, dividing out the stop and fricative portions of the cluster and affricate for comparison with the single stop and fricative. Of interest here is how the occlusion and frication phases of the affricate and cluster compare with the single consonants, i.e., the extent to which there is a compression effect in the two types of consonantal sequences.

The ANOVA revealed all consonant effects to be significant. In the cluster, both the stop and the fricative regions were shorter than in the single consonants. In the affricate, the closure duration was about the same as for the single stop, whereas the fricative region was shorter than in the singleton. Thus, compression was observed for both segments in the cluster, but only for the fricative portion of the affricate.

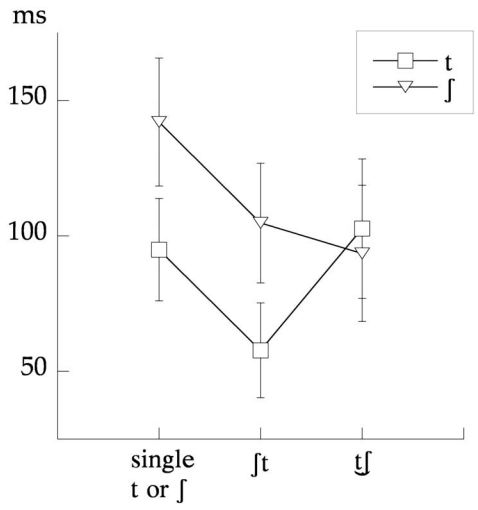


FIG. 11. Average acoustic durations for single /t/ and /∫/ compared with the stop and fricative regions of the cluster and affricate.

To assess the degree to which stop and fricative placement changed in the cluster and affricate compared to their single productions for each speaker, proportional changes in COG were obtained as described in Sec. II D 3. Positive values indicate COG proportions greater than 100%, or more anterior placement in the cluster/affricate than the singleton. Negative values indicate lower COG and more posterior placement. These individual speaker values supplement the average data shown above in Fig. 8, and speak to the question of whether the fricative and stop show similar degrees of coarticulatory influence (or resistance) in these consonant sequences.

The results, presented in Table IV, indicate that /t/ was 8%–11% more posterior when produced in combination with /∫/ than when produced alone. This retraction was a bit more extensive in the cluster than the affricate on average, but three of the eight speakers showed the reverse pattern (F3, M1, and M3). Compared to the single fricative, /∫/ in the affricate was on average 6% more anterior, but 11% more

TABLE IV. Relative amount of coarticulation in articulatory place, measured as a percentage of COG. Negative values indicate more posterior places of articulation in the cluster or affricate as compared to the single stop or fricative.

Speaker	/∫/-/∫t/ diff.	/∫/-/t͡∫/ diff.	/t/-/∫t/ diff.	/t/-/t͡∫/ diff.
F1	−4	3	−4	0
F2	−9	3	−14	−8
F3	−10	6	−3	−8
M1	0	32	3	−9
M2	−6	9	−24	−15
M3	−5	8	−1	−7
M4	−1	9	−28	−10
M5	−15	15	−15	−6
Mean	−6	11	−11	−8

posterior in the cluster. Although /ʃ/ demonstrated less coarticulatory change on average than /t/, the difference was very small, and not all speakers showed this pattern. For example, speaker M1 had much larger coarticulatory changes for the fricative than for the stop portion of /t͡ʃ/. Speaker-specific differences in placement of the single consonants may have contributed to some of this variability. For example, the three speakers who had the most posterior movement of /ʃ/ in the cluster (F2, F3, and M5) were among those who had the most anterior articulation of the single fricative (COG values 3.5–3.6, within the range of 2.19–4.11 for all eight speakers). Thus, these speakers may have retracted their /ʃ/ articulation so as to allow more tongue-tip movement toward the /t/.

IV. DISCUSSION

A. Influences on IOP changes

The first question in this study was the extent to which IOP variation could be predicted from articulatory contact patterns (PC, COG) and constriction duration. Results from the stepwise regression indicated that the EPG indices and overall consonant durations explained, on average, about 64% of the variance. These rather strong relationships were obtained despite the fact that the EPG indices used here provide fairly global information on articulation. For all speakers, changes in PC (i.e., PCdiff) accounted for most of the IOP variation in consonant sequences. Differences in place of articulation, as reflected by the COGdiff values, contributed little to IOP variation. The general pattern for PC and IOP was /ʃ/ < /ʃt/ < /t/, /t͡ʃ/. The differences between the single stop and fricative were in accord with the expectations laid out in the Introduction.

Considerable interspeaker variation was observed in the strength of these relationships, however. One way to quantify this variation is to consider the slopes for the simple correlations. As indicated in Sec. III A, slope values for PC ranged from 2.9 (speaker M1) to 7.0 (speaker F2). Given these slopes and the associated intercept values, a pressure change of 100 Pa corresponded to a PC difference of 35% for M1 vs 14% for F2. In terms of electrode activation, this reflects about 21 electrodes contacted for M1 vs about 9 for F2. There are at least two likely sources for this variation. One is the magnitude and, possibly, timing of glottal area changes for the voiceless consonant targets. More extreme changes in glottal opening could offset less extreme changes in oral constriction areas to yield a similar IOP increase across speakers. Another is individual variation in oral anatomy. Recent EPG studies have obtained high correlations between measures of palatal doming and percentage of articulatory contact (Brunner *et al.*, 2009; Fuchs and Toda, 2008). The combination of EPG and IOP data may ultimately help clarify how individual speakers achieve aerodynamic conditions in obstruents despite morphological variation.

B. Aerodynamics and articulatory contact in stops, fricatives, clusters, and affricates, with particular focus on affricates

The average IOP data showed that both stops and fricatives had the highest values when they were the last member of a consonant sequence, suggesting that duration contributes to the IOP value obtained in a cluster or affricate. Miller and Daniloff (1977), observing poorer correlations between duration and IOP in voiceless stops than in voiced ones, proposed that IOP rapidly reaches a ceiling level in the voiceless stops, reducing such correlations. In the current data, pressure typically rose quickly early in the stop, but it did not always quickly reach a plateau. Further, IOP could increase throughout most of the duration of an affricate for some speakers, and indeed reached a slightly higher peak, on average, in the affricate than in the plain stop. Subtelny *et al.* (1966) also found higher pressures in /t͡ʃ/ than in /t/ for English: 7.1 vs 6.8 cm H_2O, respectively. This corresponds to a difference of about 300 Pa, or 4% over the value for single /t/. This appears to be consistent with the 7% difference found here between /t͡ʃ/ and /t/. An extended plateau of IOP was observed during the affricate for some speakers; other speakers had a steadily rising pattern. Both of these indicate that pressure is not vented rapidly after the release of the occlusion phase. In a modeling study of affricates, Stevens (1993) obtained good fits with acoustic data using a relatively constant constriction area for about 50 ms following stop burst, combined with large glottal areas. Oral airflow data in that study also showed a rather gradual increase following affricate release. Again, the current data appear to be consistent with this past work. Stevens (1993) also noted that rates of constriction release may vary considerably across speakers, which would account for the varying IOP profiles seen across the eight speakers recorded here.

C. Compression and coarticulation: Clusters vs affricates

Several past studies have reported durational data for clusters and affricates. Since the degree of temporal compression in such sequences may vary as a function of the articulatory place of the consonants involved (Borden and Gay, 1979; Haggard, 1973; O'Shaughnessy, 1981), comparisons with the present results will be limited to alveolar stops combined with alveolar or postalveolar (sibilant) fricatives. The shortening of stops and fricatives seen here in clusters is consistent with most past work (Haggard, 1973; Hawkins, 1979; Klatt, 1974, 1976), although Crystal and House (1988) found compression of fricatives but not stops in clusters. Data on compression in the affricate has mostly been limited to measures of the stop portion, and findings here are conflicting. Some authors have observed compression of the occlusion phase (Byrd, 1993; Hoelterhoff and Reetz, 2007), but Liker *et al.* (2007) found, as in the current data, that stop regions did not shorten in affricates.

An assessment of the degree of coarticulatory place change as a proportion of the COG value for the singleton did not reveal clear differences between /t/ and /ʃ/ in the context of an affricate or cluster. Values for the fricative varied less on average than those of the stop, but the difference was very small and inconsistent across speakers. In the affricate, both the stop and the fricative portions showed coarticulatory effects. This contrasts with previous work that has observed retracted placement of /t/ in the affricate, but a

fricative similar in place to single /ʃ/ (Dixit and Hoffman, 2004; Fletcher, 1989; Liker *et al.*, 2007). The data therefore did not support predictions of greater coarticulatory resistance (more stable placement) for /ʃ/ given its postalveolar place of articulation and the necessity of a precisely shaped constriction for the fricative (Recasens, 1984, 1985; Recasens *et al.*, 1993, 1997). Such differences across studies may arise from variation in measurement methods or possibly from cross-linguistic differences in coarticulatory behavior (Recasens and Espinosa, 2007). Another possibility is that some speakers may increase lip rounding to achieve frequencies appropriate for /ʃ/ in the context of a somewhat fronted articulation. That is, speakers may accomplish some aspects of the acoustic characteristics of /ʃ/ using articulators other than the tongue. Finally, recent work suggests that /t/ is subject to some restrictions on coarticulation as well; Mooshammer *et al.* (2006) argued that apical stops had to maintain an anterior place of articulation in order to ensure appropriately high frequencies in the release burst. Thus, a variety of factors may come into play in determining coarticulatory behavior in addition to coarticulatory resistance: speaker-specific differences in anatomy and speech sound production, language-specific characteristics, the extent to which motor equivalence yields multiple possibilities for achieving a sound's characteristic acoustics, and biomechanical constraints on articulatory activities for sequential sound production.

The phonological distinction between affricates and clusters, namely, a difference between one segment and two, could suggest that compression and/or coarticulation should be more extreme in the affricate than the cluster. The current data do not show either of these effects. The affricate was usually durationally longer than the cluster, so temporal compression was not more extensive. It was the case that the affricate showed coarticulatory effects for both the stop and the fricative regions, whereas only the stop portion of the cluster showed a place change relative to the singleton. One might argue that finding coarticulatory processes for both stop and fricative portions of the affricate provides some support for greater phonological unity in the affricate. More likely, however, is that the lack of coarticulation seen for the /ʃ/ portion of the cluster simply reflects biomechanical constraints on articulating an apical stop when the body of the tongue is elevated for a postalveolar fricative.

D. Conclusions

There have essentially been no previous studies comparing IOP and EPG in obstruents; thus this work was somewhat exploratory. The results indicate that much of the variance in overall IOP change in an obstruent sequence can be predicted simply from consideration of changes in articulatory contact, with little influence of articulatory place, at least for the small consonant set investigated here. Both EPG and IOP data showed that stop and fricative regions in clusters and affricates differed from singleton productions in ways that reflected coarticulatory behavior as well as durational effects and biomechanical effects. The data did not support expectations of greater coarticulatory resistance in /ʃ/ vs /t/, or greater cohesion within an affricate than a cluster. These data will provide baseline information for considering more complex consonantal strings in the full corpus, and provide input to future modeling of low-frequency consonantal aerodynamics in running speech.

ACKNOWLEDGMENTS

We express our appreciation to our speakers; to Anke Busler, for performing the acoustic measurements; to Jörg Dreyer, for designing the experimental setup; to Phil Hoole and Jorge Lucero for discussions on data processing and filtering; to Ralf Winkler for assistance with statistics in *R*; and to Marzena Zygis for discussions on the phonological status of the affricate. This work was supported by a grant from the German Research Council (DFG) and the Federal Ministry of Education and Research (BMBF), and a grant from the French-German University (DFH Saarbrücken) to the PILIOS project.

[1]For Italian and Catalan, Recasens and colleagues use the term alveopalatal to characterize /ɲ/ and /ʎ/ as well as /ʃ/. The fricative /ʃ/ and affricate /t͡ʃ/ in German are traditionally characterized as postalveolar (Pompino-Marschall, 2003). The main point in the present context is that /ʃ/, having more posterior articulation, involves the tongue body to a greater extent than /t/.

[2]The PCQUIRER program automatically adjusts sampling rates depending on the frequency range chosen for analysis. What is important for present purposes is that the pressure data were sampled at a rate much higher than the low-frequency variations of interest.

[3]In the aov function, data from all repetitions are included, rather than entering means for each subject. The repetition error term accounts for the repeated measures within speaker. This option is not available in some other statistical packages such as SPSS.

Arkebauer, H. J., Hixon, T. J., and Hardy, J. C. (**1967**). "Peak intraoral air pressures during speech," J. Speech Hear. Res. **10**, 196–208.

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

Boersma, P., and Weenink, D. (**2006**). "Praat: Doing phonetics by computer (Version 4.4.20) [Computer program]," http://www.praat.org/ (Last viewed May, 2006).

Borden, G. J., and Gay, T. (**1979**). "Temporal aspects of articulatory movements for /s/-stop clusters," Phonetica **36**, 21–31.

Brunner, J., Fuchs, S., and Perrier, P. (**2009**). "On the relationship between palate shape and articulatory behavior," J. Acoust. Soc. Am. **125**, 3936–3949.

Butcher, A., and Weiher, E. (**1976**). "An electropalatographic investigation of coarticulation in VCV sequences," J. Phonetics **4**, 59–74.

Byrd, D. (**1993**). "54,000 American stops," UCLA Working Papers in Phonetics **83**, 97–116.

Byrd, D. (**1996**). "Influences on articulatory timing in consonant sequences," J. Phonetics **24**, 209–244.

Clements, G. N. (**1999**). "Affricates as noncontoured stops," in *Proceedings of LP '98: Item Order in Language and Speech*, edited by O. Fujimura, B. D. Joseph, and B. Palek (Karolinum, Prague), pp. 271–299.

Crystal, T. H., and House, A. S. (**1988**). "Segmental durations in connected-speech signals: Current results," J. Acoust. Soc. Am. **83**, 1553–1573.

Dixit, R. P., and Hoffman, P. R. (**2004**). "Articulatory characteristics of fricatives and affricates in Hindi: An electropalatographic study," J. Int. Phonetic Assoc. **34**, 141–159.

Edels, Y. (**1983**). *Laryngectomy: Diagnosis to Rehabilitation* (Croom Helm, London).

Fletcher, S. G. (**1989**). "Palatometric specification of stop, affricate, and sibilant sounds," J. Speech Hear. Res. **32**, 736–748.

Fuchs, S., and Toda, M. (**2008**). "Inter-speaker variability and the articulatory-acoustic relations in German and English /ʃ/," J. Acoust. Soc. Am. **123**, 3079.

Haggard, M. (**1973**). "Abbreviation of consonants in English pre- and post-

vocalic clusters," J. Phonetics **1**, 9–24.
Hardcastle, W. J., Gibbon, F., and Nicolaidis, K. (**1991**). "EPG data reduction methods and their implications for studies of lingual coarticulation," J. Phonetics **19**, 251–266.
Hawkins, S. (**1979**). "Temporal co-ordination of consonants in the speech of children: Further data," J. Phonetics **7**, 235–267.
Hirose, H., Yoshioka, H., and Niimi, S. (**1978**). "A cross language study of laryngeal adjustment in consonant production," Ann. Bull., Res. Inst. Logopedics and Phoniatrics, Faculty of Medicine, Univ. Tokyo **12**, 61–71.
Hoelterhoff, J., and Reetz, H. (**2007**). "Acoustic cues discriminating German obstruents in place and manner of articulation," J. Acoust. Soc. Am. **121**, 1142–1156.
Hoole, P., Fuchs, S., and Dahlmeier, K. (**2003**). "Interarticulator timing in initial consonant clusters," in *Proceedings of the Sixth International Seminar on Speech Production*, edited by S. Palethorpe and M. Tabain (Macquarie University, Sydney), pp. 101–106.
Howe, M. S., and McGowan, R. S. (**2005**). "Aeroacoustics of [s]," Proc. R. Soc. London, Ser. A **461**, 1005–1028.
Huber, J., Stathopolous, E. T., and Sussman, J. E. (**2004**). "The control of aerodynamics, acoustics, and perceptual characteristics during speech production," J. Acoust. Soc. Am. **116**, 2345–2353.
Jakobson, R., Fant, G., and Halle, M. (**1951**). *Preliminaries to Speech Analysis* (MIT, Cambridge).
Johnson, K. (**2008**). *Quantitative Methods in Linguistics* (Wiley, Oxford).
Kagaya, R. (**1974**). "A fiberscopic and acoustic study of the Korean stops, affricates and fricatives," J. Phonetics **2**, 161–180.
Klatt, D. H. (**1974**). "The duration of [s] in English words," J. Speech Hear. Res. **17**, 51–63.
Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.
Koenig, L. L., and Lucero, J. C. (**2008**). "Stop consonant voicing and intraoral pressure contours in women and children," J. Acoust. Soc. Am. **123**, 1077–1088.
Koenig, L. L., Löfqvist, A., Gracco, V. L., and McGowan, R. S. (**1995**). "Articulatory activity and aerodynamic variation during voiceless consonant production," J. Acoust. Soc. Am. **97**, 3401.
Krane, M. H. (**2005**). "Aeroacoustic production of low-frequency unvoiced speech sounds," J. Acoust. Soc. Am. **118**, 410–427.
Liker, M., Gibbon, F. E., Wrench, A., and Horga, D. (**2007**). "Articulatory characteristics of the occlusion phase of /tʃ/ compared to /t/ in adult speech," Adv. Speech-Lang. Path. **9**, 101–108.
Lindqvist, J. (**1972**). "Laryngeal articulation studied on Swedish subjects," Speech, Music and Hearing Quart. Prog. and Status Rept., Royal Inst. Tech. (KTH), Stockholm, Sweden **2–3**, 10–27.
Lisker, L. (**1970**). "Supraglottal air pressure in the production of English stops," Lang Speech **13**, 215–230.
Lisker, L., Abramson, A. S., Cooper, F. S., and Schvey, M. H. (**1969**). "Transillumination of the larynx in running speech," J. Acoust. Soc. Am. **45**, 1544–1546.
Löfqvist, A. (**1992**). "Acoustic and aerodynamic effects of interarticulator timing in voiceless consonants," Lang Speech **35**, 15–28.
Löfqvist, A., and Yoshioka, H. (**1984**). "Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production," Speech Commun. **3**, 279–289.
Lombardi, L. (**1990**). "The nonlinear organization of the affricate," Nat. Lang. and Ling. Theory **8**, 375–425.
Malécot, A. (**1966**). "The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates," Phonetica **14**, 65–81.
Malécot, A. (**1968**). "The force of articulation of American stops and fricatives as a function of position," Phonetica **18**, 95–102.
Malécot, A. (**1969**). "The effect of syllabic rate and loudness on the force of articulation of American stops and fricatives," Phonetica **19**, 205–216.
Miller, C. J., and Daniloff, R. (**1977**). "Aerodynamics of stops in continuous speech," J. Phonetics **5**, 351–360.
Mooshammer, C., Hoole, P., and Geumann, A. (**2006**). "Interarticulator cohesion within coronal consonant production," J. Acoust. Soc. Am. **120**, 1028–1039.
Müller, E. M., and Brown, W. S. (**1980**). "Variations in the supraglottal air pressure waveform and their articulatory interpretation," in *Speech and Language Advances in Basic Research and Practice*, edited by N. Lass (Academic, Madison), Vol. **4**, pp. 318–389.
Nguyen, N., Hoole, P., and Marchal, A. (**1994**). "Regenerating the spectral shapes of [s] and [ʃ] from a limited set of articulatory parameters," J. Acoust. Soc. Am. **96**, 33–39.
Ohala, J. J. (**1983**). "The origin of sound patterns in vocal tract constraints," in *The Production of Speech*, edited by P. F. MacNeilage (Springer-Verlag, New York), pp. 189–216.
Öhman, S. E. G. (**1966**). "Coarticulation in VCV utterances: Spectrographic measurements," J. Acoust. Soc. Am. **39**, 151–168.
O'Shaughnessy, D. A. (**1981**). "A study of French vowel and consonant durations," J. Phonetics **9**, 385–406.
Pompino-Marschall, B. (**2003**). *Einführung in die Phonetik*, 2nd ed. (De-Gruyter, Berlin).
Prosek, R. A., and House, A. S. (**1975**). "Intraoral air pressure as a feedback cue in consonant production," J. Speech Hear. Res. **18**, 133–147.
Recasens, D. (**1984**). "Vowel-to-vowel coarticulation in Catalan VCV sequences," J. Acoust. Soc. Am. **76**, 1624–1635.
Recasens, D. (**1985**). "Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences," Lang Speech **28**, 97–114.
Recasens, D., and Espinosa, A. (**2007**). "An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects," J. Int. Phonetic Assoc. **37**, 143–172.
Recasens, D., Pallarès, M. D., and Fontdevila, J. (**1997**). "A model of lingual coarticulation based on articulatory constraints," J. Acoust. Soc. Am. **102**, 544–561.
Recasens, D., Farnetani, E., Fontdevila, J., and Pallarès, M. D. (**1993**). "An electropalatographic study of alveolar and palatal consonants in Catalan and Italian," Lang Speech **36**, 213–234.
Ridouane, R., Fuchs, S., and Hoole, P. (**2006**). "Laryngeal adjustments in the production of voiceless obstruent clusters in Berber," in *Speech Production: Models, Phonetic Processes, and Techniques*, edited by J. Harrington and M. Tabain (Psychology, New York), pp. 275–297.
Rosenfield, D. B., Viswanath, N., Herbrich, K. E., and Nudelman, H. B. (**1991**). "Evaluation of the speech motor control system in amyotrophic lateral sclerosis," J. Voice **5**, 224–230.
Rubach, J. (**1994**). "Affricates as strident stops in Polish," Ling. Inq. **25**, 119–143.
Sagey, E. (**1986**). "The representation of features and relations in non-linear phonology," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge.
Scully, C., Castelli, E., Brearley, E., and Shirt, M. (**1992**). "Analysis and simulation of a speaker's aerodynamic and acoustic patterns for fricatives," J. Phonetics **20**, 39–51.
Shadle, C. (**1990**). "Articulatory-acoustic relationships in fricative consonants," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 187–209.
Stevens, K. N. (**1990**). "Vocal-fold vibration for obstruent consonants," in *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, edited by Jan Gauffin and Britta Hammarberg (Singular, San Diego), pp. 29–36.
Stevens, K. N. (**1993**). "Modelling affricate consonants," Speech Commun. **13**, 33–43.
Subtelny, J. D., Worth, J. H., and Sakuda, M. (**1966**). "Intraoral pressure and rate of flow during speech," J. Speech Hear. Res. **9**, 498–518.
Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., and Wilhelms-Tricarico, R. (**1997**). "Tongue surface displacement during bilabial stops," J. Acoust. Soc. Am. **102**, 562–571.
Tabain, M. (**2000**). "Variability in fricative production and spectra: Implications for the hyper- and hypo- and quantal theories of speech production," Lang Speech **44**, 57–94.
Warren, D. W., and Hall, D. J. (**1973**). "Glottal activity and intraoral pressure during stop consonant productions," Folia Phoniatr. **25**, 121–129.
Warren, D. W., Rochet, A. P., Dalston, R., and Mayo, R. (**1992**). "Controlling changes in vocal tract resistance," J. Acoust. Soc. Am. **91**, 2947–2953.
Westbury, J. R. (**1979**). "Aspects of the temporal control of voicing in consonant clusters in English," Texas Linguistic Forum **14**, 1–304.
Westbury, J. R. (**1983**). "Enlargement of the supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**, 1322–1336.
Yoshioka, H., Löfqvist, A., and Hirose, H. (**1981**). "Laryngeal adjustments in the production of consonant clusters and geminates in American English," J. Acoust. Soc. Am. **70**, 1615–1623.

Ranking vocal fold model parameters by their influence on modal frequencies

Douglas D. Cook[a)]
School of Mechanical Engineering, Purdue University, 140 South Martin Jischke Drive, West Lafayette, Indiana 47907

Eric Nauman
Weldon School of Biomedical Engineering, Department of Basic Medical Science, Purdue University, 206 South Martin Jischke Drive, West Lafayette, Indiana 47907

Luc Mongeau
Department of Mechanical Engineering, McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada

(Received 19 January 2009; revised 26 June 2009; accepted 30 June 2009)

The purpose of this study was to identify, using computational models, the vocal fold parameters which are most influential in determining the vibratory characteristics of the vocal folds. The sensitivities of vocal folds modal frequencies to variations model parameters were used to determine the most influential parameters. A detailed finite element model of the human vocal fold was created. The model was defined by eight geometric and six material parameters. The model included transitional boundary regions to idealize the complex physiological structure of real human subjects. Parameters were simultaneously varied over ranges representative of actual human vocal folds. Three separate statistical analysis techniques were used to identify the most and least sensitive model parameters with respect to modal frequency. The results from all three methods consistently suggest that a set of five parameters are most influential in determining the vibratory characteristics of the vocal folds. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3183592]

PACS number(s): 43.70.Bk [AL] Pages: 2002–2010

I. INTRODUCTION

Highly detailed numerical models of human voice production are powerful tools for phonation research. Such models have been used, for example, to predict mechanical stresses and strains in vocal fold tissue (Gunter, 2003; Tao *et al.*, 2006; Tao and Jiang, 2007). Computational models are becoming more accurate and more realistic. It is anticipated that detailed finite element models, in conjunction with laboratory experiments, may yield a better understanding of the formation of polyps and nodules, vocal fold damage, and healing. Models may also be useful in evaluating potential prosthetic devices and in improving articulatory models for voice synthesis.

It is well known that vocal fold vibrations are highly affected by the elastic constants used to characterize the mechanical deformation of the vocal folds in detailed models (Alipour-Haghihi and Titze, 1985). Numerical simulations of phonation have demonstrated that variations in elastic constants can lead to chaos (Berry *et al.*, 1994) or biphonation (Tao and Jiang, 2006). The mechanical properties of vocal fold tissue may vary by orders of magnitude between subjects (Kakita *et al.*, 1981, Chan and Titze, 1999, Zhang *et al.*, 2006). Vocal fold geometry is similarly variable. Uncertainty in tissue and geometric parameters does contribute to overall model errors. A recent literature review of studies involving computational vocal fold models revealed that *ad hoc* estimates have been used for the majority of tissue parameter inputs in computational vocal fold models (Cook, 2009). Many tissue parameter estimates are used repeatedly throughout the literature with no rigorous verification. For example, one *ad hoc* estimate of the longitudinal shear modulus of the vocal ligament (40 kPa) has been used in most, if not all, previous studies. Considering the previously mentioned variability in vocal fold tissue parameter values, it is unlikely that this particular parameter has a unique value.

In general, the effects of tissue parameter uncertainty have not been investigated over the full range of plausible values in vocal fold models. While parametric methods have been used occasionally (Berry and Titze, 1996; Cook and Mongeau, 2007), this approach has not been widely adopted. Perhaps one reason is because systematic parametric studies are often prohibitively expensive, especially when fluid-structure interactions are modeled.

The purpose of the present study was to investigate the vibration response of a vocal fold model across a broad range of model parameters. The approach was based on principles of stochastic modeling: Parameter *ranges* were used rather than discrete values in order to obtain a broader understanding of the influence of structural parameters on vocal folds resonance frequencies. The objective was to identify the most and least sensitive model parameters. The underlying general hypothesis is that vocal fold models are insensitive to certain model parameters, moderately sensitive to others, and highly sensitive to a select group of parameters. Identifica-

[a)]Author to whom correspondence should be addressed. Electronic mail: ddcook@purdue.edu

 0001-4966/2009/126(4)/2002/9/$25.00

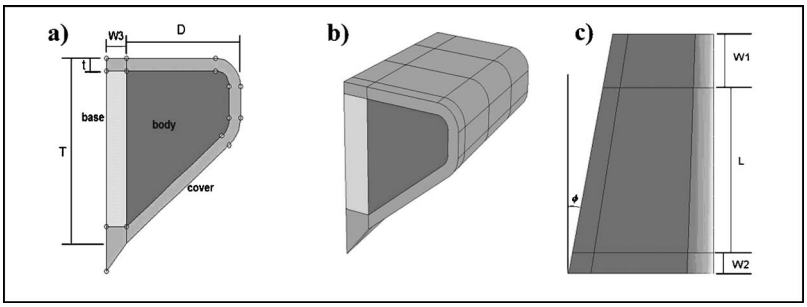


FIG. 1. Vocal fold model geometry: (a) coronal cross-section at the center of the mid-membranous region, (b) isometric view, and (c) superior view. Geometric symbol descriptions are listed in Table I.

tion and ranking of these parameters may provide valuable information for future model creation and may guide research in other areas where model complexity must be reduced. By focusing on the most sensitive parameters (and neglecting the least sensitive parameters), models may be created that are accurate and efficient while minimizing uncertainty. Furthermore, *a priori* confidence in such models may be enhanced when accounting for uncertainty over the broad parameter ranges found in vocal fold tissues.

II. VOCAL FOLD MODEL

A. Geometry

A three-dimensional body-cover model of the vocal folds was created after the two-layer models proposed by Hirano *et al.* (1981) and Story and Titze (1995). The vocal ligament was not assigned a distinct region but was assumed to be included as part of the cover. The model geometry was based on the two-dimensional M5 profile defined by Scherer *et al.* (2001). The cover was assumed to have a constant thickness over the medial, inferior, and superior surfaces. The two-dimensional coronal geometry [see Fig. 1(a)] was extruded in the anterior/posterior direction to obtain a three-dimensional geometry, as shown in Fig. 1(b). The depth (D) of the vocal fold decreased linearly from the anterior to the posterior ends to approximate the anterior/posterior asymmetry of the human vocal folds.

Vocal fold tissue has been reported to increase in stiffness toward the cartilage attachment points (Hirano *et al.*, 1987; Sataloff, 2005). Nearly all finite element vocal fold models (93%) have utilized the rigid rectangular boundary conditions of Titze and Strong (1975), with vocal fold tissue properties held constant throughout the structure. Hunter *et al.* (2004) questioned the accuracy of the rigid rectangular boundary conditions and used more realistic boundary conditions to investigate posturing of the vocal folds. Based on similar ideas, transitional boundary regions were added to both posterior and anterior of the vocal fold boundaries. The widths of these regions are labeled as W1, W2, and W3 in Fig. 1.

The paraglottic region (connective tissue lateral to the thyroarytenoid muscle and medial to the laryngeal cartilage (Hirano and Sato, 1993) was included in the model through the addition of an isotropic region at the base of the body region [indicated as "base" in Fig. 1(a)]. Rigid boundary conditions were applied on the anterior, posterior, and lateral faces, representative of the interface between the cartilages of the larynx and soft tissues of the vocal folds.

B. Material parameters

All tissues were modeled as incompressible. This approximation is commonly applied in vocal fold models and has been shown to introduce only minor errors in modal analysis (Cook *et al.*, 2008). The incompressibility assumption reduces the number of independent material parameters required to define the elastic properties of each tissue region. All tissues were assumed to be linearly elastic for the same reasons set forth in the previous reference.

The behavior of isotropic tissues is entirely defined by Young's modulus, E. Incompressible transversely isotropic tissues are defined by three material parameters: the transverse Young's modulus (E), the longitudinal Young's modulus (E'), and the longitudinal shear modulus (G'). The same density was assumed for all tissue types.

The vocal fold model was divided into seven regions, each with distinct material parameters. The vocal fold body and cover included anterior, mid-membranous, and posterior regions for a total of six regions. The seventh region was the paraglottic region. The cover regions and the paraglottic region were modeled as isotropic, with the body regions modeled as transversely isotropic.

All parameter values were assumed to be constant within the mid-membranous regions. Tissue parameters were assumed to vary continuously between the mid-membranous region and anterior or posterior faces. Although the cartilage was not explicitly modeled, the tissue property values at anterior and posterior faces of the model were representative of cartilage tissue. The spatial variation in all independent tissue parameters was determined as

$$S_{x/\mathrm{ant}} = S_{x/\mathrm{mid}} + (S_{\mathrm{cart}} - S_{x/\mathrm{mid}})\left(\frac{x - L - w_2}{w_1}\right)^P, \tag{1}$$

$$S_{x/\mathrm{pos}} = S_{x/\mathrm{mid}} + (S_{\mathrm{cart}} - S_{x/\mathrm{mid}})\left(\frac{w_2 - x}{w_2}\right)^P. \tag{2}$$

In these equations, $S_{x/\mathrm{ant}}$ and $S_{x/\mathrm{pos}}$ represent the varied parameter S_x in the anterior and posterior boundary regions, respectively. The corresponding cartilage value is S_{cart}, and the parameter value at the mid-membranous region is $S_{x/\mathrm{mid}}$. The equation for the variation of the cover stiffness in the anterior region was

$$E_{c/\mathrm{ant}} = E_c + (E_{\mathrm{cart}} - E_c)\left(\frac{x - L - w_2}{w_1}\right)^P. \tag{3}$$

A similar relation was used for other material parameters: E_c, E_b, E', and G'.

TABLE I. Parameter descriptions, symbols, ranges, and references. The first eight parameters refer to the model geometry; the last seven refer to material properties.

Parameters	Symbol	Range	Reference
Length	L	6.8–10.2 mm	Friederich *et al.*, 1993; Titze, 2006
Depth	D	7–10 mm	Estimated
Thickness	T	5.9–8.8 mm	Titze, 2006
Cover thickness	t	0.84–1.26 mm	Titze, 2006; Stiblar-Martincic, 1997
Angle	φ	0°–28°	Estimated
Length of W 1	W1	3.6–5.4 mm	Friederich *et al.*, 1993
Length of W 2	W2	1.0–1.6 mm	Friederich *et al.*, 1993
Depth of W 3	W3	1.6–2.4 mm	Inferred from Hirano and Sato, 1993
Cover stiffness	E_c	4–50 kPa	Zhang *et al.*, 2006
Transverse Young's modulus	E_b	1–30 kPa	de Vries *et al.*, 1999; Gunter, 2003
Longitudinal Young's modulus	E'	E_b–E_{cart}	E_b and E_{cart}
Stiffness of W3	E_{base}	50–1000 kPa	Estimated
Longitudinal shear modulus	G'	5–100 kPa	Rosa *et al.*, 2003; Berry and Titze, 1996
Degree of transition function	P	1–66	Estimated
Cartilage Young's modulus	E_{cart}	10 MPa (fixed)	Rains *et al.*, 1992; Roberts *et al.*, 1998

For all isotropic regions, Poisson's ratio was held constant, $\upsilon = \frac{1}{2}$, and the shear modulus varied as a function of local stiffness. For example, in the case of cover shear modulus,

$$G_{c/\text{ant}} = \frac{E_{c/\text{ant}}}{2(1+\nu)}, \tag{4}$$

$$G_{c/\text{pos}} = \frac{E_{c/\text{pos}}}{2(1+\nu)}. \tag{5}$$

For transversely isotropic regions, the shear modulus varied as a function of (local) longitudinal and transverse Young's modulus values,

$$\mu = \frac{E_b E'}{4E' - E_b}. \tag{6}$$

The Poisson's ratios also varied spatially according to the following equations:

$$\nu' = \tfrac{1}{2}, \tag{7}$$

$$\nu^{+} = \frac{E}{2E'}, \tag{8}$$

$$\nu = 1 - \frac{E}{2E'}. \tag{9}$$

The above methods allowed the tissue stiffness, Poisson's ratios, and shear moduli to vary continuously between the mid-membranous region and the cartilage values at the anterior and posterior surfaces. The independent parameters governing these transitions were the mid-membranous material parameters (E_c, E_b, E', G'), the cartilage tissue parameter (E_{cart}), and the exponent P which controlled the rate of transition between regions.

1. Parameter ranges

The vocal fold model described above is defined by eight geometric and seven material (tissue) parameters. To obtain a robust description of the vocal fold structure, parameter ranges were established for all independent parameters based upon available measured data or *ad hoc* estimates. The guiding principle in the range selection was to encompass approximately 75% of human vocal fold values.

For example, the medial/lateral thickness of the vocal folds was reported by Titze (2006) as having a mean value of 9.34 mm with a standard deviation of 1.63 mm. Assuming a normal distribution, 75% of vocal fold thicknesses should lie between 5.87 and 8.81 mm. When mean and standard deviation values were not available, *ad hoc* estimates based on experimental data from measurements of similar tissues or previous vocal fold models were used.

Vocal fold tissues are known to exhibit nonlinear stress-strain relationships (Chan and Titze, 1999; Zhang *et al.*, 2006). Nonlinear relations are available for only a few of the parameters identified in Sec. II B. The use of nonlinear constitutive models requires additional model parameters, most of which are extremely difficult or impossible to estimate. This problem was addressed by choosing linear material parameters over ranges that were sufficiently broad to encompass both zero strain and moderate (i.e., 10%–15%) strain parameter values. All geometric and material parameters, with their associated ranges, are listed in Table I.

C. Numerical implementation

This study employed more than 2000 unique vocal fold models. The commercial finite element software COMSOL, along with MATLAB Version 3.4 was used. A custom subroutine created each finite element model based on unique sets of 14 parameters (Table I). For each set of parameters, a vocal fold geometry was defined, tissue properties were assigned, a finite element mesh was created, boundary conditions were imposed, and modal analysis was performed. The output consisted of eight unique eigenvectors and eight associated eigenvalues or resonance frequencies. Eigenvectors contained the x-, y-, and z-displacement values for each node.

TABLE II. Modal analysis comparisons for three different methods. Rightmost column indicates the maximum percent difference between COMSOL predictions and the remaining two methods.

Mode	Mode shape[a]	Modal frequency (Hz)			Maximum difference
		Experimental	ABAQUS	COMSOL	
1	z-10	75	74.17	74.52	−6.9%
2	z-20	···	106.55	107.08	0.50%
3	y-10	···	125.57	125.67	0.08%
4	*x*-10	···	133.99	134.63	0.48%
5	*x*-11	···	150.38	150.91	0.35%
6	*x*-20	···	151.32	152.08	0.50%

[a]Mode shape convention: Berry *et al.* (1994).

For each vocal fold model, the mesh consisted of approximately 1500 quadratic finite elements with over 28 000 degrees of freedom. A mesh refinement study was performed to verify spatial convergence. The model was validated through comparisons with measured data for a synthetic physical model of the vocal folds (Chen *et al.*, 2008). The synthetic model geometry was similar to that of the model described above, with a single isotropic layer and no transitional boundary regions. Due to high damping in the synthetic model, experimental modal analysis results were obtained for only the first mode of vibration. Computational modal analysis was also performed using the commercial software package ABAQUS. The experimental studies and ABAQUS simulations were performed independently. Table II provides comparisons between these methods. Very good agreement was found between the COMSOL and ABAQUS models, both of which adequately predicted the experiment.

III. DATA ANALYSIS TECHNIQUES

Three different techniques were applied in the analysis of the output data. The first was a screening method, as set forth by Cotter (1979). The second examined the correlation between parameter values and modal frequencies. The final method was based on descriptive statistics of 100 local sensitivity simulations. These methods are described in Secs. III A–III C.

A. Cotter's method

So-called screening methods are one class of methods used to perform sensitivity analysis (Saltelli *et al.*, 2000). Screening methods provide a ranking of model parameters based on the importance or influence of each parameter on the overall model response. Screening methods utilize a limited number of simulations to reduce computational cost but occasionally fail to identify key parameters and cannot precisely quantify the contribution of each parameter. These methods yield essentially first-order estimates of each parameter's importance.

The method proposed by Cotter (1979) involves $2n+2$ total simulations, where n is the number of model parameters. A two-level factorial design was used with all parameter values set to their highest or lowest value. The first simulation (case zero) is performed with all parameters set to low values. The final simulation is performed with all parameters set to high values. Cases 1 through n are based on the zero case with all parameters at their lowest values and with one parameter at a time set to its highest value. Cases $(n+1)$ through $2n$ are variations of the final case with one parameter at a time set to its lowest value.

Cotter provided a method for estimating the importance of each factor based on the $2n+2$ simulation outputs designated as y_i. The relative importance of the jth parameter, $M(j)$, is given by

$$M(j) = |y_{2n+1} - y_{n-j} + y_j - y_0| + |y_{2n+1} - y_{n+j} - y_j - y_0|. \tag{10}$$

The leftmost absolute value term quantifies the effect of varying only one parameter at a time. The second (right) absolute value term accounts for interaction effects.

B. Correlation between modal frequencies and model parameters

The ranges given in Table I were used to define uniform probability density functions (PDFs) for each parameter. Vocal fold models were created by randomly selecting each parameter from its corresponding distribution. A total of 100 vocal fold models were created in this fashion $(n=100)$. Pearson correlations were used to determine possible relationships between simulation outputs (modal frequencies) and each of the input parameters. The correlation coefficient, r_{ij}, was calculated for each of the first eight modal frequencies (subscript i) and for each of the 14 input parameters (subscript j). This yielded 112 total correlation values. The coefficient of determination (r^2) was also calculated. Finally, p-values were obtained via a hypothesis test. The null hypothesis (H_0) was that no relationship existed between modal frequencies and parameter values $(H_0: r=0)$. The alternative hypothesis is that modal frequencies and model parameters were correlated with an r-value greater than or equal to each respective r_{ij}-value $(H_A: r \geq r_{ij})$. The test statistic was $t=(n-2)r^2/(1-r^2)^{1/2}$, which follows a standard t-distribution. Since the alternative hypothesis was that $r \geq r_{ij}$, a one-sided t-distribution was used for determination of p-values.

C. Local sensitivity analysis

The final analysis method involved computing local sensitivity values for each mode and parameter. The local sensitivity of each modal frequency, M_i, to each parameter, X_j, was defined as the partial derivative,

$$S_{ij} = \frac{\partial M_i}{\partial X_j}. \tag{11}$$

The sensitivity, S_{ij}, was evaluated using a finite difference approximation. First, a nominal case, M^{nom}, was randomly selected, and its modal frequencies were calculated. Next, each parameter X_j was increased slightly, with all other parameters held at their nominal values. Finally, the sensitivity to each modal frequency was computed as

$$S_{ij} = \frac{\partial M_i}{\partial X_j} \approx \frac{M_i' - M_i^{\text{nom}}}{X_j' - X_j^{\text{nom}}}, \tag{12}$$

where the nom superscript refers to nominal values and the primed quantities are associated with incremented X_j values.

The units of the sensitivity depend upon the parameter of interest. To allow comparison between sensitivities of different parameters, each parameter was non-dimensionalized as follows:

$$X_j^* = \frac{X_j}{R_j}, \tag{13}$$

where R_j refers to the range of parameter j. Using P_j^*, the sensitivity was expressed as

$$S_{ij} = \frac{\partial M_i}{\partial X_j^*} \approx \frac{M_i' - M_i^{\text{nom}}}{(X_j' - X_j^{\text{nom}})/R_j}. \tag{14}$$

This sensitivity can be interpreted as the change in modal frequency due to a unit percent change in parameter j. This approach was used to obtain 112 total sensitivities for each nominal case. The sample of 100 vocal fold models used for correlation analysis was used as the nominal set, with sensitivities calculated for each unique model. This process required the creation of 1500 individual vocal fold models (100 nominal cases, with 1400 additional cases required for sensitivity calculations). The computational time was approximately 16 h using two (dual) 2.2 GHz XEON processors with 1 Gbyte random access memory.

IV. RESULTS

A. Cotter's method

Cotter's method yields a single importance value for each parameter-mode pair for a total of 112 importance values. Importance values were observed to depend primarily on each parameter, with no apparent dependence on modal frequency. Importance value statistics are presented in Table III for each parameter, with distributions of importance values indicated by their standard deviation as well as their maximum and minimum values.

Cotter's method identified the cover stiffness, E_c, as the most important parameter, followed by P, G', E_b, and L. The five least important parameters were identified as W1, t, φ, W2, and W3. The importance values for individual modes were remarkably consistent among the five most important parameters. In fact, 85% of individual importance values from the lead group were greater than the highest importance value found in the remaining nine parameters. A similar consistency was found among the five least important parameters.

TABLE III. Importance values obtained from Cotter's method. Each row based upon eight parameter-modal frequency pairs ($n=8$).

	Importance value statistics			
Parameter	Mean (st. dev)		Max/min	Rank
L	21.03	(5.40)	31.5/16.1	5
D	19.23	(1.63)	21.9/17.6	6
T	8.85	(2.53)	13.0/3.8	9
t	5.99	(2.54)	9.9/2.7	11
φ	5.15	(2.05)	9.4/2.9	12
W1	7.47	(2.58)	12.0/4.2	10
W2	2.93	(0.70)	4.0/2.2	13
W3	1.79	(0.41)	2.7/1.4	14
E_c	54.14	(14.9)	74.9/33.1	1
E_b	31.07	(8.17)	40.2/15.8	4
E'	13.62	(6.30)	23.6/3.4	7
E_{base}	9.09	(2.48)	12.6/5.8	8
G'	33.20	(6.21)	42.5/24.6	3
P	35.56	(5.84)	44.7/30.2	2

B. Correlation results

For a sample size of 100 models, Pearson correlation coefficients, r_{ij}, coefficients of determination, $(r_{ij})^2$, and p-values were calculated. A total of 112 values were obtained for each of these three statistical coefficients. The results were again primarily dependent on parameter values, with little dependence on mode number. Means, standard deviation, maximum, and minimum values are reported for the data corresponding to each parameter in Table IV. The Pearson correlation coefficients ranged from $r_{3,14}=-0.45$ to $r_{8,12}=0.60$. The absolute values of correlation coefficients ranged from $r_{4,7}=0.0047$ to $r_{8,12}=0.60$. The mean and standard deviations of r-values for all modes are given in Table IV. Positive r-values indicate that the parameters are positively correlated with modal frequency, while negative values indicate negative correlations. Although r-values were much lower than typically encountered in engineering experiments, statistical significance was due to a large number of degrees of freedom (df=99).

The five most correlated with model frequency were E_c, E_b, P, G', and L. The five parameters least correlated with modal frequency were φ, E', t, W2, and W1. As expected, tissue stiffness parameters typically showed a positive correlation with modal frequency, while geometric parameters tended to have a negative correlation. Some exceptions to these generalizations included E', which had a weak negative correlation, as well as T and t, which exhibited weakly positive correlations.

The r^2-values can be interpreted as the percentage of variation in modal frequency associated with each parameter.

TABLE IV. Statistics of r, r^2, and p-values. Means and standard deviations were based on the eight modal frequency values for each parameter. Values in bold highlights the r, r^r, and p-values of significantly correlated parameters.

	r-values			r^2-values			p-values			
Parameters	Mean (st. dev)		Max/min	Mean (st. dev)		Max/min	Mean (st. dev)		Max/min	Rank
L	**−0.27**	(0.03)	−0.24/−0.31	**0.08**	(0.02)	0.10/0.06	**0.004**	(0.00)	0.007/0.001	**5**
D	**−0.20**	(0.02)	−0.17/−0.22	**0.04**	(0.01)	0.05 /0.03	**0.027**	(0.01)	0.044/0.013	**7**
T	0.09	(0.05)	0.15/0.02	0.01	(0.01)	0.02 /0.00	0.211	(0.13)	0.412/0.070	10
t	0.04	(0.03)	0.07/−0.01	0.00	(0.00)	0.01 /0.00	0.354	(0.09)	0.477/0.234	12
φ	−0.09	(0.03)	−0.05/−0.13	0.01	(0.01)	0.02 /0.00	0.198	(0.08)	0.296/0.094	9
W1	0.00	(0.01)	0.02/−0.01	0.00	(0.00)	0.00 /0.00	0.458	(0.03)	0.498/0.430	14
W2	0.02	(0.01)	0.03/0.00	0.00	(0.00)	0.00 /0.00	0.429	(0.04)	0.482/0.375	13
W3	**−0.23**	(0.02)	−0.20/−0.26	**0.05**	(0.01)	0.07 /0.04	**0.012**	(0.00)	0.020/0.005	**6**
E_c	**0.55**	(0.04)	0.60/0.50	**0.30**	(0.04)	0.36 /0.25	**0.000**	(0.00)	0.000/0.000	**1**
E_b	**0.44**	(0.05)	0.54/0.38	**0.20**	(0.05)	0.29 /0.14	**0.000**	(0.00)	0.000/0.000	**2**
E'	−0.07	(0.02)	−0.05/−0.09	0.01	(0.00)	0.01 /0.00	0.245	(0.06)	0.324/0.181	11
E_{base}	0.11	(0.02)	0.13/0.08	0.01	(0.00)	0.02 /0.01	0.141	(0.04)	0.202/0.095	8
G'	**0.31**	(0.03)	0.35/0.27	**0.09**	(0.02)	0.13 /0.07	**0.001**	(0.00)	0.003/0.000	**4**
P	**−0.38**	(0.04)	−0.35/−0.45	**0.15**	(0.03)	0.20 /0.12	**0.000**	(0.00)	0.000/0.000	**3**

This suggests that approximately 30% of the modal frequency variation is due to E_c, with E_b, P, G', and L contributing 20%, 15%, 9%, and 8%, respectively.

p-values were also calculated to test the null hypothesis that no relationship existed between parameters and modal frequencies. Seven parameters possessed p-values lower than 0.025, with the majority of p-values in this group lower than 0.001. For these parameters, the null hypothesis was rejected in favor of the alternative hypothesis. Several parameters had p-values greater than 0.025, indicating a non-negligible possibility that the null hypothesis may be true. These parameters included T, t, φ, W1, W2, E', and E_{base}.

C. Local sensitivity results

As described in Sec. III C, a total of 11 200 total local sensitivities were calculated. These sensitivities were then grouped into 800 sensitivities for each parameter. The absolute value of the 10% trimmed mean was chosen to rank parameter influence since this value is more moderate than either the mean or median. The most sensitive parameters were P, E_c, G', E_b, and L. The least sensitive parameters were T, E_{base}, W2, φ, and W3.

The model parameters are grouped into geometric (L, D, T, t, φ, W1, W2, and W3) and tissue parameters (E_c, E_b, E', E_{base}, and G'). The *signs* of the 10% trimmed mean values reveal some expected characteristics of the model. For tissue parameters, all mean sensitivity values were positive, and 99.8% of all sensitivities associated with these parameters were also positive. This is in agreement with expectations that an increase in stiffness increases modal frequency. Geometric parameters typically exhibited negative local sensitivities, as expected, with 80% of geometric sensitivities below zero. Although increased mass typically reduces modal frequencies, geometric parameters also influence the model stiffness through boundary conditions. Tradeoffs between added mass and added boundary area vary, thus causing some geometric parameters such as thickness and extrusion angle to have positive trimmed means. The same effect is responsible for the slight positive (but insignificant) correlations observed for certain geometric parameters (see Table IV).

A high sensitivity to the rate of transition in the transitional boundary regions (P) was observed. Low values of P caused the boundary effects to extend further inward from each fixed boundary. For high values, the boundary regions act as extensions of the mid-membranous region, thus effectively lengthening the vibrating length of the vocal folds.

D. Comparison between methods

The three methods utilized are not directly comparable. Each relies upon different assumptions, incorporates different amounts of data, and produces different rankings. However, the methods can be compared in a relative sense. To this end, the results of each method were normalized by the sum of all primary results. This approach preserves relative importance within each method while allowing the various rankings to be compared with each other. For example, the trimmed mean values of Table V were summed, after which each individual value was divided by the sum. Similar approaches were used for the mean values of Table III and the r-values of Table IV. The results are shown graphically in Fig. 2 below. The parameters were ordered along the ordinate axis according to the average of each parameter's three normalized values.

As seen in Fig. 2, the three methods are in general agreement, but they do not always agree. The most obvious differences are observed for P and W3. The P parameter was consistently ranked among the top 5 but exhibited an unusually high value from the local sensitivity method. This seems due to the fact that the model was observed to be extremely sensitive to low values of P (i.e., 1–5), and less sensitive to values above 5. The third transitional boundary region (W3) was identified as the sixth most influential parameter by the correlation method, whereas both Cotter's method and the local sensitivity approach ranked this parameter among the least sensitive parameters. These discrepancies may be due

TABLE V. Statistics for local sensitivity values. Each row is based upon local sensitivity values ($n=800$).

Parameters	Local sensitivity statistics			
	Mean (st. dev)	Trimmed mean	Median	Rank
L	−48.63 (20.3)	−50.7	−51.6	5
D	−13.10 (9.64)	−14.1	−14.9	6
T	7.17 (10.9)	7.0	6.8	10
t	−6.30 (11.0)	−7.1	−7.9	9
φ	1.55 (57.3)	2.4	4.6	13
W1	−11.70 (6.97)	−12.0	−12.3	7
W2	−3.51 (4.15)	−3.6	−3.8	12
W3	−0.92 (3.91)	−1.1	−1.4	14
E_c	75.66 (58.7)	81.3	85.3	2
E_b	47.38 (97.9)	55.4	61.9	4
E'	4.18 (159.)	8.4	29.0	8
E_{base}	2.88 (48.9)	5.0	8.1	11
G'	49.33 (82.0)	60.4	65.2	3
P	−310.9 (387)	−362.7	−395.6	1

to sample size, the choice of parameter ranges, or unidentified interaction effects with other parameters. The general agreement between methods, especially concerning the most sensitive parameters, was judged to be more compelling than the discrepancies.

V. DISCUSSION

A. Difficulty in mode shape identification

The mode shapes of the human vocal folds are often described using the x-10, x-11 nomenclature of Titze and Strong (1975). These mode shapes take on different modal frequencies depending upon model parameters, and modal frequencies have been shown to cross as model parameters are varied [see Berry and Titze (1996), Fig. 3]. Because eigenvalue solvers generally order results according to increasing modal frequency, eigenvectors (mode shapes) are not immediately differentiable based on modal frequencies alone. For example, the x-11 mode might appear as the third eigenvector for one set of model parameters but as the fourth eigenvector for a different set of model parameters.

With hundreds of distinct models created over the course of this study, mode shape identification was a major challenge. Simultaneous variation in all model parameters over broad ranges exacerbated this problem, leading to high variance in modal frequencies. It was observed that the PDF of modal frequencies for any given mode shape exhibited substantial overlap with the PDFs of other mode shapes. An analogous situation was also observed in the eigenvector space. Although all finite element models had the same mesh (to allow eigenvector comparisons), eigenvectors exhibited a high level of variance, making mode shape identification difficult. Several comparative schemes were attempted but produced unacceptable rates of false-positive and/or false-negative identifications.

Because of this difficulty, the modal frequencies were analyzed as obtained from the eigenvalue solver (i.e., ordered by modal frequency). It was anticipated that this approach would have a moderating effect on analysis outcomes since highly sensitive mode shapes would occasionally be combined with less sensitive ones, and vice versa. For the purposes of identifying the most and least sensitive parameters at first-order accuracy, this shortcoming was deemed acceptable. The general agreement between the three analysis methods suggests that this approach was justified.

B. Variation in model parameters in relation to previous models

Cook (2009) reported that the average number of parameters for vocal fold models was 20 with an average of three parameters varied in each study. Of the tissue parameters used in this study (E_c, E_b, E', and G'), an average of five unique parameter values have been reported in the literature (Cook, 2009). While this figure is certainly an underestimate (many values were found to be unreported), the present study utilized over 100 unique values for each of 14 parameters. All parameters were varied simultaneously, thus allowing for a broader sampling of the entire parameter space than can be obtained by varying one parameter at a time as is typically

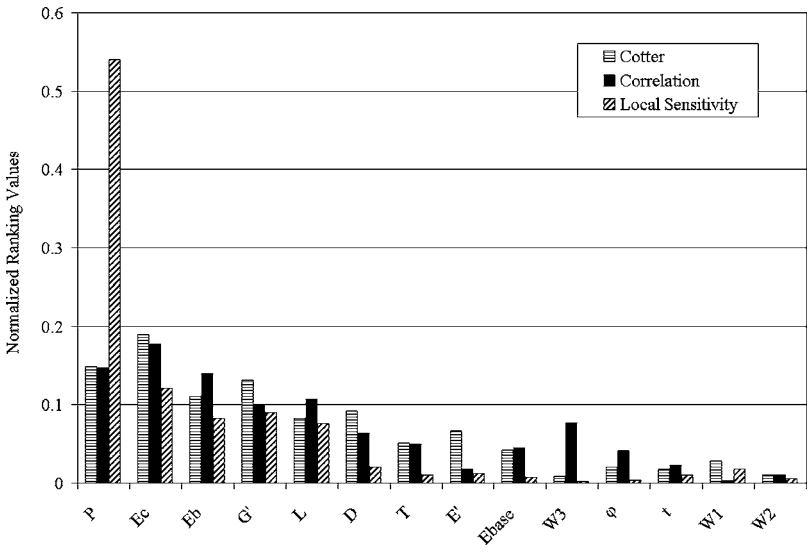


FIG. 2. Normalized ranking values for comparison between methods.

done. As such, the models described in this study account for more variation in vocal fold structure than all previous computational vocal fold studies combined.

C. Discussion of results

The reliability of the results presented above is partially dependent on the number of data points utilized by each method. Cotter's method relied upon 240 total modal frequency values, while the correlation and local sensitivity approaches utilized 800 and 11 200 modal frequencies, respectively. From this perspective, the latter two approaches are more reliable since they provide more comprehensive descriptions of how the model behaves across the entire spectrum of possible parameter combinations.

The most striking feature observed in the application of these three methods was the consistency with which the lead group of parameters, consisting of E_c, P, E_b, G', and L, were identified as influential on modal frequencies. Four of these parameters are related to tissue properties, and only one parameter (L) is a geometric parameter. There are two reasons for identifying these *five* parameters as the most sensitive. First, each of these parameters was ranked within the top five by each of the three methods of analysis. Second, certain parameters known from prior research to strongly influence vocal fold vibration should be expected to appear among the group of most sensitive parameters. Such parameters included the stiffness of the vocal fold cover (E_c) and vocal fold length (L). The mechanical properties of the cover are known to be a dominant factor in healthy phonation (Chan and Titze, 1999). Vocal fold length has been attributed to the difference between male and female fundamental frequencies of phonation (Friederich *et al.*, 1993; Perkins and Kent, 1986). Finally, the presence of E_c and L among the most sensitive parameters indicates the importance of the remaining parameters, all of which were observed to be more influential than the vocal fold length.

The least sensitive parameter ratings were less consistent than the most sensitive parameters. The parameters W1, W2, and t appeared among the five least sensitive parameters for each method. The vocal fold thickness (T) was ranked 9th, 10th, and 10th by the three respective methods. The longitudinal body stiffness (E') was ranked as 11th by both regression and correlation methods but was ranked 7th by Cotter's method. Both T and E' were included with the least sensitive parameters based on their consistently low ranking by both correlation and local sensitivity approaches. Thus, the five least sensitive parameters were found to be T, E', t, W1, and W2.

It is difficult to make any definitive statement about the remaining parameters (D, W3, E_{base}, and φ). Some of these parameters (such as D) may have a substantial influence on the dynamic response of the vocal folds, while others (such as E_{base}) may have little influence. More advanced analysis, involving fluid loading of the vocal fold structure and/or sorting of modal frequency data, will be required to provide a more accurate ranking of these parameters.

D. Application of results

The above results have several interrelated applications. First, these results can be used to guide the experimental measurement of vocal fold tissues. The measurement of highly sensitive parameters should be prioritized over the measurement of less sensitive parameters. Furthermore, sensitive parameters require both more accurate and more comprehensive measurements than less sensitive parameters. In other words, the ranges, distributions, and gender- and age-dependent relationships of highly sensitive parameters should be characterized with greater detail.

Second, the identification of the most and least sensitive model parameters has great importance for computational models of phonation. Models should be created with all parameter sensitivities in mind. This is a double-edged issue. On one hand, the identification of non-influential parameters allows those parameter values to be chosen with relative impunity since these choices will have little effect of model response. On the other hand, a great deal of care and attention must be paid to highly sensitive model parameters, many of which have not yet been characterized in sufficient detail. In fact, previous models of phonation have typically utilized *ad hoc* estimates for several of the most sensitive material parameters (Cook *et al.*, 2009) in review, supplemental document]. Highly sensitive parameters should either be based on experimental data or varied parametrically to account for uncertainty until more such data are available.

Third, the ranking of model parameters may be useful in the creation of patient-specific models of phonation and (indirectly) in the evaluation of individuals suffering from voice disorders. Based on the ranking presented above, the least sensitive parameters would be based on general population trends, and even some moderately sensitive parameters could be estimated. However, a sufficiently accurate patient-specific model would require patient-specific data for the most sensitive vocal fold parameters. The achievement of patient-specific models will undoubtedly require a great deal of research focused on methods for obtaining *in vivo* measurement of the most sensitive parameter values. The ranking of model parameters provides some guidance on this front, suggesting which parameters should be targeted initially. This issue is related to the characterization of parameter distributions. For example, detailed experimental characterization of vocal fold parameters may reveal that certain parameters (such as P), although highly influential, exhibit relatively narrow distributions and thus may be estimated without introducing a great deal of error in the analysis.

Finally, the identification of several highly influential parameters may suggest new approaches for treating certain voice disorders, especially thyroplasty implants and injections. For example, vocal fold medialization implant designs could be improved to mimic the most influential human vocal fold characteristics as reported above. New medialization injection materials could also be developed to exhibit sensitivities similar to those of the actual human vocal folds. Both approaches would utilize a more complete understanding of the human vocal fold structure to create treatments that restore functionality of the human voice.

VI. CONCLUSIONS

The application of three separate methods was used to identify the most and least sensitive parameters of a vocal fold model. All three methods identified the following parameters as most sensitive: the stiffness of the cover (E_c), the stiffness of the body (E_b), the rate of transition between midmembranous and cartilage stiffness (P), the longitudinal shear stiffness of the body (G'), and the vocal fold length (L). The least sensitive model parameters included the following (in descending order): E', T, t, W1, and W2. Moderately sensitive parameters were identified as D, W3, E_{base}, and φ. Such ranking of vocal fold parameters is useful to guide (a) experimental measurements and characterization of vocal fold parameters, (b) the creation of accurate computational models of phonation in general, (c) the development of techniques for obtaining patient-specific models of phonation, and (d) further improvements in the treatment of voice disorders.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the National Institute on Deafness and Other Communication Disorders (Grant Nos. R01 DC005788 and R01 DC008290), the National Science Foundation (Grant No. CBET-0828903), and the ASME Teaching Fellowship. Li-Jen Chen graciously provided experimental and computational modal analysis results for validation of the models in this study. Portions of this paper were originally presented at the 6th International Congress on Voice Physiology and Biomechanics in Tampere, Finland 6–9 August 2008.

Alipour-Haghihi, F., and Titze, I. R. (**1985**). "Simulation of particle trajectories of vocal fold tissue during phonation," in *Vocal Fold Phyisiology: Biomechanics, Acoustics, and Phonatory Control*, edited by I. R. Titze and R. S. Scherer (Denver Center for the Performing Arts, Denver), pp. 183–190.

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

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

Chan, R. W., and Titze, I. R. (**1999**). "Viscoelastic shear properties of human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008.

Chen, L. J., Zanartu, M., Cook, D., and Mongeau, L. (**2008**). "Effects of acoustic loading on the self-oscillations of a synthetic model of the vocal folds," Proceedings of the Ninth International Conference on Flow-Induced Vibration, Prague, Czech Republic, 30 June–3 July, edited by I. Zolotarev and J. Horacek.

Cook, D. D. (**2009**). "Systematic structural analysis of human vocal fold models," Ph.D. dissertation, Purdue University, IN.

Cook, D. D., and Mongeau, L. (**2007**). "Sensitivity of a continuum vocal fold model to geometric parameters, constraints, and boundary conditions," J. Acoust. Soc. Am. **121**, 2247–2253.

Cook, D. D., Nauman, E., and Mongeau, L. (**2008**). "Reducing the number of vocal fold mechanical tissue properties: Evaluation of the incompressibility and planar displacement assumptions," J. Acoust. Soc. Am. **124**, 3888–3896.

Cotter, S. C. (**1979**). "A screening design for factorial experiments with interactions," Biometrika **66**, 317–320.

de Vries, M., Schutte, H., and Verkerke, G. (**1999**). "Determination of parameters for lumped parameter models of the vocal folds using a finite-element method approach," J. Acoust. Soc. Am. **106**, 3620–3628.

Friedrich, G., Kainz, J., and Freidl, W. (**1993**). "Zur funktionellen strucktur der menschlichen stimmlippe (Functional structure of the human vocal cord)," Laryngorhinootologie **72**, 215–224.

Gunter, H. (**2003**). "A mechanical model of vocal-fold collision with high spatial and temporal resolution," J. Acoust. Soc. Am. **113**, 994–1000.

Hirano, M., Kurita, S., and Nakashima, T. (**1981**). "The structure of the vocal folds," in *Vocal Fold Physiology*, edited by K. Stevens and M. Hirano (University of Tokyo, Tokyo), pp. 33–41.

Hirano, M., and Sato, K. (**1993**). *Histological Color Atlas of the Human Larynx* (Singular Publishing Group, San Diego, CA).

Hirano, M., Yoshita, T., Kurita, S. Kiyokawa, K., Sato, K., and Tateishi, O. (**1987**). "Anatomy and behavior of the vocal process," in *Laryngeal Function in Phonation and Respiration*, edited by T. Baer, C. Sasaki, and K. Harris (College-Hill, Boston, MA), pp. 1–13.

Hunter, E. J., Titze, I. R., and Alipour, F. (**2004**). "A three-dimensional model of vocal fold abduction/adduction," J. Acoust. Soc. Am. **115**, 1747–1759.

Kakita, Y., Hirano, M., and Ohmaru, K. (**1981**). "Physical properties of the vocal fold tissue," in *Vocal Fold Physiology*, edited by K. Stevens and M. Hirano (University of Tokyo, Tokyo, Japan).

Perkins, W., and Kent, R. (**1986**). *Functional Anatomy of Speech, Language, and Hearing: A Primer* (Allyn and Bacon, Boston, MA).

Rains, J. K., Bert, J. L., Roberts, C. R., and Pare, P. D. (**1992**). "Mechanical properties of human tracheal cartilage," J. Appl. Physiol. **72**, 219–225.

Roberts, C. R., Rains, J. K., Pare, P. D., Walker, D. C., Wiggs, B., and Bert, J. L. (**1998**). "Ultrastructure and tensile properties of human tracheal cartilage," J. Biomech. **31**, 81–86.

Rosa, M. O., Pereira, J. C., Grellet, M., and Alwan, A. (**2003**). "A contribution to simulating a three-dimensional larynx model using the finite element method," J. Acoust. Soc. Am. **114**, 2893–2905.

Saltelli, A., Chan, K., and Scott, E. M. (**2000**). *Sensitivity Analysis* (Wiley, Chichester, NY).

Sataloff, R. T. (**2005**). *Voice Science* (Plural Publishing, San Diego, CA), p. 62.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Stiblar-Martincic, D. (**1997**). "Histology of laryngeal mucosa," Acta Oto-Laryngol., Suppl. **527**, 138–141.

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.

Tao, C., and Jiang, J. J. (**2006**). "Anterior-posterior biophonation in a finite element model of vocal fold vibration," J. Acoust. Soc. Am. **120**, 1570–1577.

Tao, C., and Jiang, J. J. (**2007**). "Mechanical stress during phonation in a self-oscillating finite-element vocal fold model," J. Biomech. **40**, 2191–2198.

Tao, C., Jiang, J. J., and Zhang, Y. (**2006**). "Simulation of vocal fold impact pressures with a self-oscillating finite-element model," J. Acoust. Soc. Am. **119**, 3987–3994.

Titze, I. R. (**2006**). *The Myoelastic Aerodynamic Theory of Phonation* (National Center for Voice and Speech, Denver, CO), p. 84.

Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **57**, 736–744.

Zhang, K., Siegmund, T., and Chan, R. W. (**2006**). "A constitutive model of the human vocal fold cover for fundamental frequency regulation," J. Acoust. Soc. Am. **119**, 1050–1062.

Acoustic-articulatory mapping in vowels by locally weighted regression

Richard S. McGowan[a)]
CReSS LLC, 1 Seaborn Place, Lexington, Massachusetts 02420

Michael A. Berger[b)]
Department of Linguistics, University of Rochester, 503 Lattimore Hall, Rochester, New York 14627

(Received 10 December 2008; revised 5 May 2009; accepted 30 June 2009)

A method for mapping between simultaneously measured articulatory and acoustic data is proposed. The method uses principal components analysis on the articulatory and acoustic variables, and mapping between the domains by locally weighted linear regression, or loess [Cleveland, W. S. (1979). J. Am. Stat. Assoc. **74**, 829–836]. The latter method permits local variation in the slopes of the linear regression, assuming that the function being approximated is smooth. The methodology is applied to vowels of four speakers in the Wisconsin X-ray Microbeam Speech Production Database, with formant analysis. Results are examined in terms of (1) examples of forward (articulation-to-acoustics) mappings and inverse mappings, (2) distributions of local slopes and constants, (3) examples of correlations among slopes and constants, (4) root-mean-square error, and (5) sensitivity of formant frequencies to articulatory change. It is shown that the results are qualitatively correct and that loess performs better than global regression. The forward mappings show different root-mean-square error properties than the inverse mappings indicating that this method is better suited for the forward mappings than the inverse mappings, at least for the data chosen for the current study. Some preliminary results on sensitivity of the first two formant frequencies to the two most important articulatory principal components are presented.
 [DOI: 10.1121/1.3184581]

PACS number(s): 43.70.Bk [DAB] Pages: 2011–2032

I. INTRODUCTION

The acoustic response to articulatory behavior is determined by physical law. Given the physical state of the vocal tract, it is possible to predict the acoustic output using deterministic equations, which may be considered mappings from articulation to acoustic output. Further, these mappings can be inverted with optimization procedures to predict articulatory configuration from acoustics, possibly in a non-unique way.

Articulatory-acoustic models, based on physics, as simple as four-tube models (e.g., Fant, 1960; Badin *et al.*, 1990; Stevens, 1998; McGowan, 2006) and as sophisticated as an articulatory synthesizer (e.g., Mermelstein, 1973; Maeda, 1982, 1990) can be employed to understand the lawful variations between articulation and formant frequencies. As much as such models are useful for conceptual understanding of speech production, they are not direct measures of the articulatory-acoustic relations and are not sufficient for a complete understanding of human articulatory behavior and output acoustics.

An empirical approach is taken here to determine mappings between articulation and output acoustics during vowel production. This approach has the virtues of being based on actual human behavior, of not relying on simplified models of the vocal tract acoustics, and of not relying on published parameters based on measurements of various individuals for which simultaneous articulatory-acoustic data have not been obtained. The empirical approach in this paper relies solely on a corpus of simultaneous acoustic recordings and articulatory measurements.

Several techniques may be used to generate continuous articulatory measurements. These include flesh point measurements such as electromagnetic articulography and X-ray microbeam, and imaging techniques such as magnetic resonance imaging (MRI) and ultrasound. Three-dimensional (3D) MRI (e.g., Engwall and Badin, 1999) can be employed to generate detailed vocal tract shapes that provide much of the information necessary to determine the output acoustics, although additional properties would need to be provided before the output could be accounted for completely, such as vocal tract wall impedance, nasal tract properties, glottal configuration, and sub-glottal properties.

Instead of using three-dimensional imaging by MRI, less comprehensive articulatory data in the form of flesh point data are commonly employed (e.g., Kiritani, 1986; Perkell *et al.*, 1992). One reason for this is the fact that flesh point data are most often recorded simultaneously with acoustic output, whereas MRI data are generally not recorded simultaneously with acoustic output (see, however, Bresch *et al.*, 2006). Further, MRI technology has relatively low temporal resolution, whereas flesh point technologies are faster. This means that the MRI data sets are generally small, but with

[a)]Author to whom correspondence should be addressed. Electronic mail: rsmcgowan@cressllc.net
[b)]Present address: The Centre for Speech Technology Research, University of Edinburgh, Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom.

point measure technologies the data sets can be larger and contain more natural running speech. In the present work, the point measures in the University of Wisconsin X-ray Microbeam Speech Production Database (XRMB-SPD) (Westbury, 1994) were chosen because of the magnitude of the data set and variety of utterances, and the fact that acoustic signals were recorded simultaneously.

Because the XRMB-SPD only tracks midsagittal flesh points, the articulatory data are impoverished in the sense that acoustic output cannot be determined from these data using the physical theory of acoustics. Midsagittal shape does not determine an area function and output acoustics. Furthermore, even the midsagittal shape of the vocal tract is not completely measured in the XRMB-SPD: the most posterior pellet is located on the tongue dorsum. While it is possible to infer aspects of the midsagittal shape from the points that are measured, such as pharyngeal cross dimension (Whalen *et al.*, 1999; Jackson and McGowan, 2008), other aspects are not determined by the point measures, such as larynx height. On the other hand, it is generally assumed that there is a regular, if non-unique, relationship between midsagittal configuration and area function, and hence formant frequencies.

The intent of the present work is to construct mappings—forward mappings from articulation to acoustics, and inverse mappings from acoustics to articulation—for individual speakers in the XRMB-SPD. The forward mappings are empirically determined analogs to Fant's nomograms, which were derived from mathematical tube models (Fant, 1960, pp. 76–77 and 82–84). Instead of tube lengths and areas, variables derived from flesh point positions will serve as independent variables in the forward mappings in the current work. The nomograms tell us the specific response of formant frequencies to changes in tube parameters, and thus, the variations in sensitivity of the acoustic output to changes in tube geometry. These sensitivities (i.e., magnitudes of slopes in the nomograms) vary across the tube parameter space because the mapping between tube geometry and formant frequency is non-linear. Non-linearity, or changes in sensitivity, can be expected in the mappings between articulatory parameters derived from flesh point data and acoustic output. Thus, one of the goals of this study is to quantify the sensitivity of acoustic parameters to changes in articulatory parameters. This is an important consideration in phonetics, as, for instance, in Stevens' quantal theory (Stevens, 1972, 1989; Wood, 1979).

Another major area of research in the speech sciences is in the speech inverse problem: inferring articulatory information from speech acoustics in an algorithmic manner (e.g., Atal *et al.*, 1978; McGowan and Cushing, 1999; Hogden *et al.*, 2007). The inverse mappings derived in the present work provide a data-driven model to predict acoustics from articulation. They also provide a means of checking the predictions of models that are not derived from simultaneously recorded articulatory and acoustic data.

Our choices in methodology for building empirical mappings are informed by previous related work in speech and mathematical statistics. The two most important parts of the methodology are (1) a method for ordering the importance of independent variables, both acoustic and articulatory, and reducing the number of articulatory degrees-of-freedom from Cartesian pellet coordinates to some smaller, but essential, number; and (2) a least-squares method for approximating a smooth mapping between articulation and acoustics given simultaneously recorded articulatory and acoustic data points. The first element can be a form of function or data decomposition, and the latter element is a form of mathematical regression. The following is a review of previous work on articulatory-acoustic relations that employ one or both of these elements.

Principal components analysis (PCA), factor analysis, or other forms of function or data decomposition in the articulatory domain have become widely used when mapping between articulatory parameters and acoustics (Mermelstein, 1967; Shirai and Honda, 1976; Ladefoged *et al.*, 1978; Maeda, 1990; Yehia and Itakura, 1996; Story and Titze, 1998; Mokhtari *et al.*, 2007; and Story, 2007). These analyses reduce the number of articulatory parameters from an initially large number. This is particularly important when there are a limited number of acoustic parameters, such as three formant frequencies.

Mermelstein (1967) proposed using a Fourier cosine expansion of the log-area function in a largely theoretical study of the relation between area function and formant frequencies. He concluded that if the log-area function is spatially band-limited, then a unique area function can be determined from admittance function poles and zeros. However, while the poles correspond to resonance frequencies when the mouth is open, the zeros correspond to resonance when the vocal tract is closed. The latter are not observable during speech production. Building on Mermelstein's (1967) study, Yehia and Itakura (1996) decomposed the log-area function with a Fourier cosine series. In their mapping from formant frequencies to area functions they employed morphological and "least effort" constraints to alleviate ambiguities in the mapping. They tested their method for inferring area function from formant frequencies on data derived from X-rays of one speaker's 12 French vowels.

Shirai and Honda (1976) measured articulatory parameters taken from X-ray cineradiography of a speaker of Japanese, such as tongue shape, lip position, and jaw angle. The tongue shape was decomposed with PCA, and they related the first two PCA components and other articulatory measures to the first two formant frequencies. They were able to approximately recover the articulation of vowels from the first two formant frequencies using a non-linear regression technique, where the mapping was fitted on a set of 300 simultaneous articulatory and acoustic data frames of the speaker.

Ladefoged *et al.* (1978) used parallel factor analysis (PARAFAC) to decompose two-dimensional tongue shapes measured from X-rays of five speakers' pronunciation of ten American English vowels. They extracted two components for tongue position in the middle of vowels—front raising and back raising—and went on to use multiple regression to specify the two factors in terms of ratios among pairs of the three formant frequencies. Reasonable tongue positions could be inferred from formant frequencies using this

method. The study evaluated the predicted tongue positions in terms of correlations with original midsagittal shapes and in terms of root-mean-square error (RMSE).

Maeda (1990) took the approach of subtracting off important factors in articulatory movement, such as jaw movement, before performing PCA on tongue movement. He termed this "arbitrary factor analysis." In this way, it was easier to assign specific articulatory movements to changes in observed acoustic output than it would have been had PCA been applied to all the data without factoring out certain articulatory movement.

Story and Titze (1998) measured area functions of a single speaker's American English vowels with MRI and decomposed them into principal components, or what they termed "empirical orthogonal modes." They were able to obtain a mapping between the two modes that accounted for most of the variance in the area function and the first two formant frequencies in the form of a two-dimensional grid of iso-coefficient curves (coefficients of the two area function modes) in the formant plane. While ten points in the grid were determined from human empirical data, the remaining data, with 2500 grid intersections, were determined from area function modes by acoustic perturbation theory (Schroeder, 1967).

The recent work of Mokhtari *et al.* (2007) used human subjects for MRI full-volume scans during Japanese vowel production. Because of the noise of the machine, the acoustic recordings were taken separately from the MRI imaging. Also, the linear regressions between formant frequencies and principal components of human area functions were based on less than 40 samples of transitions between the vowels.

Story (2007) used pellet data from four speakers in the XRMB-SPD producing both static vowels and vowel-to-vowel transitions to find two principal components of the cross distances in the front of the mouth of each talker. "Cross distances" are the distances between, say, the tongue and the palate in the midsagittal plane. The amount of data per speaker was greater than the data used by any of Ladefoged *et al.* (1978), Story and Titze (1998), or Mokhtari *et al.* (2007), and it was shown that two PCA components were sufficient to characterize the data set. These components could be mapped to the first two formant frequencies in a largely unambiguous manner. However, the formants were calculated from a normalized area function and the two articulatory PCA components were not directly mapped to the formant data. Further, even this data set is limited and does not account for the sonorant portions of consonant-vowel transitions.

In order to obtain a robust mapping for each individual speaker, between 10 000 and 20 000 data points per speaker for four speakers were taken for the present work. These were simultaneous XRMB-SPD pellet positions and speech acoustic data that included all portions of vowels, including consonant-vowel transitions. PCA was chosen as the method of data decomposition, providing a set of independent variables ordered by the amount of variance accounted for in the data. Thus PCA can reduce the number of independent variables when the higher order components are neglected. For global linear regression, PCA alleviates problems in partial correlation in the independent variables.

Some of the work reviewed above relates articulatory and acoustic parameters using regression (e.g., Shirai and Honda, 1976; Ladefoged *et al.*, 1978; Mokhtari *et al.*, 2007). A form of regression was used in the present work to map both from articulatory coordinates to acoustic coordinates and vice-versa. The regression technique employed here is an adaptation of a method known as locally weighted regression, or "loess" (Cleveland, 1979; Cleveland and Devlin, 1988; Cleveland *et al.*, 1988), which produces a regression that is locally linear but globally non-parametric and non-linear. [Both Mermelstein (1967) and Yehia and Itakua (1996) used the property of local linearity between acoustics and articulation to obtain the inverse mapping.] In standard loess, least-squares regression is performed many times—indeed, for every point in the independent variable space at which we want to evaluate the mapping. For each point being evaluated the data are weighted differently, where the weight assigned to each data point is inversely related to its Euclidean distance to the point under evaluation. This provides a rational method for performing regression when the form of the function relating the independent and dependent variables is not specified, but the variation in local regression parameters can be presumed to be smooth.

The remainder of this paper is as follows. Section II details the procedures of speech segmentation, formant analysis, PCA applied to both pellet and formant data for each speaker, and loess. The loess method is described in some detail because it is novel to studies of speech production. Its particular implementation for constructing forward and inverse mappings, including optimization of loess parameters, is outlined. Section III presents the results of these analyses, starting with the articulatory principal components and their relation to well-known articulatory degrees-of-freedom. The rest of the results pertain to the optimum forward and inverse mappings found for each speaker, and examples of these mappings are shown. The causes of error and the differences in error between the forward and inverse mappings are examined. The sensitivities of formant frequencies to articulatory parameter changes are presented. Section IV provides a discussion focusing on the articulatory PCA, error, and sensitivity before final conclusions are drawn in Sec. V.

II. METHOD

A. Data

Simultaneously recorded acoustic and articulatory data from the XRMB-SPD (Westbury, 1994) were used. The database consists of time-aligned audio and midsagittal pellet position recordings from 57 adult American English speakers, comprising about 15 h of recorded speech. Speech tasks include reading of citation words, sentences, paragraphs, number sequences, and vowel sequences, with some tasks performed at deliberately slower or faster speaking rates.

Pellet coordinates are referenced to a speaker-specific Cartesian coordinate system whose axes are based on ana-

tomical features in the speaker's head. [See Westbury (1994) for the definition of the coordinate system.] Each speaker's time-varying articulatory data contains horizontal and vertical coordinates in the midsagittal plane for each of eight moving pellets. There are four pellets approximately evenly spaced along the tongue centerline; the most anterior tongue pellet is close to the tongue tip and the most posterior is the furthest back the pellet could be placed without inducing a gag reflex. The other four pellets include one pellet on the upper lip and one on the lower lip, and two on the mandible. These pellets give a partial representation of the vocal tract; no information is available about lateral tongue or jaw movement, the velum, or the posterior vocal tract (e.g., pharyngeal dimensions or larynx height), factors that also affect acoustic output. Pellet positions were originally measured at variable rates between 20 and 160 samples/s, but in the database all pellet positions are resampled at an equal rate of 160 samples/s (Westbury, 1994, p. 57).

In the present study only the positions of the six pellets on the tongue and lips were used, resulting in a total of 12 degrees of freedom. We examined four speakers, two males (JW11 and JW18) and two females (JW14 and JW16).

B. Segmentation

All vowel tokens were included in the analysis except those adjacent to nasal, lateral, or rhotic consonants. Also, very short reduced vowels that had little or no formant structure were excluded. Vowels in the context of nasals and laterals were excluded because velar or lateral movement during these vowels would not be captured by any pellet; hence the acoustic changes would not be matched by any articulatory changes. Vowels in rhotic contexts were excluded because the relatively rare but extreme retroflex articulations could potentially complicate reduction in the dimensionality of the articulatory data as well as the acoustic-articulatory regression analysis.

The audio recordings were manually annotated to segment and label the vowels not excluded by the criteria above using PRAAT TextGrids (Boersma and Weenink, 2007). The acoustic and articulatory data within the demarcated vowel intervals formed the basis of the study. The following conventions were used for the placement of interval boundaries. Between the vowel and a preceding obstruent, the boundary was generally placed at the first glottal pulse after the closure release. Similarly, between the vowel and a following obstruent, the boundary was placed at the last glottal pulse before the onset of closure. For the less straightforward case of dividing the vowel from an adjacent oral sonorant, we sought the midpoint of the transition between the two sounds based on auditory judgments and visible formant transitions.

C. Formant tracking, editing, and data pruning

After defining the vowel intervals to be used in the analysis, automated formant tracking was performed on the vowels using PRAAT. The formant analysis was linear predictive coding (LPC)-based using the Burg algorithm. A 25 ms window was used and the centers of neighboring windows were 6.25 ms apart so that the frame rate was the same as the pellet sampling rate of 160 Hz. The maximum number of formants that were identified per frame was 5, with a maximum frequency of 5000 Hz for males and 5500 Hz for females. Pre-emphasis was applied to frequencies above 50 Hz. Only $F1$, $F2$, and $F3$ values were used from the acoustic analysis. The formant tracks were resampled to align acoustic samples with the articulatory samples.

To ensure the best quality formant tracking, the $F1$, $F2$, and $F3$ formant tracks within all of the vowel segments were visually inspected and manually corrected at each sample point where the automatic tracking was deemed faulty. To make this process efficient we developed a graphical formant editor that allows point-and-click corrections to formant values in a dual spectrogram/spectrum display. The spectrum display includes a function for refining an estimated formant frequency by fitting a parabola to the three harmonics closest to the estimate and shifting the formant frequency to the peak of the parabola.

Once the formant analysis of the vowel segments was complete, the samples from all of the segments were pooled for each speaker. Each sample comprised a 3-dimensional acoustic vector and a 12-dimensional articulatory vector. Before continuing in the analysis, each articulatory-acoustic data point was examined automatically for completeness. Any sample missing acoustic data ($F1$, $F2$, or $F3$) or pellet data (any of the two coordinates for the six pellets) was excluded from further analysis. Table I summarizes the data included in the study: the gender of each speaker, the number of vowel segments used, and the number of samples drawn from these segments. The subsequent analyses (PCA and regression) were performed separately on each speaker's data set. Data from different speakers were not pooled. Although multiple data points in a speaker's data set may originate from the same time series, i.e., a vowel trajectory, it is assumed that the error distribution for each measured quantity is statistically independent from others over the same time series.

TABLE I. Gender and the amount of usable data collected for each speaker.

	Gender	No. of segments	No. of samples
JW11	M	763	13 003
JW14	F	828	16 013
JW16	F	974	19 770
JW18	M	666	11 295

D. PCA

For each speaker, PCA was performed twice: once on the speaker's articulatory data points and once on the acoustic data points. The principal components of the articulatory data are denoted $K1, K2, K3, \ldots, K12$, and those of the formant data are denoted $A1$, $A2$ and $A3$. The lower-order principal components were utilized as variables in the subsequent regression analysis: the first four articulatory components $K1-K4$, and all three acoustic components.

E. Locally weighted regression (loess)

1. General method

For each speaker's data set we computed both forward (articulatory to acoustic) and inverse (acoustic to articulatory) mappings using locally weighted regression. Locally weighted regression or loess (Cleveland, 1979; Cleveland and Devlin, 1988; Cleveland *et al.*, 1988) is a form of nonparametric regression by data smoothing. It is computationally intensive but allows one to represent the relationship between one or more independent variables and a dependent variable with few assumptions about the form of that relationship. This makes it suitable for fitting complex regression surfaces for which a suitable parametric function is not known. The general method of loess is now described.

Let $S=\{(\mathbf{x}_i,y_i):i=1\ldots n\}$ be a given set of data points, where y_i is a measurement of the dependent variable and $\mathbf{x}_i$ is a measurement of a *p-tuple* of independent variables. A regression model relating these variables is

$$y_i = g(\mathbf{x}_i) + \varepsilon_i, \tag{1}$$

where ε_i is a zero-mean, normal random error term. In classical regression analysis, g is assumed to belong to some parametric class of functions, such as polynomials, which places practical limits on the variety of surfaces that can be modeled. For example, in the case that a linear relation is assumed to hold between the independent variable $\mathbf{x}$ and dependent variable y, the function g can be estimated in a least-squares sense by $\hat{g}$,

$$y \approx \hat{g}(\mathbf{x}) = \mathbf{x}\cdot\boldsymbol{\beta} + \gamma, \tag{2}$$

where parameters $\boldsymbol{\beta}$ and γ are determined by the data points $(\mathbf{x}_i,y_i)$ and a chosen weighting function of those points, say, w_i (e.g., Chaterjee and Hadi, 1988).

In loess, g is not limited to being a parametric function; it is only assumed to be a smooth function of $\mathbf{x}$. Accordingly, the estimate of g, $\hat{g}$, is computed without fitting a parametric function to the entire data set. Rather the smoothness property of g is exploited to estimate g by locally fitted functions. The smoothness property allows us to assume that for any point $\mathbf{x}$ in the space of the independent variables, $\hat{g}(\mathbf{x}')\approx l_{\mathbf{x}}(\mathbf{x}')$ for $\mathbf{x}'$ near $\mathbf{x}$, where $l_{\mathbf{x}}$ is a locally fitted, low-order polynomial. Strict equality holds for $\mathbf{x}'=\mathbf{x}$; that is, $\hat{g}(\mathbf{x})=l_{\mathbf{x}}(\mathbf{x})$. $l_{\mathbf{x}}$ is obtained by a least-squares fit to the data based on a local weighting function $w_i(\mathbf{x})$ that heavily weights data points $(\mathbf{x}_i,y_i)$ close to $\mathbf{x}$. The locally weighted regression function $l_{\mathbf{x}}$ may be linear or non-linear; in the present study a linear model was used:

$$l_{\mathbf{x}}(\mathbf{x}') = \mathbf{x}'\cdot\boldsymbol{\beta}(\mathbf{x}) + \gamma(\mathbf{x}), \tag{3}$$

where $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$ are computed just as in standard linear least squares, except that they now depend on $\mathbf{x}$ because the weighting function depends on $\mathbf{x}$. Consequently, the least-squares procedure must be repeated at each value of $\mathbf{x}$ for which we want to solve $\hat{g}(\mathbf{x})$, which makes loess computationally intensive.

To perform loess, one must choose a weight function $w_i(\mathbf{x})$ that assigns weights to data points in S based on distance from $\mathbf{x}$: data points close to $\mathbf{x}$ have large weight, while those far from $\mathbf{x}$ have small weight. A distance metric is also needed. For the distance metric, this study used Euclidean distance in the space of the independent variables, after first scaling the variables by dividing each by its own standard deviation. The weight function was the standard one used by Cleveland (1979), which guarantees a fixed neighborhood size (number of positively weighted data points) regardless of data distribution around $\mathbf{x}$. The weight function has a parameter b between 0 and 1, known as the bandwidth, that expresses neighborhood size as a proportion of the data. Thus, the larger the value of b, the more data points influence the local regression at $\mathbf{x}$. Using a nearest-neighbor algorithm, the neighborhood size is used to determine a neighborhood radius. The radius of the neighborhood of $\mathbf{x}$ for a given bandwidth b, denoted $d(\mathbf{x},b)$, is defined to be the distance from $\mathbf{x}$ to the qth nearest data point, where q is equal to bn rounded to the nearest integer. The weight function is zero beyond this radius. The weight assigned to data point $(\mathbf{x}_i,y_i)$ for the locally weighted regression at $\mathbf{x}$, using bandwidth b is

$$w_i(\mathbf{x},b) = W\left(\frac{\|\mathbf{x}_i-\mathbf{x}\|}{d(\mathbf{x},b)}\right), \tag{4}$$

in which W is the tricube function: $W(u)=(1-u^3)^3$ for $0\leqslant u\leqslant 1$ and $W(u)=0$ otherwise.

2. Computationally efficient loess

In loess, weighted least-squares estimation must be performed for every value of $\mathbf{x}$ at which we want to know $\hat{g}(\mathbf{x})$. This makes the technique computationally expensive for operations requiring many samples, such as plotting. An efficient alternative is to pre-compute $\hat{g}(\mathbf{x})$ at sample values of $\mathbf{x}$, and then interpolate for intermediate values. For the sample set one may choose the original data points, or some strategically selected set of points, such as the vertices of a kd-tree constructed on the data (Cleveland *et al.*, 1988). In the present study the pre-computed sample set was simply the data points. However, rather than directly interpolating the dependent variable at the sample points, better results were obtained by interpolating the local regression parameters computed at those points and generating the dependent variable from the interpolated regression parameters.

Following this approach, a loess model in the current study was fitted to data set S by computing a locally weighted regression at each data point. The resulting model then included one set of local regression parameters for each data point, accompanied by an interpolation scheme, which was as follows. For each data point $(\mathbf{x}_i,y_i\in S)$, let $\bar{\boldsymbol{\beta}}(\mathbf{x}_i)$ and $\bar{\gamma}(\mathbf{x}_i)$ be the pre-computed local regression parameters fitted to the neighborhood of $\mathbf{x}_i$. The regression parameters $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$ at a novel point $\mathbf{x}$ were estimated by a weighted average of the parameters $\bar{\boldsymbol{\beta}}(\mathbf{x}_i)$ and $\bar{\gamma}(\mathbf{x}_i)$ at the data points, where the data points were weighted using the tricube-based weight function $w_i(\mathbf{x},b)$ defined above. The interpolated regression parameters $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$ at $\mathbf{x}$ were thus given as

$$\boldsymbol{\beta}(\mathbf{x}) = \frac{\sum_{i=1}^{n} w_i(\mathbf{x},b)\bar{\boldsymbol{\beta}}(\mathbf{x}_i)}{\sum_{i=1}^{n} w_i(\mathbf{x},b)}, \quad \gamma(\mathbf{x}) = \frac{\sum_{i=1}^{n} w_i(\mathbf{x},b)\bar{\gamma}(\mathbf{x}_i)}{\sum_{i=1}^{n} w_i(\mathbf{x},b)}. \tag{5}$$

The predicted value $\hat{g}(\mathbf{x})$ was then calculated using the interpolated parameters $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$. Note that the bandwidth b used to weight data points for interpolation may differ from that used for the weighted least-squares fit of the local regressions. Thus each loess model constructed in this study had two distinct bandwidth parameters: a *regression bandwidth* b_R, which was used to weight the data points for the local regressions, and an *interpolation bandwidth* b_I, which was used to interpolate the local regression parameters. It will be assumed henceforth that the loess models in this study generate output $\hat{g}(\mathbf{x})$ by interpolating the precomputed regression parameters at the data points in this manner.

3. Grid sampling

Given a loess model $\hat{g}(\mathbf{x})$ with regression parameters $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$, it is useful to be able to sample these functions in a regular grid over the space of the independent variables. Grid sampling is useful both for generating plots of the regression surfaces and for studying the variation in the regression parameters. The *full grid* associated with a particular data set S has d evenly spaced columns of vertices in each independent variable, generally with $d=20$. For example, a data set with four independent variables will have a $20\times20\times20\times20$ full grid. The span of the grid in each dimension is from -2.5 to 2.5 standard deviations from the mean, using the standard deviation and mean of that dimension in S. A *subgrid* associated with a particular data set is the same as the full grid but missing one or more coordinates of each vertex. Thus it occupies a subspace of the independent variable space. Subgrids are useful for reducing data complexity and for generating 3D plots as a function of the first two independent variables.

Functions are defined for sampling any function of the independent variables at grid vertices. If $f(\mathbf{x})$ is a function of the independent variables, then $[f(\mathbf{x})]=f(\mathbf{x})$ for any $\mathbf{x}$ that is one of the vertices of the full grid. To sample at the vertices of a subgrid, the dimensions that are present are indicated by a superscript; for example, superscript (1,2) indicates that the subgrid occupies only the first two dimensions of the full grid, and values are averaged over the missing dimensions. (The absence of any superscript indicates sampling on the full grid.) Thus, for any vertex $\mathbf{z}$ of the subgrid, $[f(\mathbf{z})]^{(1,2)}$ is equal to the average $[f(\mathbf{x})]$ over all vertices $\mathbf{x}$ of the full grid that agree with $\mathbf{z}$ in the first and second coordinates. The grid sampling functions may be used to sample $\hat{g}(\mathbf{x})$, $\boldsymbol{\beta}(\mathbf{x})$, and $\gamma(\mathbf{x})$ at grid vertices.

4. Model evaluation

Given a loess model $\hat{g}$ fit to data set S, which we may call the fitting set, the performance of the model was evaluated using a completely separate, randomly selected set of data that was held out from the fitting set, called a test set $T=\{(\mathbf{x}_j,y_j):j=1,\ldots,m\}$. The loess model is fitted to the fitting set, without using any data from the test set. The test set is used to evaluate the prediction error of the model using RMSE:

$$\mathrm{RMSE}(\hat{g},T) = \sqrt{\frac{\sum_{j=1}^{m}(\hat{g}(\mathbf{x}_j)-y_j)^2}{m}}. \tag{6}$$

5. Optimum model selection

Of course, prior to testing, the loess model has to be selected by choosing appropriate values for the regression bandwidth b_R and interpolation bandwidth b_I. Too large a value for either bandwidth parameter will generate too smooth a regression surface, while too small a value will lead to overfitting of the data. To optimize model parameters without overfitting to data, a standard technique is to select parameter values that optimize the prediction rate for a separate set of data that was not used during the construction of the model, called a *validation set*. Similar to a test set, a validation set is a randomly selected set external to the fitting set; thus it allows us to see which model best generalizes to new data. However, the validation set may not actually be part of the test set since it is used to select the model parameters and is thus not "unseen" prior to testing of the selected model.

However, when the data pool is limited, it may be difficult to draw an adequately sized validation set separate from both fitting and test data. A common method for validating a model without having a separate validation set is k-fold cross-validation. In this method, the fitting set is randomly partitioned into k equal subsets, and the model is fitted to the data k times. For each fitting iteration i, the ith subset is held out as the validation set, and the remaining $k-1$ subsets are combined to form a reduced fitting set; validation error is equal to RMSE in Eq. (6) calculated over the data in the ith subset. Thus, each data point gets used for validation once and for fitting $k-1$ times. Validation error is averaged over the k trials.

k-fold cross-validation was employed on the fitting set to optimize the two bandwidth parameters b_R and b_I. An exhaustive search of the two-dimensional parameter space was conducted, and for each pair of bandwidth values in the search, validation error was computed using k-fold cross-validation with $k=10$. The regression and interpolation bandwidths yielding the minimum validation error were deemed the optimal bandwidths. The space of bandwidth values to search was determined empirically. If a minimum was not attained in a given search space, then the search space was shifted in the direction of decreasing validation error observed in previous searches. Figure 1 shows the results of a typical search for optimal bandwidths. The loess model being optimized is that of the inverse mapping from the acoustic variables $A1$, $A2$, and $A3$ to the articulatory variable $K1$, for subject JW16. The search space consisted of 15 values of regression bandwidth between 0.002 and 0.03, and 20 values of interpolation bandwidth between 0.0001 and 0.002 for a

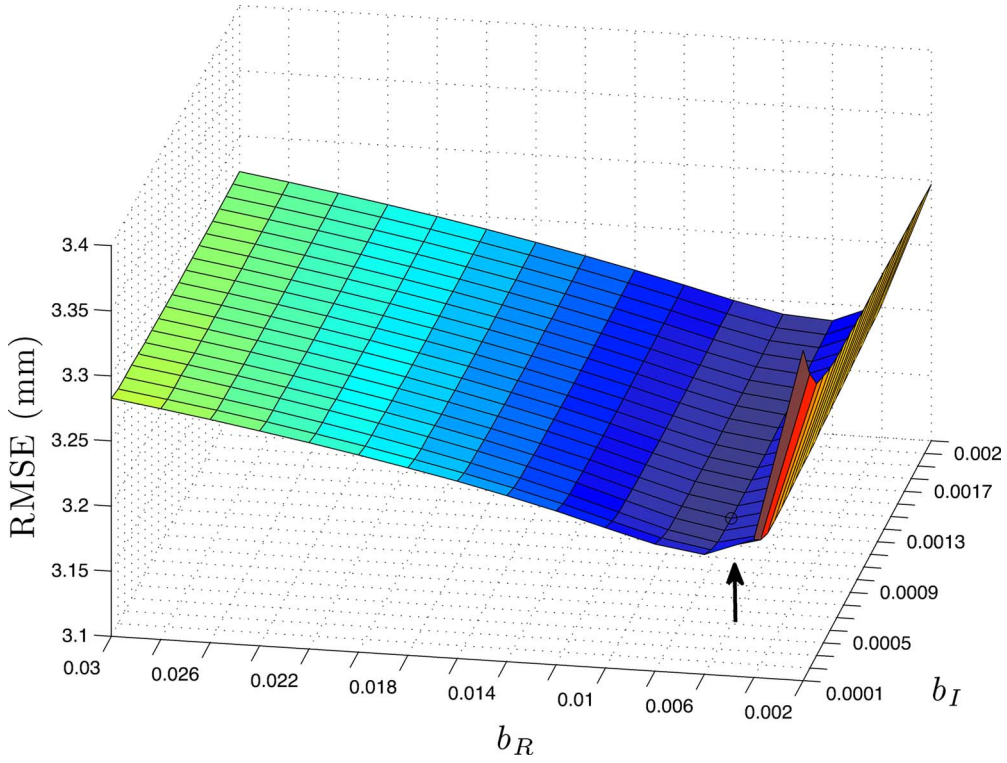


FIG. 1. (Color online) RMSE surface as a function of regression bandwidth b_R and interpolation bandwidth b_I. The loess model being optimized is that of the inverse mapping from the acoustic variables $A1$, $A2$, and $A3$ to the articulatory variable $K1$ for subject JW16. The arrow indicates the location of the minimum RMSE.

total of 300 different trials. The minimum validation RMSE of 3.181 mm was found at b_R=0.006, b_I=0.0005 (indicated by the arrow).

Once the optimal bandwidth values were found for a particular mapping, the loess model was constructed on the entire fitting set using the optimal b_R value. Then the optimal b_I value was used in the equations in Eq. (5).

It should be emphasized that in order to maintain an objective evaluation, the test data T may not be used in optimizing the model parameters. Thus k-fold cross-validation was performed only using data from the fitting set S. The validation RMSE values have no relation to the test RMSE mentioned in Sec. II E 4, which is determined only after a model has been selected and fitted.

6. Acoustic-articulatory loess

To construct loess models, each speaker's data set was divided into a fitting set S and a test set T. The test set comprised a random selection of 10% of the data points of the speaker. After discovering the optimal b_R and b_I values for a given model, as described in Sec. II E 5, it was fitted to the fitting set and evaluated against the test set in the manner described above (see Secs. II E 2 and II E 4).

For each speaker, seven loess models were constructed: three forward mappings from the articulatory PCA components $\mathbf{x}=(K1,K2,K3,K4)$ to each of the formants $y=F1$, $y=F2$, or $y=F3$, and four inverse mappings from the acoustic PCA components $\mathbf{x}=(A1,A2,A3)$ to each of the articulatory components $y=K1$, $y=K2$, $y=K3$, or $y=K4$. These totaled seven loess mappings for each of the four speakers: three forward mappings and four inverse mappings. Each model had the optimum bandwidths.

The various loess models generated for each subject are denoted as follows. In the case of a forward mapping from $\mathbf{x}=(K1,K2,K3,K4)$ to say, $y=Fj$, the loess model is denoted $\hat{g}_j^{\rightarrow}$. In the case of an inverse mapping from $\mathbf{x}=(A1,A2,A3)$ to $y=Kj$, the loess model is denoted $\hat{g}_j^{\leftarrow}$.

As shown in Eq. (3), for each locally weighted regression $\hat{g}(\mathbf{x})$ there are functions $\boldsymbol{\beta}(\mathbf{x})$ and $\gamma(\mathbf{x})$, where $\boldsymbol{\beta}(\mathbf{x})$ is a vector of slopes varying as a function of $\mathbf{x}$, and $\gamma(\mathbf{x})$ is the regression constant. $\boldsymbol{\beta}(\mathbf{x})$ will be denoted for each mapping as follows. In the case of a forward mapping, $\hat{g}_j^{\rightarrow}$, the vector of slopes is denoted $\boldsymbol{\beta}_j^{\rightarrow}=(\beta_{1j}^{\rightarrow},\beta_{2j}^{\rightarrow},\beta_{3j}^{\rightarrow},\beta_{4j}^{\rightarrow})$, where $\beta_{ij}^{\rightarrow}$ is the regression slope from Ki to Fj. While all of these quantities depend on $\mathbf{x}$, the argument has been suppressed, as it will be for the rest of the paper. The constant is denoted $\gamma_j^{\rightarrow}$. In the case of an inverse mapping $\hat{g}_j^{\leftarrow}$, $\boldsymbol{\beta}(\mathbf{x})$ is a three-dimensional vector: $\boldsymbol{\beta}_j^{\leftarrow}=(\beta_{1j}^{\leftarrow},\beta_{2j}^{\leftarrow},\beta_{3j}^{\leftarrow})$, where $\beta_{ij}^{\leftarrow}$ is the regression slope from Ai to Kj, which again depends on $\mathbf{x}$. The constant is denoted $\gamma_j^{\leftarrow}$.

The same notation is employed for the pre-computed regression parameters at the data points. Thus $\bar{\boldsymbol{\beta}}(\boldsymbol{x}_i)$ is denoted $\bar{\boldsymbol{\beta}}_j^{\rightarrow}=(\bar{\beta}_{1j}^{\rightarrow},\bar{\beta}_{2j}^{\rightarrow},\bar{\beta}_{3j}^{\rightarrow},\bar{\beta}_{4j}^{\rightarrow})$ in the case of a forward mapping and $\bar{\boldsymbol{\beta}}_j^{\leftarrow}=(\bar{\beta}_{1j}^{\leftarrow},\bar{\beta}_{2j}^{\leftarrow},\bar{\beta}_{3j}^{\leftarrow})$in the case of an inverse mapping. Similarly, $\bar{\gamma}(\mathbf{x}_i)$ is denoted $\bar{\gamma}_j^{\rightarrow}$ for a forward mapping and $\bar{\gamma}_j^{\leftarrow}$ for an inverse mapping.

F. Global linear regression

To provide a baseline of performance, the loess models were compared to standard linear regressions. Thus for each of the seven mappings for each speaker, in addition to constructing a locally weighted regression on the basis of the speaker's fitting data, a single, uniformly weighted linear regression was computed by least-squares fit to the same fitting data. This will be called the "global regression" in contrast to the locally weighted regression. The global regressions were evaluated against the test data in the same manner as the local regression models.

G. Sensitivity

One application of the mappings constructed with loess is to study the sensitivity of variables in one domain to changes in the other domain. In particular, it is of interest to study the variation in the magnitudes of slopes as a function of position in the independent variable space. The magnitudes of the slopes of the forward mappings $\boldsymbol{\beta}_j^{\rightarrow}$ are direct measures of the sensitivity of formant frequencies to changes in the principal components of articulation as a function of the position in articulatory configuration space. The slopes of the forward mappings will be used to define *empirical sensitivity functions*.

To permit systematic examination, the slopes of the forward mappings were sampled in grid format (see Sec. II E 3). However, the resulting data are complex: in the full grid of the forward mapping, there are $20^4=160\,000$ vertices, each of which is evaluated for 12 different slope functions: $\beta_{11}^{\rightarrow}$, $\beta_{12}^{\rightarrow}$, $\beta_{13}^{\rightarrow}$, $\beta_{21}^{\rightarrow}$, $\beta_{22}^{\rightarrow}$, $\beta_{23}^{\rightarrow}$, $\beta_{31}^{\rightarrow}$, $\beta_{32}^{\rightarrow}$, $\beta_{33}^{\rightarrow}$, $\beta_{41}^{\rightarrow}$, $\beta_{42}^{\rightarrow}$, and $\beta_{43}^{\rightarrow}$. To focus on some of the more important aspects of sensitivity, the degrees of freedom of the slope data needed to be reduced. The reduction presented here is simply based on the observation that $K1$ and $K2$, on average, account for 81% of the variation in pellet positions for the four speakers with a range from 77% to 89%. Thus only slopes for which the independent variable was $K1$ or $K2$ were considered, i.e., $\beta_{ij}^{\rightarrow}$ with $i=1,2$, reducing the number of slopes by half. Furthermore, slopes were sampled on the $K1,K2$ subgrid, which has only $20^2=400$ vertices, while averaging over $K3$ and $K4$ coordinates, i.e., using the slope functions $[\beta_{ij}^{\rightarrow}]^{(1,2)}$.

Because we are concerned with magnitudes, the absolute values of slopes are examined here. In fact, for reasons that will become apparent in Sec. IV, it was convenient to consider these values normalized by the predicted formant frequencies. Thus the empirical sensitivity functions, evaluated at each vertex of the $K1,K2$ subgrid, are defined as

$$\nu_{ij}^{\rightarrow}=\frac{[|\beta_{ij}^{\rightarrow}|]^{(1,2)}}{[\hat{g}_j^{\rightarrow}]^{(1,2)}} \quad \text{for } i=1,2 \text{ and } j=1,2,3.$$

For each speaker and each empirical sensitivity function $\nu_{ij}^{\rightarrow}$, two categories of $(K1,K2)$ vertices over the 20×20 subgrid were defined: those with the largest 20% $\nu_{ij}^{\rightarrow}$ and those with the smallest 20% $\nu_{ij}^{\rightarrow}$. These categories were labeled large sensitivity and small sensitivity, respectively. Critical vertices were also saved for later examination. $(K1,K2)$ was a critical vertex for $\nu_{ij}^{\rightarrow}$ if its value of $\nu_{ij}^{\rightarrow}$ is smaller than that of any of the neighboring vertices, where two vertices in a rectangular grid are neighboring if they differ by at most one row or column, or both one row and column. (This criterion was more easily satisfied for the boundary points because there were fewer comparisons to satisfy.)

To provide reference markers for the sensitivity categories and critical points in the $K1,K2$ plane, point vowels were also located in the plane as follows. The formant frequencies and the pellet positions of sustained vowel productions [ɑ], [æ], [u], and [i] of the point vowels from task 14 of the XRMB-SPD were extracted for each of the speakers. (One exception was the [æ] of JW14, which was extracted from task 24.) Using the first two formants, a time sample was chosen for each of the vowels as being the most extreme moment in the articulation of the vowel. For instance, for [ɑ], a time where the first formant frequency was high and the second formant frequency was low, relative to the other values in the file, was chosen as the time that best represented the extreme of the articulation. The pellet positions at that time were projected onto $K1$ and $K2$ to obtain a representation of the vowel articulation in the $(K1,K2)$ coordinates.

TABLE II. Percent of variance accounted for by the first five PCA components of the articulatory data.

	$K1$ (%)	$K2$ (%)	$K3$ (%)	$K4$ (%)	$K5$ (%)
JW11	53.5	23.3	10.4	7.4	1.4
JW14	75.8	13.1	5.3	3.2	1.1
JW16	58.7	21.0	7.9	6.6	1.7
JW18	56.0	24.0	10.1	5.1	1.8

III. RESULTS

A. PCA

The first four components of the PCA of the articulatory pellet data accounted for between 94% and 97% of the variance for each of the four talkers. Table II shows the percentage of variance accounted for by each of the first five components.

Figure 2 illustrates the first four articulatory components for each speaker. As evident in the figure, the first articulatory PCA component, $K1$, corresponds to a low-back to high-front degree-of-freedom in all of the talkers, except for JW16. For JW18, the front-back component is minimal. Also, there appear to be varying amounts of tongue bunching in $K1$ for each of the subjects for the high-front position. $K1$ is associated with varying degrees of lower lip height changes, except in the case of JW14. JW16's $K1$ shows a tongue height degree-of-freedom, and there appears to be a "rocking" motion of the tongue in her $K1$, which has the effect of inducing a tongue blade up-down degree-of-freedom.

For JW11 and JW14 the second articulatory PCA component, $K2$, corresponds to a high-back to a low-front tongue degree-of-freedom, with more height change for JW14 than for JW11 (Fig. 2). There is also some tongue bunching-stretching associated with tongue height changes for JW14. Interestingly, the $K2$ of JW16 shows a low-back to high-front

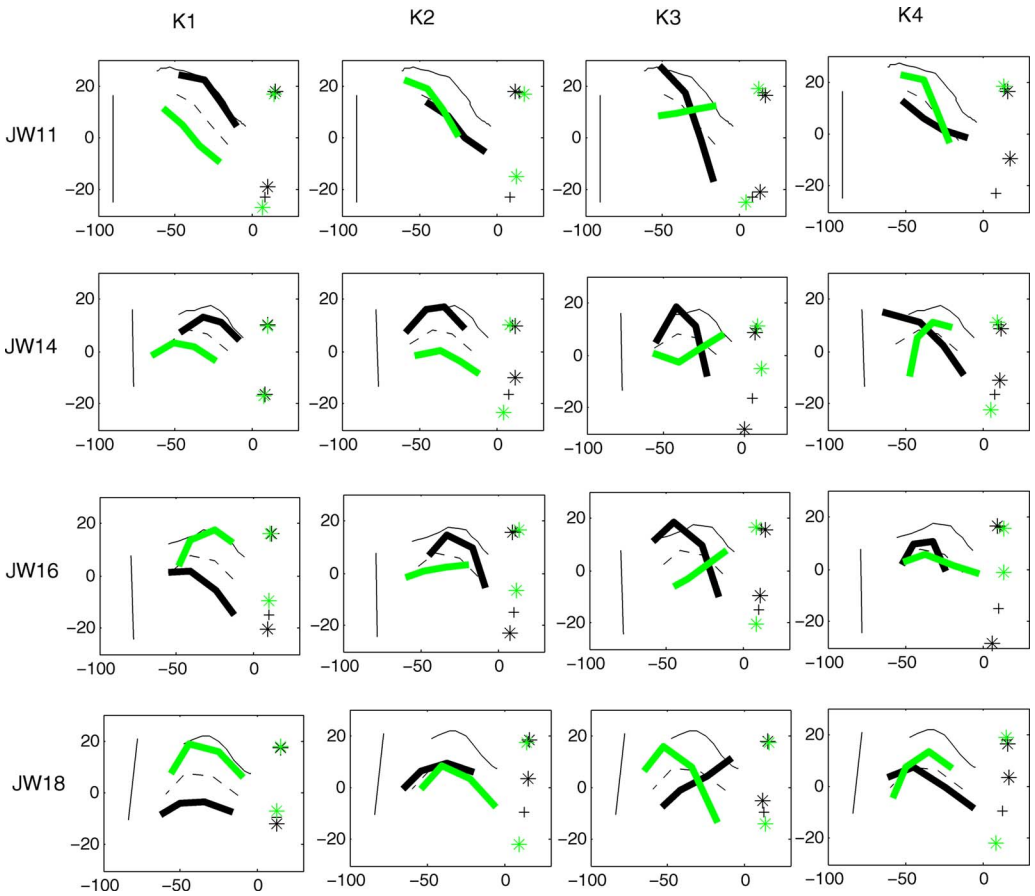


FIG. 2. (Color online) Graphical representation of the first four articulatory principal components. Shown are mean and two extreme values of each component, *K*1, *K*2, *K*3, and *K*4 (by column) for each of the four speakers, JW11, JW14, JW16, and JW18 (by row). The axes are in units of milimeters and correspond to the horizontal and vertical axes as defined in Westbury (1994). The top curve in each plot is the palate trace and the black vertical line is the approximate pharyngeal wall. The dotted curve and "+" marks represent the mean tongue and lip pellet coordinates for the principal component, while the bold curves and " *" marks represent the extreme values.

degree-of-freedom that the others show in their *K*1, with a substantial bunching-stretching degree-of-freedom. JW18's *K*2 exhibits a front-back degree-of-freedom, with some tongue rocking, again inducing a tongue blade up-down degree-of-freedom.

In *K*3, all the subjects show a movement from a high bunched tongue with the tongue blade down to a low stretched tongue with tongue blade up. Thus, *K*3 of all subjects contains simultaneous tongue blade up-down and tongue bunching-stretching degrees-of-freedom. There does not appear to be much consistency among the subjects in their *K*4 components, although bunching-stretching and tongue blade up-down degrees-of-freedom also appear as aspects of the *K*4 components.

There are differences among the subjects in the kinds of tongue shapes/positions and lip positions accounted for in the first four components. A portion of these differences can be accounted for in terms of differences in the placement of the pellets. For instance, many of the differences between JW11 and the other three subjects can be accounted for if the rearmost tongue pellet of JW11 can be considered to be equivalent to the second rearmost in the other subjects.

The relation between familiar articulatory degrees-of-freedom (such as high-low, front-back) and the PCA components can be quantified with an orthonormal set of vectors intended to represent these familiar degrees-of-freedom. These are (1) the tongue front-back degree-of-freedom, with equal weights (0.5) in the four tongue *x*-components and zeros in the four tongue *y*-components and the four lip components; (2) the tongue high-low degree-of-freedom, with equal weights (0.5) in the four tongue *y*-components and zeros in the four tongue *x*-components and the four lip components; (3) the lip-opening degree-of-freedom, with equal and opposite weights (–0.707 and 0.707) in the two lip *y*-components and zeros in the others; and (4) the lip-protrusion degree-of-freedom, with equal weights (0.707) in the two lip *x*-components and zeros elsewhere. The amount that each of the PCA components has of the tongue front-back, tongue high-low, lip-closure, and lip-protrusion degrees-of-freedom can be determined by projecting the PCA component vector onto each of the vectors associated with the familiar degrees-of-freedom, using the dot product. The proportion of the articulatory PCA components that can be attributed to each familiar degree-of-freedom are shown in Fig. 3. The signs within a speaker indicate the relative directions of these projections. For instance, *K*2 of JW11 associates a more low with a more front tongue position.

The tongue front-back movement is the familiar degree-of-freedom that is represented the most often in the PCA components at or above the 20% level. On the other hand,

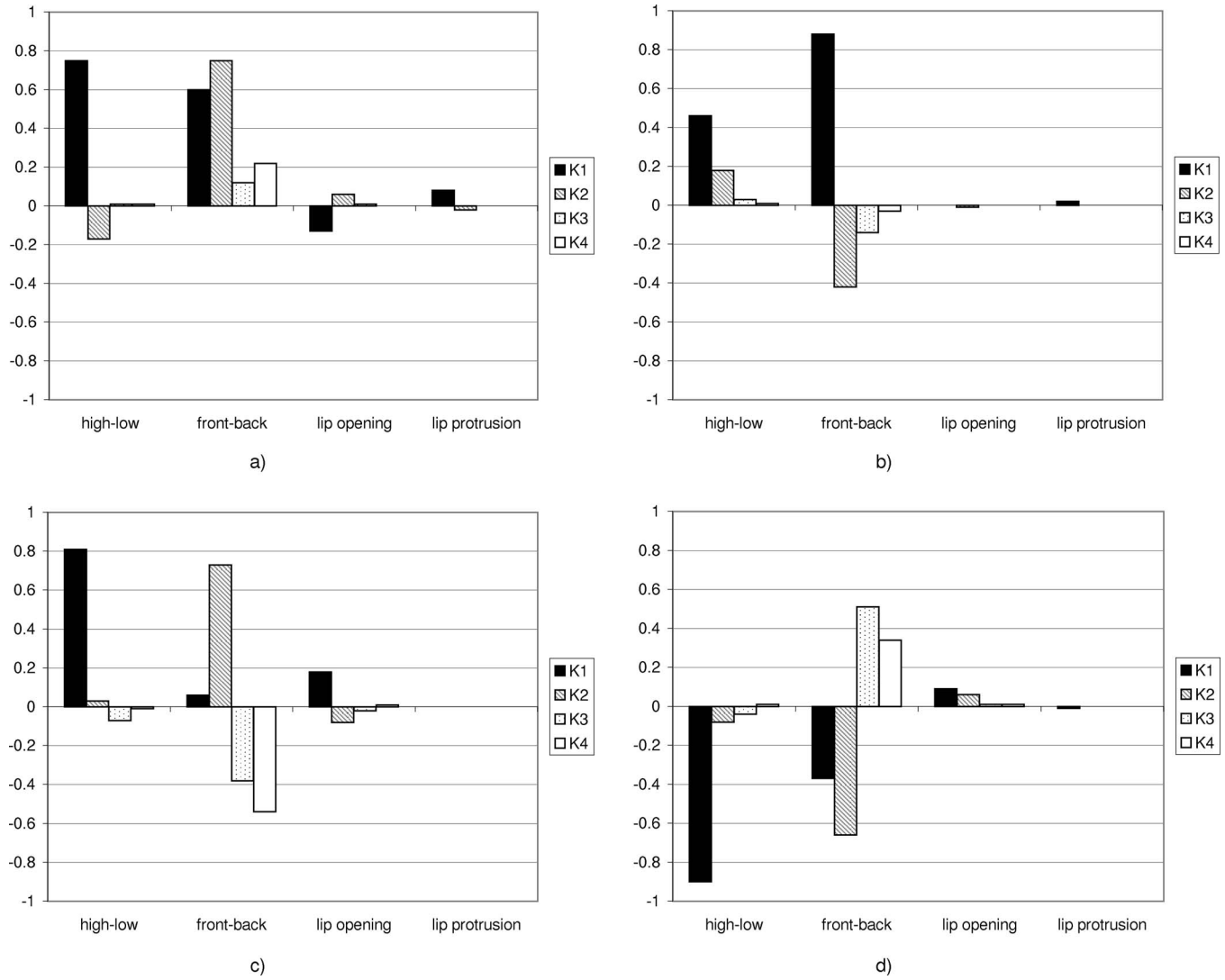


FIG. 3. The normalized projection of each articulatory component—$K1, K2, K3, K4$—onto each familiar degree-of-freedom—high-low, front-back, lip opening, and lip protrusion—for (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

the tongue high-low movement reaches the 20% level only in each speaker's $K1$. Lip protrusion and lip opening are not captured at the 20% level by any of the components for any of the speakers.

PCA analysis was performed on the acoustic data to generate a set of three orthonormal vectors that spanned the formant space. Table III shows the amount of variance accounted for by each of the PCA components, $A1$, $A2$, and $A3$. A detailed examination of the PCA results revealed that for all speakers, $A1$ contained a very large proportion of $F2$, and small amounts of $F1$ and $F3$, because $F2$ was the formant frequency with the largest variance. $F1$ and $F3$ had comparable variances in terms of absolute frequency and these formants were accounted for in $A2$ and $A3$.

The PCA components $A1$, $A2$, and $A3$ spanned the same space as the original formant frequencies $F1$, $F2$, and $F3$ and the two sets of basis vectors can be transformed from one to another with a linear, distance-preserving transformation. However, because the PCA components are orthogonal they can be used as independent variables in global linear regression analysis without the complication of partial correlations between them.

TABLE III. Percent of variance accounted for by the three acoustic PCA components.

	$A1$ (%)	$A2$ (%)	$A3$ (%)
JW11	81.0	13.4	5.6
JW14	72.9	17.0	10.1
JW16	75.8	13.6	10.6
JW18	77.2	13.6	9.2

TABLE IV. Optimum regression bandwidths b_R and the corresponding number of data points in parentheses for the forward mappings. Speakers are arranged by row and dependent variables by column.

	$F1$	$F2$	$F3$
JW11	0.004 (47)	0.004 (47)	0.004 (47)
JW14	0.003 (43)	0.003 (43)	0.003 (43)
JW16	0.003 (53)	0.002 (36)	0.002 (36)
JW18	0.004 (41)	0.003 (30)	0.003 (30)

TABLE V. Optimum interpolation bandwidths b_I and the corresponding number of data points in parentheses for the forward mappings. Speakers are arranged by row and dependent variables by column.

	*F*1	*F*2	*F*3
JW11	0.001 (12)	0.001 (12)	0.001 (12)
JW14	0.0008(12)	0.0008(12)	0.0008(12)
JW16	0.0007(12)	0.0007(12)	0.0007(12)
JW18	0.0009 (9)	0.0009 (9)	0.0012(12)

B. Optimum bandwidths for loess

The regression and interpolation bandwidths that were found optimal for the loess mappings by the cross-validation method are listed in Tables IV and V for forward mappings and in Tables VI and VII for inverse mappings. The neighborhood sizes represented by these bandwidths are found in parentheses.

The interpolation bandwidths are consistently smaller than the regression bandwidths with 7–18 data points included in the neighborhood for interpolation (Table IV versus Table V, and Table VI versus Table VII). The optimum regression bandwidths are consistently two to three times as large for the inverse mappings as for the forward mappings when comparing within subject (Table IV versus Table VI).

C. Forward mappings

It is possible to visualize the forward mappings from $K1$ and $K2$ to the formant frequencies $F1$ and $F2$ by plotting $[\hat{g}_j^{\rightarrow}]^{(1,2)}$for $j=1,2$, in which $F1$ and $F2$ values are averaged over $K3$ and $K4$. Four examples of forward mappings are presented. Figures 4 and 5 show these mappings from $K1$ and $K2$ to $F1$ and $F2$, respectively—that is, $[\hat{g}_1^{\rightarrow}]^{(1,2)}$ and $[\hat{g}_2^{\rightarrow}]^{(1,2)}$—for subject JW11. $K1$ has a relatively large effect on $F1$ with $K2$ having a slight effect. This would be expected from the projections of the $K1$ and $K2$ vectors onto the vector representing tongue height (Fig. 3). Figure 5 shows that $K1$ and $K2$ both have effects on $F2$, and that these parameters interact substantially. This can be expected from the projections of $K1$ and $K2$ onto the vector representing the tongue front-back dimension. Figures 6 and 7 show these averaged mappings from $K1$ and $K2$ to formants $F1$ and $F2$, respectively, for subject JW14. Both $K1$ and $K2$ affect $F1$, but with $K1$ having the largest effect, as can be expected by considering these vectors' projections onto the tongue height dimension. $K1$ and $K2$ also affect $F2$.

TABLE VI. Optimum regression bandwidths b_R and the corresponding number of data points in parentheses for the inverse mappings. Speakers are arranged by row and dependent variables by column.

	*K*1	*K*2	*K*3	*K*4
JW11	0.008 (94)	0.008 (94)	0.008 (94)	0.008 (94)
JW14	0.006 (86)	0.006 (86)	0.010(144)	0.008(115)
JW16	0.006(107)	0.008(142)	0.006(107)	0.006(107)
JW18	0.010(102)	0.010(102)	0.010(102)	0.012(122)

TABLE VII. Optimum interpolation bandwidths b_I and the corresponding number of data points in parentheses for the inverse mappings. Speakers are arranged by row and dependent variables by column.

	*K*1	*K*2	*K*3	*K*4
JW11	0.001 (12)	0.001 (12)	0.0012 (14)	0.0012 (14)
JW14	0.0005 (7)	0.0005 (7)	0.0012 (17)	0.0005 (17)
JW16	0.0005 (9)	0.001 (18)	0.0006 (11)	0.0006 (11)
JW18	0.0007 (7)	0.0011(11)	0.0013 (13)	0.0012 (13)

D. Inverse mappings

The inverse mappings from the formants to each of $K1-K4$ can be viewed as mappings from the $F1,F2$ plane to variables $K1-K4$ when the values are averaged over $F3$—that is, $[\hat{g}_j^{\leftarrow}]^{(1,2)}$ for $j=1,2,3,4$. (The independent variables $A1$, $A2$, and $A3$ have been transformed to $F1$, $F2$, and $F3$ here for purposes of illustration.) Instead of plotting these functions as a set of surfaces, a series of tongue shapes can be drawn in which each tongue shape corresponds to the value of $K1$, $K2$, $K3$, and $K4$ as either $F1$ or $F2$ varies. Examples of such inverse mappings are provided in Figs. 8 and 9. In Fig. 8, $F2$ is held constant, while $F1$ is increased for subjects JW11 [Fig. 8(a)] and JW14 [Fig. 8(b)]. The tongue goes from the thickest lines to the thinnest lines, and the lips from the largest asterisks to the smallest. The expected variation in tongue height and mouth opening is apparent. $F1$ is held constant in Fig. 9, while $F2$ is increased for subjects JW11 [Fig. 9(a)] and JW14 [Fig. 9(b)]. The tongues tend to move forward as $F2$ increases, and as $F2$ reaches the maximum value the mouth must open and the tongue tip drop in order to allow $F1$ to remain constant.

E. Distributions of local regression parameters

Fitting a loess model produces a set of regression parameters $\bar{\boldsymbol{\beta}}(\mathbf{x}_i)$ and $\bar{\gamma}(\mathbf{x}_i)$ at each data point $\mathbf{x}_i$ in the fitting set. We may examine the distributions of those parameters. As an

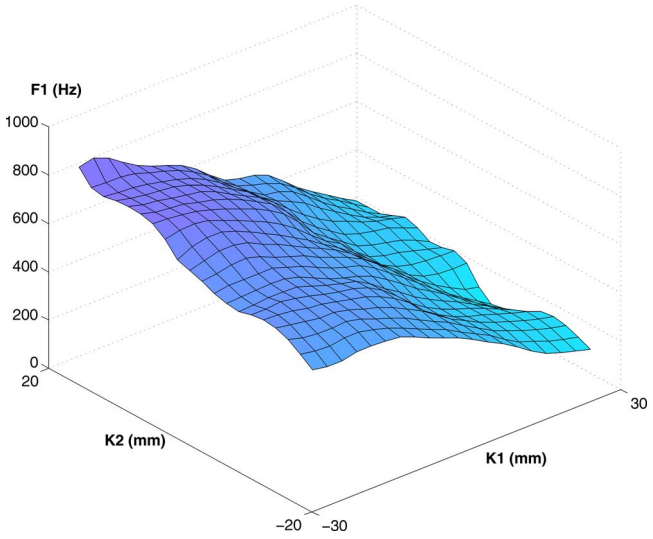


FIG. 4. (Color online) Visualization of the forward map from $K1$ and $K2$ to $F1$ for subject JW11 by a surface plot of $[\hat{g}_1^{\rightarrow}]^{(1,2)}$.

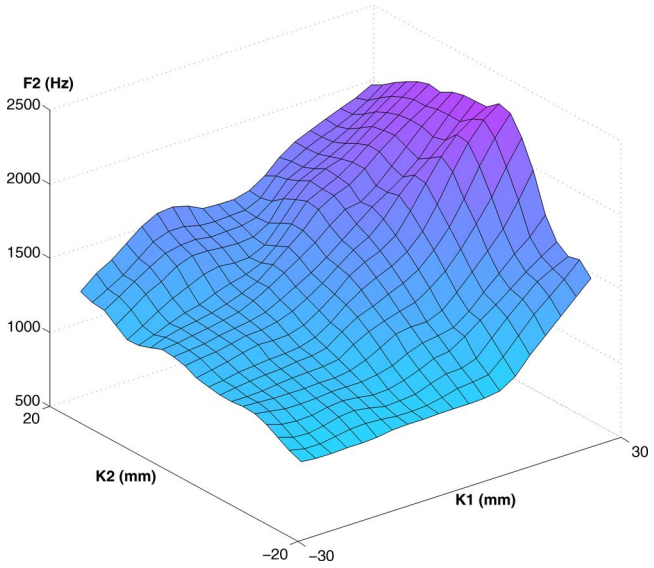


FIG. 5. (Color online) Visualization of the forward map from $K1$ and $K2$ to $F2$ for subject JW11 by a surface plot of $[\hat{g}_2^{\rightarrow}]^{(1,2)}$.

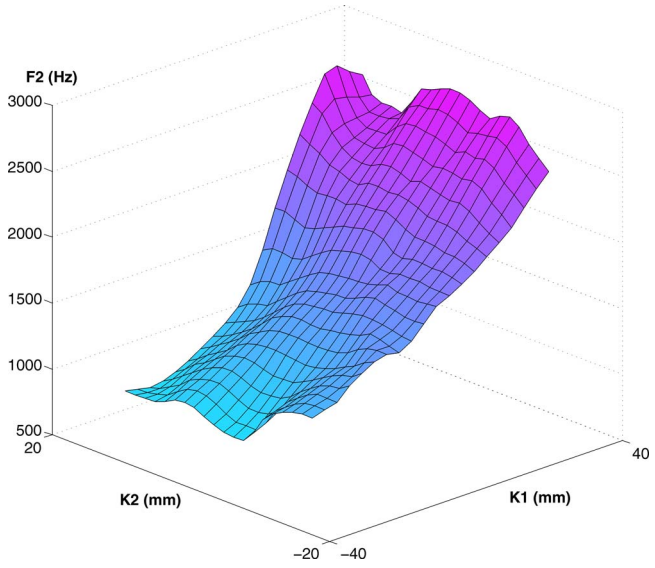


FIG. 7. (Color online) Visualization of the forward map from $K1$ and $K2$ to $F2$ for subject JW14 by a surface plot of $[\hat{g}_2^{\rightarrow}]^{(1,2)}$.

example, Fig. 10 shows the distribution of $\bar{\beta}_{11}^{\rightarrow}$ slope values for JW11. The distribution's tails have been removed in this figure.

The rest of the results are presented in terms of the means and standard deviations of the distributions of the loess regression coefficients. Figure 11 presents the means and standard deviations of the constants $\bar{\gamma}_j^{\rightarrow}$ in the forward mappings [Fig. 11(a)] and $\bar{\gamma}_j^{\leftarrow}$ in the inverse mappings [Fig. 11(b)]. The regression constants resulting from the corresponding global regressions are also indicated. The means of the loess constants and the constants from the global regression are all within one standard deviation of the distribution of constants for both the forward and inverse mappings. Of course the constants of the global regressions representing the forward mappings are simply the average formant frequencies of each speaker's data [Fig. 11(a)]. Because the variables $K1-K4$ for each speaker are the result of a PCA analysis with the mean subtracted out, the constants for the global regression representing the inverse mappings are very close to zero [Fig. 11(b)]. The standard deviations of the constants from the loess regressions for the forward mappings increase with the order of the dependent variable, formant frequency, for all subjects [Fig. 11(a)], where the "order" of formant Fj is j. There is no apparent trend between standard deviation of the constants from the loess regressions for the inverse mappings and the order of the dependent variable, articulatory PCA component [Fig. 11(b)].

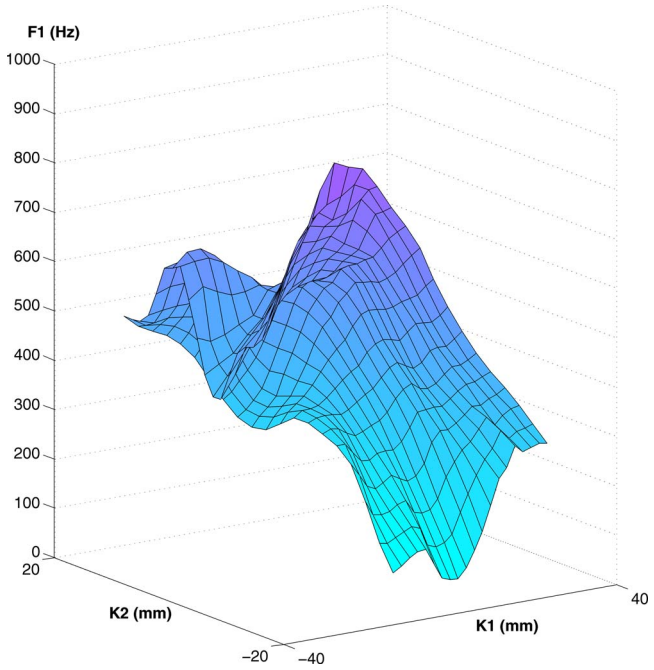


FIG. 6. (Color online) Visualization of the forward map from $K1$ and $K2$ to $F1$ for subject JW14 by a surface plot of $[\hat{g}_1^{\rightarrow}]^{(1,2)}$.

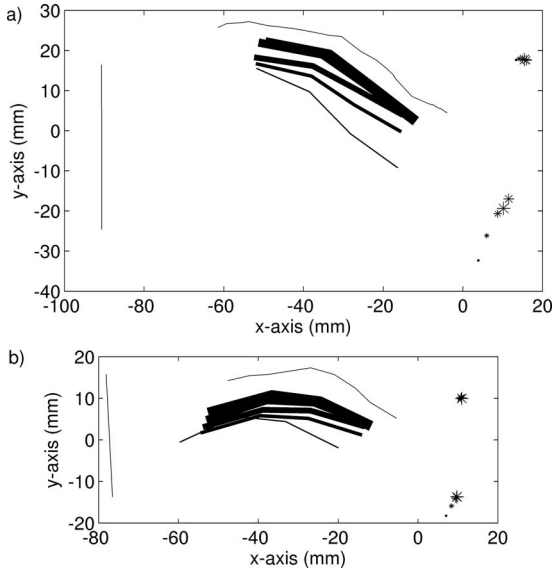


FIG. 8. Visualization of the inverse maps from $F1$ and $F2$ to $K1$, $K2$, $K3$, and $K4$. Tongue and lip positions predicted by $[\hat{g}_j^{\leftarrow}]^{(1,2)}$, $j=1,2,3,4$, are shown for a series of $F1$ increases for fixed $F2$. Tongue moves from thick line to thin. (a) JW11 with $F2=1556$ Hz and $F1$ from 145 to 877 Hz and (b) JW14 with $F2=1664$ Hz and $F1$ from 53 to 1019 Hz.

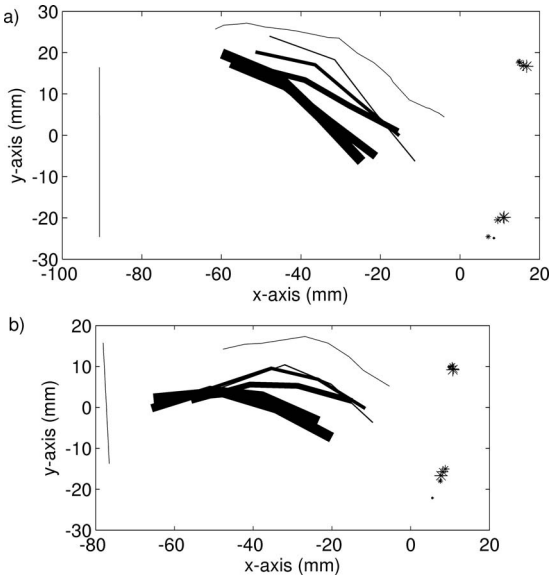


FIG. 9. Visualization of the inverse maps from $F1$ and $F2$ to $K1$, $K2$, $K3$, and $K4$. Tongue and lip positions predicted by $[\hat{s}_j^{\leftarrow}]^{(1,2)}$, $j=1,2,3,4$, are shown for a series of $F2$ increases for fixed $F1$. Tongue moves from thick line to thin. (a) JW11 with $F1=530$ Hz and $F2$ from 644 to 2377 Hz and (b) JW14 with $F1=613$ Hz and $F2$ from 522 to 2910 Hz.

Figure 12 shows the means and standard deviations of the slopes $\bar{\boldsymbol{\beta}}_j^{\rightarrow}=(\bar{\beta}_{1j}^{\rightarrow},\bar{\beta}_{2j}^{\rightarrow},\bar{\beta}_{3j}^{\rightarrow},\bar{\beta}_{4j}^{\rightarrow})$, $j=1,2,3$, of the forward mappings for each speaker. The corresponding quantities from the global regression are also indicated. The means of the loess slopes are very close to the corresponding global regression slopes when compared to the size of the standard deviations. There is a consistent trend of increasing standard deviation with the order of either the independent variable (i.e., the i in Ki and $\bar{\beta}_{ij}^{\rightarrow}$) and the order of the dependent variable (i.e., the j in Fj and $\bar{\beta}_{ij}^{\rightarrow}$). Figure 13 shows the mean

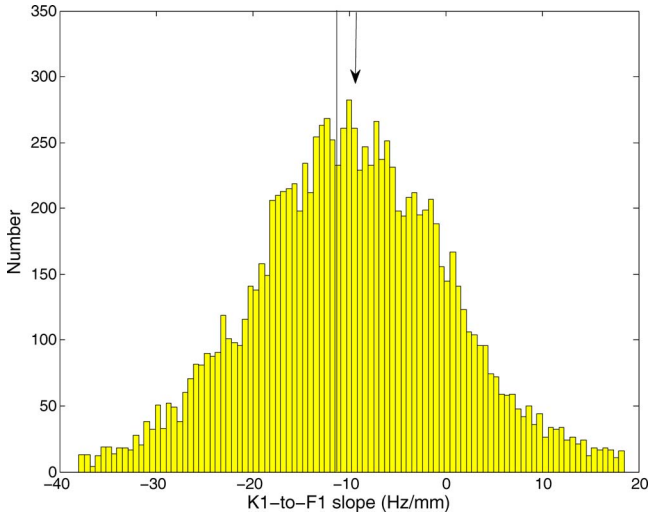


FIG. 10. (Color online) Histogram of local $K1$-to-$F1$ slope, or $\bar{\beta}_{11}^{\rightarrow}$, at the data points in the forward mapping for JW11. The arrow denotes the mean of the distribution, and the vertical line indicates the value of the $K1$-to-$F1$ slope obtained in global regression. The slope data have been trimmed to cut the tails from the distributions. The tails for a distribution X are defined as follows. Let x_q=value of X that marks the qth percentile of the X distribution, and let x_{mean}=mean of the X distribution. Then y is in one of the tails if $x_{25}-y>1.5|x_{25}-x_{\text{mean}}|$ or $y-x_{75}>1.5|x_{75}-x_{\text{mean}}|$.

and standard deviations of the loess slopes $\bar{\boldsymbol{\beta}}_j^{\leftarrow}=(\bar{\beta}_{1j}^{\leftarrow},\bar{\beta}_{2j}^{\leftarrow},\bar{\beta}_{3j}^{\leftarrow})$, $j=1,2,3,4$, of the inverse mappings for each speaker, along with the corresponding slopes from the global regressions. Again, the slopes from the global regressions are very close to the means of the corresponding slopes from the loess regressions. While there is a general trend for increasing standard deviation with the order of the independent variable (i.e., the i in Ai and $\bar{\beta}_{ij}^{\leftarrow}$), there is no apparent trend with the order of the dependent variable (i.e., the j in Kj and $\bar{\beta}_{ij}^{\leftarrow}$).

A sample of correlation coefficients among the slopes and constant was calculated. This was done for JW11 and JW14 in their loess forward mappings to $F1$ and $F2$. The independent variables $K1$ and $K2$ were the focus. Thus, the three correlation coefficients among constants, $\bar{\gamma}_j^{\rightarrow}$, and slopes, $\bar{\beta}_{1j}^{\rightarrow}$ and $\bar{\beta}_{2j}^{\rightarrow}$, for fixed $j=1$ or 2 were calculated for each JW11 and JW14. The three correlation coefficients among the absolute values of these quantities were also calculated for both speakers. The correlation coefficients in Tables VIII and IX indicate varying degrees of co-variation among loess parameters and their absolute values. Given the large number of data points, all of these correlation coefficients are significant at the $p<0.001$ level (Bickel and Docksum, 1977, pp. 221 and 472).

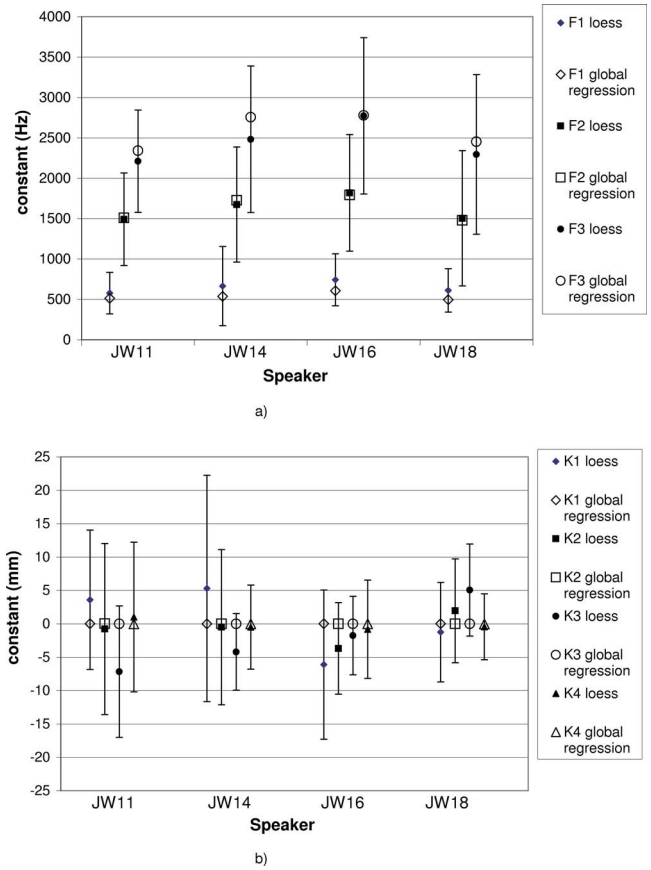


FIG. 11. (Color online) Means and standard deviations of the local regression constants ($\bar{\gamma}_j^{\rightarrow}$ and $\bar{\gamma}_j^{\leftarrow}$) at the data points. The constants obtained in global regression are also shown. Labels refer to the dependent variable. (a) $\bar{\gamma}_j^{\rightarrow}$ (forward mappings) and (b) $\bar{\gamma}_j^{\leftarrow}$ (inverse mappings).

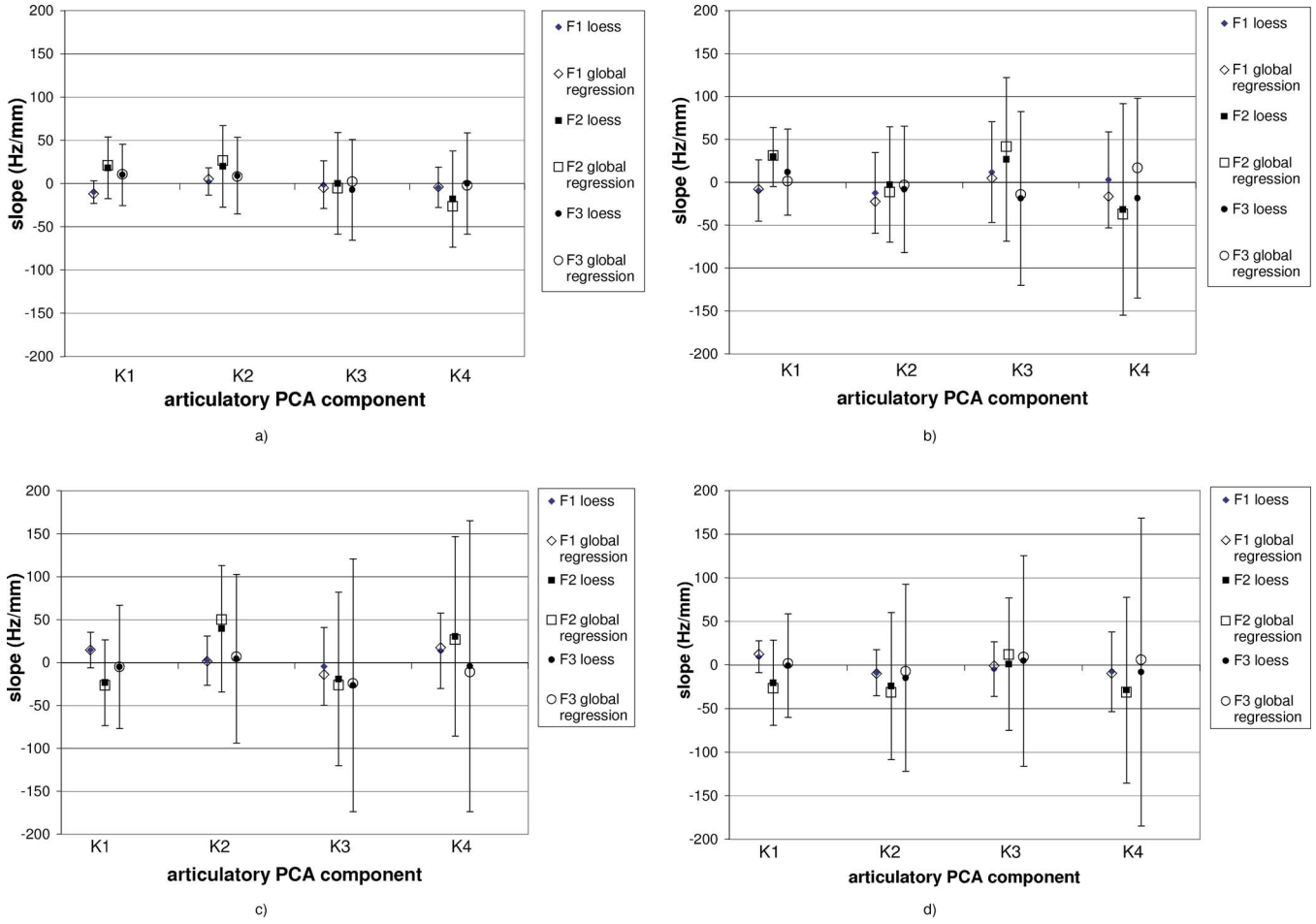


FIG. 12. (Color online) Means and standard deviations of the local regression slopes of the forward mappings at the data points (i.e., distributions of $\bar{\beta}_{ij}^{\rightarrow}$ for $i=1,2,3,4$ and $j=1,2,3$). Corresponding global regression slopes are also plotted. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

F. Test RMSE

Figure 14 shows the RMSE results from applying the forward loess mappings to each speaker's test data set. If it is assumed that the formant frequency values are distributed in a normal distribution about a mean, then the RMSE is one standard deviation of the distribution. There is a general trend for increasing RMSE with the order of the dependent variable, or formant frequency number, for each speaker. Figure 15 shows the percent decrease in the RMSE of the forward loess compared to the corresponding global regression applied to the same test data. The improvement in fit from the global regression to the loess regression is consistently greatest for $F2$ in terms of percent reduction in RMSE for each subject.

Figure 16 exhibits the test RMSE values for the inverse loess mappings. There was a general trend, though not completely consistent, of decreasing RMSE error from $K1$ and $K2$ to $K3$ and $K4$ for each speaker. This can be contrasted with increases in RMSE from $F1$ to $F3$ in the forward mappings for each speaker. Figure 17 shows the percent decrease in RMSE for the loess from the global regression for the inverse mapping. The amount of decrease for the inverse mappings is generally less than for the forward mappings (Fig. 15).

G. Sensitivity

The results of the sensitivity analysis are presented by sectioning a schematic two-dimensional articulatory vowel space of each speaker into regions of large and small sensitivity of acoustic variables to articulatory variables. For each speaker, schematic articulatory vowel spaces were constructed by connecting the four point-vowel articulations projected onto their $(K1,K2)$ coordinates to form a quadrilateral. Adjoined to this quadrilateral was another quadrilateral representing the region between the projected [i] and [u] articulations and the palate. The resulting polygon formed a schematic of the articulatory vowel space in $K1-K2$ space.

The schematic articulatory vowel space of each speaker was divided into regions of large and small sensitivity according to each empirical sensitivity function $\nu_{ij}^{\rightarrow}$, $i=1,2$, $j=1,2$. Boundaries of the regions were defined by visual inspection. The results were grouped into four figures with Fig. 18 corresponding to $F1$ sensitivity to changes in $K1$ (magnitudes of $\nu_{11}^{\rightarrow}$), Fig. 19 corresponding to $F1$ sensitivity to changes in $K2$ (magnitudes of $\nu_{21}^{\rightarrow}$), Fig. 20 corresponding to $F2$ sensitivity to changes in $K1$ (magnitudes of $\nu_{12}^{\rightarrow}$), and Fig. 21 corresponding to $F2$ sensitivity to changes in $K2$ (magnitudes of $\nu_{22}^{\rightarrow}$).

Figure 18 shows a general tendency for the regions of

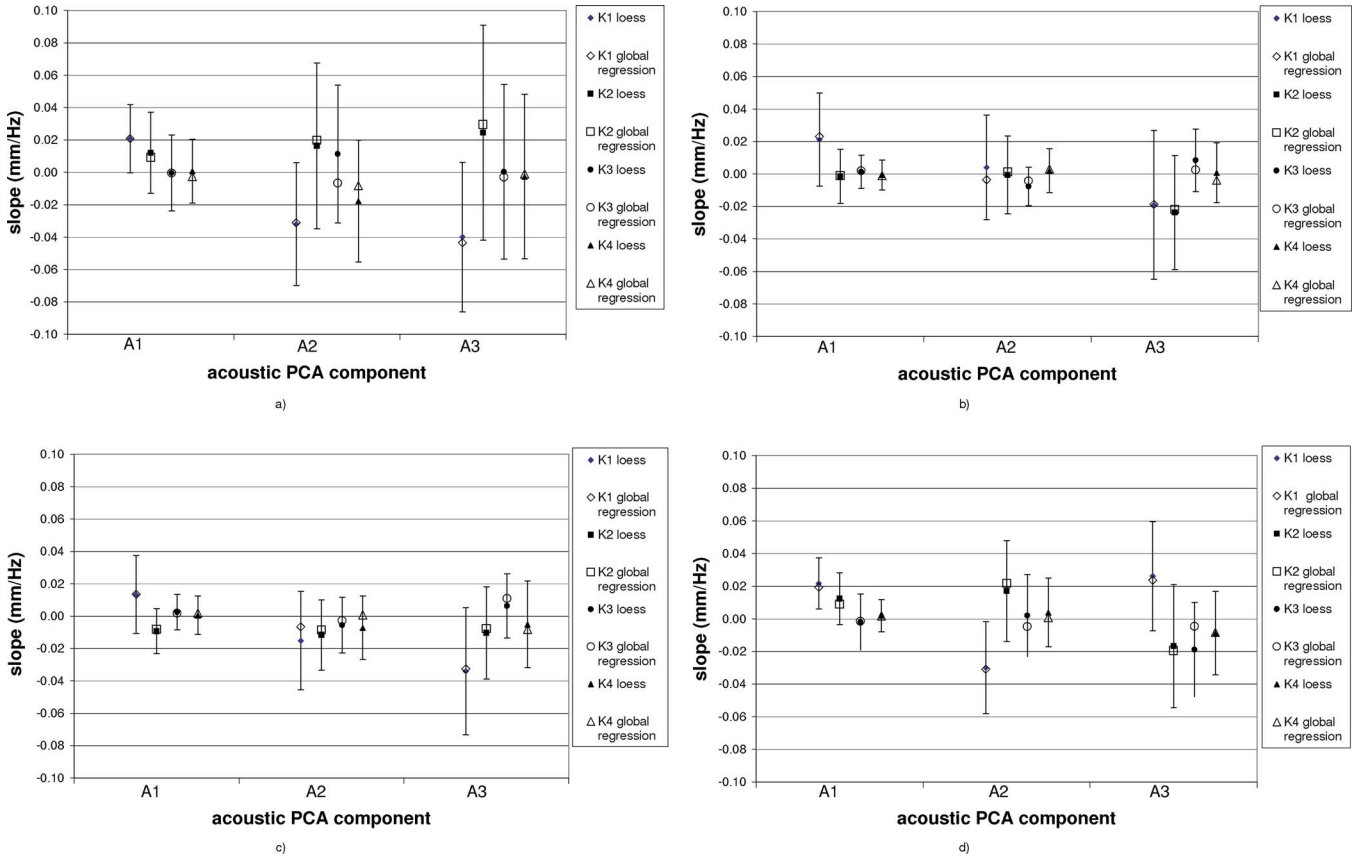


FIG. 13. (Color online) Means and standard deviations of the local regression slopes of the inverse mappings at the data points (i.e., distributions of $\bar{\beta}_{ij}^{\leftarrow}$ for $i=1,2,3$ and $j=1,2,3,4$). Corresponding global regression slopes are also plotted. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

largest sensitivity of $F1$ to $K1$, or largest $\nu_{11}^{\rightarrow}$, to be associated with high vowels and the least sensitive with low vowels. There are differences among the speakers, with JW11 [Fig. 18(a)] and JW18 [Fig. 18(d)] exhibiting similar patterns. (These are sketches where differences in the amount of area covered by the shadings are not a reliable indication of differences in sensitivity.) JW14 [Fig. 18(b)] appears to have the pattern of JW11 and JW18 rotated counterclockwise through her vowel space. JW16 has neither large nor small sensitivity in the front part of her vowel space. Figure 19 for the sensitivity of $F1$ to $K2$ change again shows that regions of large sensitivity, or large $\nu_{21}^{\rightarrow}$, are in the high part of the vowel space, and regions of small sensitivity are in the low part of the vowel space. In fact, for JW11 the sensitivity regions are quite similar between $\nu_{11}^{\rightarrow}$ and $\nu_{21}^{\rightarrow}$. In contrast, JW14's pattern for $\nu_{21}^{\rightarrow}$ appears to be a clockwise rotation of her pattern for $\nu_{11}^{\rightarrow}$ [Figs. 18(b) and 19(b)]. JW18 shows a similar small rotation [Figs. 18(d) and 19(d)]. JW16 exhibits a sensitivity pattern in $\nu_{21}^{\rightarrow}$ similar to that of JW11, except her region of small sensitivity appears to extend higher into her vowel space [Figs. 19(c) and 19(a)].

The sensitivities of $F2$ to $K1$ change and $K2$ change (Figs. 20 and 21) show more regions of large and small sensitivity than the sensitivity of $F1$ to $K1$ change and $K2$ change (Figs. 18 and 19). This is to be expected because the higher frequency formant depends on finer-scaled details of articulation than the lower frequency formant does. All of the speakers appear to have large $F2$ sensitivity to $K1$ change, or large $\nu_{12}^{\rightarrow}$, in the high-back region of the vowel space and a region of small sensitivity somewhere in the low region of the vowel space with, often, an accompanying region of large sensitivity (Fig. 20). On the other hand, the regions of large sensitivity of $F2$ to $K2$ change, or large $\nu_{22}^{\rightarrow}$, are in the back part of the vowel spaces, except for JW16 (Fig. 21). The regions of small sensitivity of $F2$ to $K2$ change are most notably in the front of the vowel spaces, except for JW11, who shows small sensitivity around the low-back vowel (Fig. 21).

The critical articulation points in the $K1,K2$ subgrid—i.e., those subgrid vertices with $\nu_{ij}^{\rightarrow}$ values smaller than that of any of the neighboring vertices—were examined in the context of the point vowels projected onto the $K1,K2$ plane. The critical articulations that corresponded closely to point vowels for each speaker are noted in Table X.

TABLE VIII. Correlation coefficients for local regression parameters at the data points with means subtracted in the forward mappings. The subscript j denotes the dependent variable (formant) number.

	$\bar{\gamma}_j^{\rightarrow}$ by $\bar{\beta}_{1j}^{\rightarrow}$	$\bar{\gamma}_j^{\rightarrow}$ by $\bar{\beta}_{2j}^{\rightarrow}$	$\bar{\beta}_{1j}^{\rightarrow}$ by $\bar{\beta}_{2j}^{\rightarrow}$
JW11, $j=1$	−0.174	−0.188	0.133
JW11, $j=2$	−0.471	−0.271	0.303
JW14, $j=1$	0.232	0.181	−0.032
JW14, $j=2$	−0.217	−0.075	0.048

TABLE IX. Correlation coefficients for absolute values of the local regression parameters at the data points with means subtracted in the forward mappings. The subscript j denotes the dependent variable formant number.

	$\lvert\vec{\gamma}_j\rvert$ by $\lvert\vec{\beta}_{1j}\rvert$	$\lvert\vec{\gamma}_j\rvert$ by $\lvert\vec{\beta}_{2j}\rvert$	$\lvert\vec{\beta}_{1j}\rvert$ by $\lvert\vec{\beta}_{2j}\rvert$
JW11, $j=1$	0.683	0.365	0.362
JW11, $j=2$	0.719	0.618	0.599
JW14, $j=1$	0.545	0.500	0.272
JW14, $j=2$	0.529	0.424	0.301

The results show that there is little consistency among the speakers as to the point vowels that could be considered to have critical $(K1, K2)$ coordinates.

IV. DISCUSSION

A. Articulatory PCA

The lower-order articulatory PCA components in the present study accounted for less of the variance than is typical for PCA analyses of a small number of vowels. Typically two principal components will account for at least 95% of the variance for a small number of static vowel images (on the order of 10) (e.g., Jackson, 1988; Mokhtari *et al.*, 2007; Jackson and McGowan, 2008). On the other hand, for subject JW16, more of the tongue position variance was accounted for in the present study than was accounted for in the Beaudoin and McGowan (2000) study involving the same subject, also with thousands of data points. That is, the sum of the variances accounted for in the first one to four components is always greater in the present work than in the Beaudoin and McGowan (2000) study, where tokens were chosen automatically based on a criterion of minimum tongue-to-palate distance. It appears that the present data set based on vowels alone and consonant-vowel transitions that excluded nasal, lateral, and rhotic contexts was more restrictive than the earlier study.

The degrees-of-freedom that were dominant in the $K1$ and $K2$ components were familiar high-low and front-back degrees-of-freedom, except that the $K1$ and $K2$ components of JW16 additionally contained a large amount of tongue blade up-down and tongue bunching-stretching degrees-of-freedom. $K3$ and $K4$ contain varying amounts of tongue bunching-stretching as well as tongue blade up-down degrees-of-freedom. These degrees-of-freedom would be expected to appear because transitions to and from consonants have been included in the data set. None of the articulatory PCA components have substantial amounts of lip opening or lip protrusion. This could be due to the fact that there is only one phonologically rounded monophthong in American English, and that the vowels in the XRMB-SPD appeared with both unrounded and rounded consonants. The latter fact may mean that lip position was not correlated with tongue position.

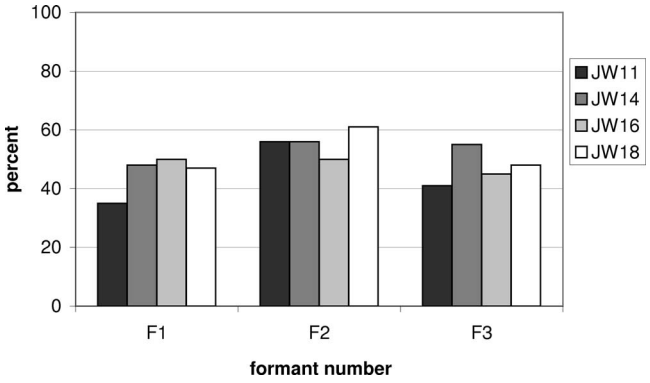


FIG. 15. Percent improvement in test RMSE of loess over global regression for the forward mappings.

B. Distributions of local regression parameters

In Sec. III E, the wide distributions of constants and slopes of loess were noted. It is of some interest to investigate whether there is co-variation among the constant and slope values in order to know whether the values in the tails of their distributions could be accounted for in terms of compensation. Evidence for at least a small amount of such covariation was found.

However, a comparison between corresponding cells in Tables VIII and IX show that the correlation coefficients for the absolute values are larger, by a factor of 1.5–8.5, than their corresponding coefficients of the signed parameters. This indicates that while there may be a negative (or positive) relation between two parameters, there are substantial numbers of data points where a positive (or negative) relation holds between them. Further, there is a strong correlation in magnitudes no matter the sign. The covariance among loess parameters cannot be neglected, yet the sign of the covariance is only weakly determined compared to the covariance in magnitude.

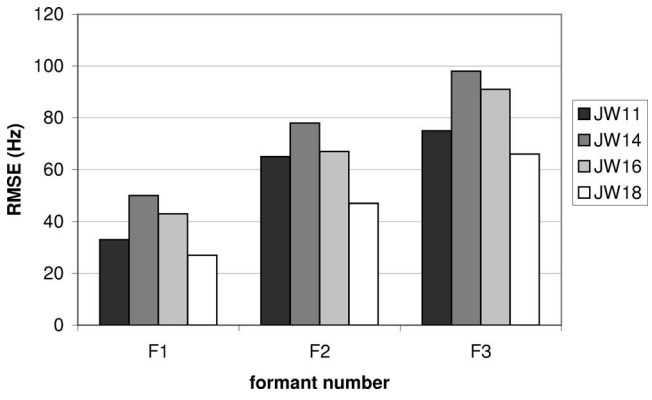


FIG. 14. Test RMSE of the forward loess mappings.

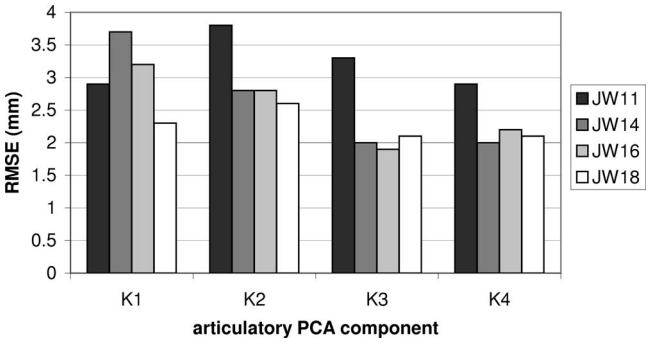


FIG. 16. Test RMSE of the inverse loess mappings.

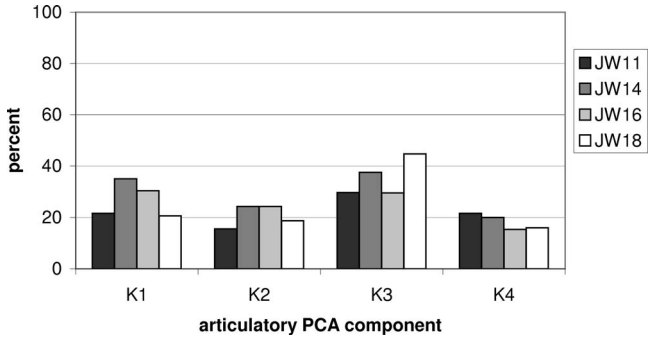


FIG. 17. Percent improvement in test RMSE of loess over global regression for the inverse mappings.

C. Measurement error and test RMSE

The different contributions to error—in measuring articulatory coordinates, in measuring formant frequencies, and in relating the two domains—are now considered. Much of the discussion will focus on how measurement error relates to the test RMSE of the loess models. It will be seen that the RMSEs for the forward loess mappings are better than expected based on measurement error considerations, but the RMSEs for the inverse loess maps are not as good as would be expected based on estimated measurement errors. Explanations for the latter phenomenon will be offered.

There are two sources of formant variability that cannot be attributed to measured articulatory variability. One source is the movement of the speech articulators whose positions were not measured. An example of an articulator coordinate that was not measured for the XRMB-SPD is larynx height. Westbury (1983), using X-ray cineradiography, measured changes in larynx height of 17 mm during speech. Wood (1979) measured a maximum range in larynx height of only 5 mm for all the speakers he studied in 15 different languages and dialects. Perkell (1969, pp. 38–42) measured larynx height changes of nearly 15 mm in isolated utterances of a single speaker. However, he did state that these variations were diminished in running speech. Changes in larynx position can have a substantial effect on the formant frequencies. For example, Lindblom and Sundberg (1971) showed in experiments with a synthesizer that $F1$ and $F2$ in vowel production could change by up to 8% with a 10 mm change in larynx height. There are other articulatory dimensions that are not measured, such as pharyngeal dimensions, velum position, and all lateral dimensions. But these effects have been minimized here. It has been shown that in English and Swedish, pharyngeal dimensions are largely predictable from the positions of the front of the tongue during static vowel production (Whalen *et al.*, 1999; Jackson and McGowan, 2008). Furthermore, restricting vowel contexts to exclude nasal and lateral consonants has minimized the amount of velar and lateral variation for any given set of tongue pellet coordinates. Thus, it would seem to be reasonable to assume that the unseen articulatory coordinates should cause a variation of at most 10%–15% in formant frequencies for a given set of tongue and lip pellet coordinates.

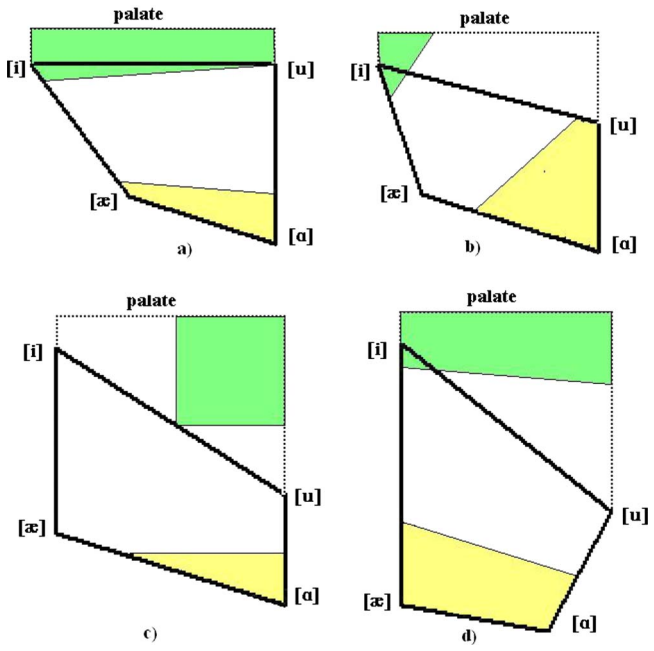


FIG. 18. (Color online) Schematic articulatory vowel spaces with regions of large (dark shading) and small (light shading) empirical sensitivity of $F1$ to $K1$ ($\vec{\nu_{11}}$). The vowel spaces are defined by the articulation of the point vowels projected onto ($K1$,$K2$) coordinates. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

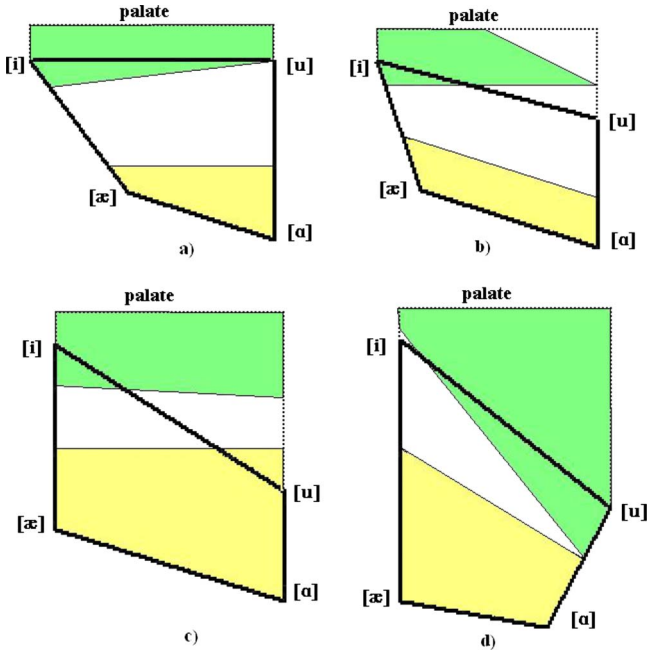


FIG. 19. (Color online) Schematic articulatory vowel spaces with regions of large (dark shading) and small (light shading) empirical sensitivity of $F1$ to $K2$ ($\vec{\nu_{21}}$). The vowel spaces are defined by the articulation of the point vowels projected onto ($K1$,$K2$) coordinates. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

Another source of variability for formant frequency values is measurement error incurred in using LPC analysis with human correction. The errors in the formant frequency measurements themselves are estimated to be less than about 5%. In total, one can expect a "cloud" of $F1$, $F2$, and $F3$ measures of radius of about 15%–20% of the mean to be associated with each data point in the articulatory space. In terms of formant frequency values, this would correspond to

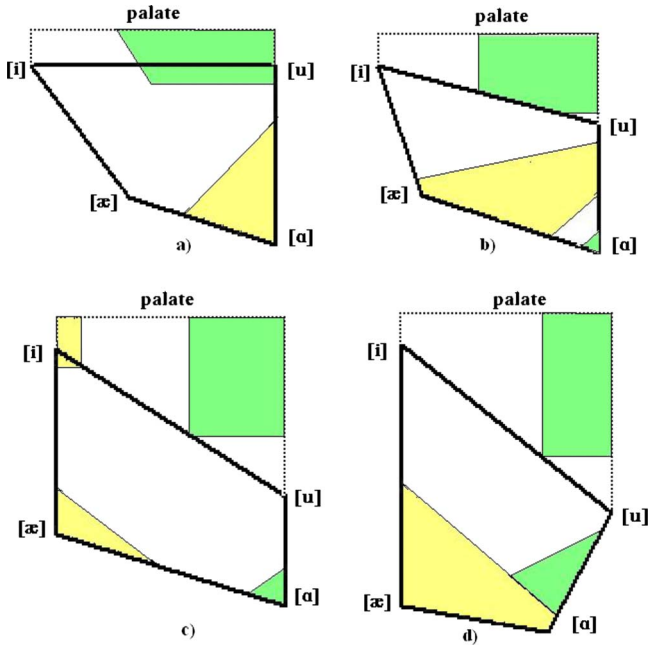


FIG. 20. (Color online) Schematic articulatory vowel spaces with regions of large (dark shading) and small (light shading) empirical sensitivity of $F2$ to $K1$ ($\nu_{12}^{\rightarrow}$). The vowel spaces are defined by the articulation of the point vowels projected onto ($K1$,$K2$) coordinates. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

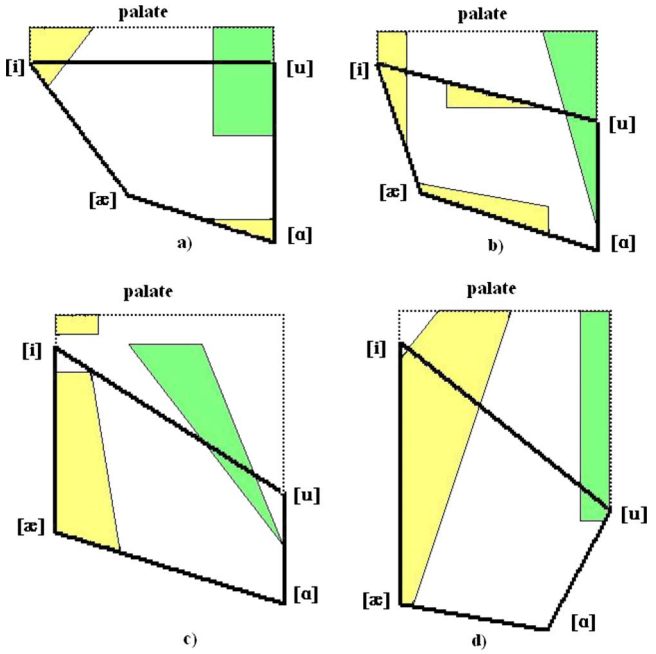


FIG. 21. (Color online) Schematic articulatory vowel spaces with regions of large (dark shading) and small (light shading) empirical sensitivity of $F2$ to $K2$ ($\nu_{22}^{\rightarrow}$). The vowel spaces are defined by the articulation of the point vowels projected onto ($K1$,$K2$) coordinates. (a) JW11, (b) JW14, (c) JW16, and (d) JW18.

radii of 75–100 Hz in $F1$, 225–300 Hz in $F2$, and 375–500 Hz in $F3$ based on means of 500, 1500, and 2500 Hz for $F1$, $F2$, and $F3$ respectively.

We may compare the above estimates of the radii of the formant clouds with the RMSEs of the forward loess mappings. If the error estimates can be understood to be on the order of two standard deviations of a normal formant frequency error distribution (accounting for 84% of the error), and the RMSEs are expected to be one standard deviation of the error distribution, then the error estimate should be twice the RMSE. In the loess mappings, doubling the RMSE for $F1$ gave a value in the range from 54 to 100 Hz, for $F2$ from 94 to 156 Hz, and for $F3$ from 132 to 196 Hz (Fig. 14). These data indicate that twice the standard deviation is on the order of 10%–20% of the mean for $F1$ and 6%–10% of the mean for $F2$ and $F3$. Thus, forward loess mappings for $F2$ and $F3$ perform better than would be expected based on error estimates. Thus it is reasonable to reduce the estimated standard deviation in the formant error distribution to coincide with the forward mappings' RMSEs.

We have estimated the RMSE of $F1$ to be 5%–10% of the mean value of $F1$ and the RMSEs of $F2$ and $F3$ to be 3%–5% of the mean values of $F2$ and $F3$, respectively. These values compare favorably with the error obtained by Mermelstein (1967) when he replaced an area function for six vowels with the first six cosine components of that area function. This replacement produced maximum errors of about 3% in the formant frequencies (Mermelstein, 1967). In contrast to Mermelstein's (1967) investigation from a single set of six model area functions, the present investigation is based on large numbers of articulatory configurations from human talkers with four orthogonal articulatory components spanning the articulatory space.

The estimated positional error for stationary pellets is 0.15 mm (Westbury, 1994, p. 69). Further, it can be expected that fast moving articulators (400 mm/s) will move 0.6 mm during the time it takes for the tracking raster to be generated, thus creating an error due to recording time delay. Another possible source of error is the time delay between the articulatory position and the recorded acoustics. This is error caused by the fact that the speed of sound is finite and the distance of the microphone to the mouth was about 110 mm. It is estimated that the acoustic signal is recorded 0.3 ms after it is generated by the vocal tract, which corresponds to about 5% of the time between sampled frames of data (at 160 Hz). For a fast moving articulator this would correspond to only 0.03 mm of movement, which is small compared to the error caused by the recording time delay. With these considerations, a reasonable estimate of maximum position measurement error is 0.6 mm for each pellet coordinate. If the position measurement errors are independent across the pellet coordinates, then the maximum total magnitude of error for the six pellets would be about 2 mm. In particular, this can be considered to be the measurement error in each of the articulatory PCA components, $K1$, $K2$, $K3$, and $K4$.

TABLE X. The point vowels projected onto $K1$,$K2$ coordinates that closely matched critical values of ($K1$, $K2$) for each empirical sensitivity function $\nu_{ij}^{\rightarrow}$.

	JW11	JW14	JW16	JW18
$\nu_{11}^{\rightarrow}$	[i]		[ɑ]	[ɑ]
$\nu_{12}^{\rightarrow}$	[æ]	[i]	[æ]	[u]
$\nu_{12}^{\rightarrow}$	[ɑ], [æ], [i]		[æ]	[ɑ]
$\nu_{22}^{\rightarrow}$	[æ], [i]		[ɑ]	[u]

If, as in the estimate of the acoustic error, the estimated articulatory measurement error is about two standard deviations of the articulatory error distribution, then it is possible to compare this estimate with twice the test RMSE in inverse loess mappings. Doubling the RMSEs of the inverse mappings gives values between 3.8 and 7.6 mm. The estimated corresponding quantity for the error distribution of 2 mm is well below this range. Causes for this discrepancy and the differences between the forward and inverse mappings in terms of the relation between error distributions and RMSE are proposed below.

One factor contributing to the RMSE of the inverse mappings being two to four times larger than the estimated standard deviation of the error distribution of the articulatory dependent variables is the compensatory ability of the vocal tract. For a given triple of formant frequencies ($F1$, $F2$, and $F3$), there are many vocal tract shapes able to produce it (Atal *et al.*, 1978). Not only can $K1$, $K2$, $K3$, and $K4$ compensate for one another but the unseen articulators, such as the larynx, can do the same. Thus, a point in the acoustic space can map into a region of the articulatory space. This is without any consideration of measurement error. It can be speculated that this is the reason why the optimum regression bandwidths are at least twice as large for the inverse mappings as for the forward mappings of each speaker (Tables IV and VI). It is a reasonable possibility that more acoustic variables need to be included in the inverse mapping to ensure that the inverse mappings are as unambiguous as possible. It can be proven that the RMSE is a positively biased estimate of the error standard deviation if independent variables with significant effect have not been included in the analysis (Chatterjee and Hadi, 1988, pp. 40–42).

Another contribution to the larger RMSE of the inverse mappings might involve the assumptions made in regression analysis. When regression is performed it is assumed that the independent variables are known exactly, or at least the independent variables should be known more precisely than the dependent variables. Therefore it of interest to know how the magnitudes of the errors in articulatory variables compare to those of the acoustic variables, or how the errors in the articulatory PCA components compare to errors in the acoustic PCA components. The standard deviations of the error distribution in the formant frequencies were estimated by the RMSEs of the optimum loess forward mappings, because the original estimates of those standard deviations were found to be too large. On the other hand the 1 mm estimate will be used for the standard deviation in the error distribution of each of the articulatory PCA components. In order to compare magnitudes of errors for physical variables (articulatory and acoustic) with different dimensionalities, it is necessary to normalize both sets of variables. They are normalized by the standard deviations of the coordinate data in the data set for each speaker. For instance, the estimated standard deviation in the error distribution of $K2$ was normalized by the total standard deviation in the $K2$ data.

The normalized estimated error standard deviations averaged over the four speakers for $K1$, $K2$, $K3$, and $K4$ are 0.10, 0.17, 0.27, and 0.33, respectively. The same quantities for $A1$, $A2$, and $A3$ are 0.18, 0.27, and 0.44. The ordering of these variables by increasing magnitude of the normalized estimated error standard deviations is $K1$, $K2$, $A1$, ($A2$,$K3$), $K4$, $A3$, which indicates that the lower-order articulatory quantities are determined with less error than the lower-order acoustic quantities. The average normalized estimated error standard deviations are 0.22 ± 0.10 for the articulatory variables and 0.29 ± 0.13 for the acoustic variables. While these averages are not significantly different, the lower-order articulatory variables do appear to be known more accurately than the acoustic variables. These considerations call into question the assumption that the independent variables are known relatively precisely for the loess regressions for the inverse mappings.

To compare these results with previous work in the inverse mapping, reference can be made to Ladefoged *et al.* (1978) and Mokhtari *et al.* (2007). The magnitudes of the RMSEs for $K1-K4$ shown in Fig. 16 are close to, or less than, those of the inverse mappings constructed by Ladefoged *et al.* (1978) for five subjects each producing five vowels (Ladefoged *et al.*, 1978, Table II). Mokhtari *et al.* (2007) investigated one subject speaking a series of vowels for which there were 35 frames of data. They performed regression using three formants as the independent variables and articulatory PCA components of the area function as the dependent variables. For each PCA component they reported the RMSE normalized by the standard deviation of that component: 0.22 and 0.45 for the first and second PCA components, respectively (Mokhtari *et al.*, 2007, Fig. 4). By performing the same normalization using the standard deviations of $K1$ and $K2$, the results of the inverse loess may be compared with these values. The normalized RMSE averaged across the four subjects was 0.29 ± 0.05 for $K1$ and 0.50 ± 0.05 for $K2$.

We have provided only rough estimates of the error expected in both the acoustic and articulatory domains. However, to provide quantitative confidence intervals for both the mappings and the parameters in the mappings, such as $\boldsymbol{\beta}(\mathbf{x})$, the error in both domains should be estimated using the RMSEs and the properties of the loess mappings. There are well-known, established methods to do this with parametric linear regression (e.g., Bickel and Docksum, 1977, pp. 267–268). These methods can be extended to loess, but with some complications. In both parametric linear regression and loess, the estimates of the dependent variable may be viewed as a linear transformation of the data in the dependent variable domain. For parametric linear regression this transformation is an idempotent projection, while for loess the transformation is not idempotent, and it is not even symmetric (Cleveland and Devlin, 1988). This makes computing the estimated variances much more computationally intensive: it involves computations of the traces of products of $N \times N$ matrices, where N is the number of data points. These computations were prohibitive for the current data sets of 10 000–20 000 data points. In the future we will research methods for finding approximations to these traces. Another procedure for finding confidence intervals would be to use Monte Carlo simulation. These issues merit further research.

Another issue confronting estimation of error is covariation in the independent variables. Since the independent

variables are PCA components, they are orthogonal in the global sense, but they are not necessarily so for each locally weighted regression, performed here at each data point. It is entirely possible that the independent variables are nearly linearly dependent in certain of the local regressions. Averaging over the higher-order independent variables was used here for visualization and data-reduction purposes. However, this averaging lessened the effects of co-variation that could cause the loess mappings and the measures of formant sensitivity to be noisier. To provide quantitative confidence intervals for both the mappings and the parameters in the mappings, future work should address the issue of co-linearity of independent variables in local regions and co-variation of loess regression coefficients between regions.

D. Sensitivity

The results on sensitivity of $F1$ and $F2$ to $K1$ and $K2$ changes near the point vowels [ɑ], [æ], and [i] can be viewed in terms of the geometric properties of $K1$ and $K2$ shown in Figs. 2 and 3, and the effects of those properties in simple acoustic tube models. Sensitivities of formants to articulatory change for the vowels [ɑ] and [i] can be understood using two-tube models, while [æ] can be viewed as a slightly flared tube (Perkell, 1969, p. 55; McGowan, 2006). The tube properties of [u] are more complex and discussion of this vowel would be more speculative. Further, the discussion will focus on subjects JW11, JW14, and JW18, with only brief mentions of JW16 because of the difficulty in characterizing the $K1$ component of JW16 due to its tongue rocking motion.

Ehrenfest's theorem applied to acoustic tubes was introduced to the speech production community by Schroeder (1967). In this community it is commonly known as acoustic perturbation theory and it is useful for the discussion of sensitivity in conjunction with the two-tube models. Briefly, if, for a given resonant frequency f, cross-sectional tube area is decreased a small amount in a region where acoustic potential energy density is greater than acoustic kinetic energy density, then the resonant frequency increases, that is, $df/f > 0$. The opposite occurs if the kinetic energy density is greater than the potential energy density. To apply this theorem, one needs to compute resonant frequencies and their corresponding energy density distributions through the tube, which is done numerically. However, for a sufficiently low resonant frequency f (e.g., $F1$) and a two-tube model, there is a simplification. For a two-tube model, let the tube closest to the glottis have length L_1 and area A_1, and the tube closest to the lips have length L_2 and area A_2, and let $\alpha = A_2/A_1$. Assuming that the total tube length remains constant, it can be shown that

$$\frac{df}{f} \approx \frac{d\alpha}{\alpha} - \frac{dL_2}{L_2}. \tag{7}$$

In other words, the fractional change in frequency equals the fractional change in area ratio minus the fractional change in length of the front tube, to first order. This is a theoretically derived sensitivity function that relates sensitivity to vocal tract geometry.

JW11, JW14, and JW18 all show large $F1$ sensitivity to $K1$ changes ($\nu_{11}^{\rightarrow}$) in a region around [i] (Fig. 18). The $K1$ of each of these speakers has the effect of moving the tongue toward or away from the palate (Fig. 2). With the tongue moving toward the palate it is plausible that both the effective α decreases rapidly and L_2 increases. According to Eq. (7) these factors would provide a rapid decrease in $F1$, which would explain the large $F1$ sensitivity near [i]. Near [ɑ], on the other hand, all three speakers show small sensitivity of $F1$ to $K1$. It is plausible that near [ɑ], as $K1$ moves the tongue away from the pharynx, α still decreases, as does L_2, thus creating conditions for a small relative change in $F1$ according to Eq. (7). The fact that [æ] is also in a region of small $\nu_{11}^{\rightarrow}$ for JW11 and JW18 suggests that near [æ], the $K1$ of these two speakers tends to balance the changes in α and L_2 in a way similar to the way it balances them near [ɑ].

It was noted in Sec. III G that JW11 and JW18 had similar regions of large and small $F1$ sensitivities to $K1$ [Figs. 18(a) and 18(d)]. This agrees with geometric similarities in their $K1$ components. For both talkers, $K1$ possesses a high-low degree-of-freedom that is proportionately greater than the front-back degree-of-freedom, and the two degrees-of-freedom are of the same polarity, in that "high" belongs with "front" (Fig. 3). The rotated shading pattern for JW14 noted in Sec. III G [Fig. 18(b)] may be due to the fact that her $K1$ has much more of a front-back degree-of-freedom compared with the high-low degree-of-freedom.

Considering the $\nu_{21}^{\rightarrow}$ sensitivity patterns in Fig. 19, JW11 and JW14 both show large $F1$ sensitivity to $K2$ changes ($\nu_{21}^{\rightarrow}$) in a region around [i] [Figs. 19(a) and 19(b)], and JW18 just fails to include [i] in the region of large sensitivity [Fig. 19(d)]. All the speakers include [ɑ] and [æ] in their region of small sensitivity. The same perturbation-theory pictures explaining the sensitivity of $F1$ to $K1$ also seem to apply to the sensitivity of $F1$ to $K2$. JW11 and JW14 share the property that high corresponds to back in their $K2$ [Figs. 3(a) and 3(b)]. On the other hand, JW18 possesses the opposite correspondence between the high-low and front-back degrees-of-freedom (i.e., high corresponds to front) [Fig. 3(d)], and this could be the reason for the clockwise rotation of his regions of sensitivity that excludes [i] from his region of large sensitivity.

The sensitivities of $F2$ to both $K1$ and $K2$ ($\nu_{21}^{\rightarrow}$ and $\nu_{22}^{\rightarrow}$) have more fragmented sensitivity regions as well as more proximate regions of small and large sensitivity (Figs. 20 and 21). Some of the observed trends in $F2$ sensitivity can be explained in terms of changes in constriction size. Changes in constriction size produce relatively large changes in formant frequencies, as can be seen, for example, in perturbation theory. The high-back (palato-velar) regions have large $F2$ sensitivity to both $K1$ and $K2$, with the exception of JW16's $F2$ sensitivity to $K2$ [Fig. 21(c)]. The explanation for this consistently large sensitivity is that both high-low and front-back degrees-of-freedom will change constriction size in this region, thus substantially affecting $F2$ (see also Stevens, 1998, pp. 366–367). Whether or not "front" is coincident with "high" does not seem to matter in the palato-velar region. On the other hand, in the region of [ɑ], the two degrees-of-freedom do not have the same effect on constric-

tion degree. Thus, the reason that $F2$ has large sensitivity to $K1$ in this region for JW14 [Fig. 20(b)] but small sensitivity for JW11 and JW18 [Figs. 20(a) and 20(d)] has to do with the different proportions of high-low and front-back degrees-of-freedom in the different speakers' $K1$s. JW14's $K1$ has a larger proportion of front-back than the other two speakers, and thus her $K1$ affects constriction degree more than the $K1$s of JW11 or JW18. In the region of [i], the reason that high-front vowels have small $F2$ sensitivity to $K2$ in the cases of JW11 and JW14 [Figs. 21(a) and 21(b)] may be due to the $K2$ components of these speakers possessing a relatively large front-back degree-of-freedom with a small high-low degree-of-freedom polarized so that high corresponds to back. This means that $K2$ has the tongue make a tighter constriction while increasing the back cavity length; these two actions have opposite effects on $F2$ and so largely cancel each other out. Further, with high corresponding to back the constriction degree changes at the palate are minimal, and they are dominated by changes in the place of constriction as the tongue moves along the hard palate.

Concerning the "critical" $(K1,K2)$ coordinates defined in Sec. III G, the implications for quantal theory are not conclusive. We see only inconsistent inclusion of projected point vowels at critical coordinates (Table X). However, there are two weaknesses to our analysis of critical points: (1) defining the critical points in $K1,K2$ space and projecting the point vowels to that space may be less informative than an analysis with the critical points and vowels in full $(K1,K2,K3,K4)$ coordinates and (2) the regular 20×20 grid may not be fine enough for a good classification of critical points. Further, quantal theory applied to vowels involves consideration of place-of-constriction, whereas the current analysis does not distinguish between place- and degree-of-constriction.

V. CONCLUSION

A method for obtaining mappings from measured articulatory coordinates to measured acoustic coordinates and the inverses of these mappings has been proposed and examined in this paper. The method involves PCA and a local linear regression procedure named loess. Four speakers from the XRMB-SPD were analyzed using these procedures. PCA was performed on both their pellet coordinate data and their formant frequency data, and these PCA components were used as independent variables in the various forward and inverse mappings constructed using loess. Loess models were made more computationally efficient by pre-computing weighted least squares at the data points and interpolating regression parameters between data points. The parameters of the loess models, regression bandwidth and interpolation bandwidth, were optimized by k-fold cross-validation. By example, it was seen that the forward and inverse mappings were reasonable when viewed as mappings from the two independent variables with the largest PCA variance and averaged over the remaining one or two independent variables. There were wide distributions in the local regression parameters at the data points. Further, there is some evidence of co-variation among these parameters. A discussion of the relation between test RMSE of the loess models and the standard deviation of the error distributions revealed that the RMSEs of the forward mappings are probably a good estimate of the error standard deviations for formant data, but that the RMSEs of the inverse mappings are not a good estimate of error standard deviations for the articulatory data. There may be two fundamental problems with the inverse mappings: (1) there are not enough acoustic parameters to construct a good mapping and (2) the assumption that the independent variables are known with relatively little error is not a good assumption. One important application of loess is in the study of the sensitivity of acoustic parameters to changes in articulatory parameters. It was possible to find regions of large and small sensitivity as a function of the two lowest-order articulatory components when the slopes and formant values are averaged over the remaining two articulatory components.

There are future directions for research into this particular method of constructing empirically determined mappings between articulation and acoustics. The acoustic data were found to be lacking in the number of degrees-of-freedom and the accuracy with which they could be measured. Further, these data require some time consuming hand editing. Work in speech technology may provide some clues on where to look for improved acoustic variables, such as line spectral densities (Qin and Carreira-Perpiñán, 2007). An avenue that would make the results more easily interpretable would be to use the method of arbitrary factor analysis proposed by Maeda (1990). The high-low and front-back dimensions could be factored out of the articulatory data before PCA is performed on the residual. Another significant improvement in this method would be to be able to provide quantitative confidence intervals for both the mappings and the parameters in the mappings. This would require development of more efficient computational methods and merits further research. Finally, an analysis of local co-linearity among independent variables should be performed, and co-linearity should be removed, perhaps by allowing a reduction in the number of independent variables locally.

ACKNOWLEDGMENTS

The authors would like to thank Rebekka Puderbaugh for her help in analyzing the acoustic data. Thanks to Professor Joyce McDonough for making sure that this research proceeded and for her encouragement to the authors. This work was supported by NIDCD-001247 to CReSS LLC.

Atal, B. S., Chang, J. J., Mathews, M. V., and Tukey, J. W. (**1978**). "Inversion of articulatory-to-acoustic transformation in the vocal tract," J. Acoust. Soc. Am. **63**, 1535–1555.

Badin, P., Perrier, P., Boë, L.-J., and Abry, C. (**1990**). "Vocalic nomograms: Acoustic and articulatory considerations upon formant convergence," J. Acoust. Soc. Am. **87**, 1290–1300.

Beaudoin, R. E., and McGowan, R. S. (**2000**). "Principal component analysis of x-ray microbeam data for articulatory recovery," in *Proceedings of the Fifth Seminar on Speech Production* (Institute for Phonetics and Speech Communication, Munich, Germany), pp. 225–228.

Bickel, P. J., and Docksum, K. A. (**1977**). *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden-Day, San Francisco, CA).

Boersma, P., and Weenik, D. (**2007**). PRAAT: Doing phonetics by computer (Version 4.5.16) (computer program) (Last vewed February, 2007), from http://www.praat.org.

Bresch, E., Nielsen, J., Nayak, K., and Narayanan, S. (**2006**). "Synchronized

and noise-robust audio recordings during realtime magnetic resonance imaging scans," J. Acoust. Soc. Am. **120**, 1791–1794.
Chatterjee, S., and Hadi, A. S. (**1988**). *Sensitivity Analysis in Linear Regression* (Wiley, New York).
Cleveland, W. S. (**1979**). "Robust locally weighted regression and smoothing scatter plots," J. Am. Stat. Assoc. **74**, 829–836.
Cleveland, W. S., and Devlin, S. J. (**1988**). "Locally weighted regression: An approach to regression analysis by local fitting," J. Am. Stat. Assoc. **83**, 596–610.
Cleveland, W. S., Devlin, S. J., and Grosse, E. (**1988**). "Regression by local fitting: Methods, properties, and computational algorithms," J. Econometr. **37**, 87–114.
Engwall, O., and Badin, P. (**1999**). "Collecting and analyzing two- and three dimensional MRI data for Swedish," TMH-QPSR Report No. 40, KTH, Stockholm, Sweden.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Hogden, J., Rubin, P., McDermott, E., Katagiri, S., and Goldstein, L. (**2007**). "Inverting mappings from smooth paths through R^n to paths through R^m: A technique applied to recovering articulatory information from acoustics," Speech Commun. **49**, 361–383.
Jackson, M. T.-T. (**1988**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.
Jackson, M. T.-T., and McGowan, R. S. (**2008**). "Predicting midsagittal pharyngeal dimensions from measures of anterior tongue position in Swedish vowels: Statistical considerations," J. Acoust. Soc. Am. **123**, 336–346.
Kiritani, S. (**1986**). "X-ray microbeam method for the measurement of articulatory dynamics: Techniques and results," Speech Commun. **5**, 119–140.
Ladefoged, P., Harshman, R., Goldstein, L., and Rice, L. (**1978**). "Generating vocal tract shapes from formant frequencies," J. Acoust. Soc. Am. **64**, 1027–1035.
Lindblom, B. E. F., and Sundberg, J. E. F. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**, 1166–1179.
McGowan, R. S. (**2006**). "Perception of synthetic vowel exemplars of four year-old children and estimation of their corresponding vocal tract shapes," J. Acoust. Soc. Am. **120**, 2850–2858.
McGowan, R. S., and Cushing, S. (**1999**). "Vocal tract normalization for midsagittal articulatory recovery with analysis-by-synthesis," J. Acoust. Soc. Am. **106**, 1090–1105.
Maeda, S. (**1982**). "A digital simulation method of vocal-tract system," Speech Commun. **1**, 199–229.
Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 131–149.
Mermelstein, P. (**1967**). "Determination of the vocal-tract shape from measured formant frequencies," J. Acoust. Soc. Am. **41**, 1283–1294.
Mermelstein, P. (**1973**). "Articulatory model for the study of speech production," J. Acoust. Soc. Am. **53**, 1070–1082.
Mokhtari, P., Kitamura, T., Takemoto, H., and Honda, K. (**2007**). "Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients," J. Phonetics **35**, 20–39.
Perkell, J. S. (**1969**). *Physiology of Speech Production* (MIT Press, Cambridge, MA).
Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., and Jackson, M. (**1992**). "Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096.
Qin, C., and Carreira-Perpiñán, M. A. (**2007**). "A comparison of acoustic features of for articulatory inversion," in *Interspeech 2007*, pp. 2469–2472.
Schroeder, M. R. (**1967**). "Determination of the geometry of the human vocal tract by acoustic measurement," J. Acoust. Soc. Am. **41**, 1002–1010.
Shirai, K., and Honda, M. (**1976**). "An articulatory model and the estimation of articulatory parameters by nonlinear regression method," Electron. Commun. Jpn. **59**, 35–43.
Stevens, K. N. (**1972**). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by E. E. David, Jr. and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Story, B. H. (**2007**). "Time-dependence of vocal tract modes during production of vowels and vowel sequences," J. Acoust. Soc. Am. **121**, 3770–3789.
Story, B. H., and Titze, I. R. (**1998**). "Parameterization of vocal tract area functions by empirical orthogonal modes," J. Phonetics **26**, 223–260.
Westbury, J. R. (**1983**). "Enlargement of supraglottal cavity and its relation to stop consonant voicing," J. Acoust. Soc. Am. **73**, 1322–1336.
Westbury, J. R. (**1994**). *X-Ray Microbeam Speech Production Database User's Handbook* (University of Wisconsin, Madison, WI).
Whalen, D. H., Kang, A. M., Magen, H. S., Fulbright, R. K., and Gore, J. C. (**1999**). "Predicting midsagittal pharynx shape from tongue position during vowel production," J. Speech Lang. Hear. Res. **42**, 592–603.
Wood, S. (**1979**). "A radiological analysis of constriction locations for vowels," J. Phonetics **7**, 25–43.
Yehia, H., and Itakura, F. (**1996**). "A method to combine acoustic and morphological constraints in the speech production inverse problem," Speech Commun. **18**, 151–174.

A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning

Stéphanie Buchaillard[a] and Pascal Perrier
ICP/ GIPSA-Lab, UMR CNRS 5216, Grenoble INP, 38402 Saint Martin d'Hères, France

Yohan Payan
TIMC-IMAG, UMR CNRS 5525, Université Joseph Fourier, 38706 La Tronche, France

(Received 18 March 2008; revised 21 July 2009; accepted 23 July 2009)

A three-dimensional (3D) biomechanical model of the tongue and the oral cavity, controlled by a functional model of muscle force generation (λ-model of the equilibrium point hypothesis) and coupled with an acoustic model, was exploited to study the activation of the tongue and mouth floor muscles during the production of French cardinal vowels. The selection of the motor commands to control the tongue and the mouth floor muscles was based on literature data, such as electromyographic, electropalatographic, and cineradiographic data. The tongue shapes were also compared to data obtained from the speaker used to build the model. 3D modeling offered the opportunity to investigate the role of the transversalis, in particular, its involvement in the production of high front vowels. It was found, with this model, to be indirect via reflex mechanisms due to the activation of surrounding muscles, not voluntary. For vowel /i/, local motor command variations for the main tongue muscles revealed a non-negligible modification of the alveolar groove in contradiction to the saturation effect hypothesis, due to the role of the anterior genioglossus. Finally, the impact of subject position (supine or upright) on the production of French cardinal vowels was explored and found to be negligible.
 [DOI: 10.1121/1.3204306]

PACS number(s): 43.70.Bk, 43.70.Aj [AL] Pages: 2033–2051

I. INTRODUCTION

Speech movements and acoustic speech signals are the results of the combined influences of communicative linguistic goals, perceptual constraints, and physical properties of the speech production apparatus. To understand how these different factors combine and interact with each other requires an efficient approach that develops realistic physical models of the speech production and/or speech perception systems. The predictions of these models can then be compared with experimental data, and used to infer information about parameters or control signals that are not directly measurable or the measurement of which is difficult and not completely reliable. Such a methodological approach underlies the present work, in which a biomechanical model of the vocal tract has been used to study muscle control in vowel production, its impact on token-to-token variability, and its consequences for tongue shape sensitivity to changes in head (supine versus upright) orientation. The findings are interpreted in the light of our own experimental data and data published in the literature.

Biomechanical models of the tongue and vocal tract have been in use since the 1960s, and their complexity has increased with the acquisition of new knowledge about anatomical, neurophysiological, and physical characteristics of the tongue, as well as with the vast growth in the computational capacities of computers. All these models have significantly contributed to the increase in knowledge about tongue behavior and tongue control during speech production, and more specifically about the relations between muscle recruitments and tongue shape or acoustic signal (see, in particular, Perkell, 1996, using his model presented in Perkell, 1974; Kakita *et al.*, 1985; Hashimoto and Suga, 1986; Wilhelms-Tricarico, 1995; Payan and Perrier, 1997; Sanguineti *et al.*, 1998; Dang and Honda, 2004). With a more sophisticated three-dimensional (3D) vocal tract model, based on nonlinear continuum mechanics modeling, and taking into consideration a number of recent experimental findings, this study aims at deepening and extending these former works for vowel production.

The model consists of a 3D biomechanical model of the tongue and the oral cavity, controlled by a functional model of muscle force generation [λ-model of the equilibrium point hypothesis (EPH)] and coupled with an acoustic model. It is a significantly improved version of the model originally developed in GIPSA-Lab by Gérard and colleagues (Gérard *et al.*, 2003, 2006). The oral cavity model was developed so as to give as realistic a representation as possible of the anatomy and of the mechanical properties of the oral cavity. The original modeling was based on the data of the Visible Human Project, and further adapted to the anatomy of a specific subject. For this subject, different kinds of data [x-ray, computed tomography (CT) images, and acoustic data] were available. The parameters used in this model were either extracted from the literature, derived from experimental data, or adapted from the literature. This modeling study is inseparable from a thorough experimental approach. In addition to

[a]Author to whom correspondence should be addressed. Electronic mail: stephanie.buchaillard@gmail.com

a careful and accurate account of anatomical, mechanical, and motor control facts, the model implements a number of hypotheses about the hidden parts of the speech production system. Simulation results, their interpretation, and the corresponding conclusions aim at opening new paths for further experimental research that could validate or contest these conclusions.

The main characteristics of the model (geometry, mechanical properties, and model of control) are presented in Sec. II. The model includes improvements in the anatomical and morphological descriptions and in the strain/stress function, as well as a control model of muscle activation (Sec. II). The model is first used (Sec. III) in order to characterize the muscle activation patterns associated with the production of the French cardinal vowels. Starting from these patterns, the relation between internal muscle strain and muscle activations is systematically studied. In Sec. IV, the sensitivity of the postural control of the tongue (and hence of the formant frequencies) to changes in motor commands is precisely studied for /i/, which is often described in the literature as a very stable vowel due to specific combinations of muscle activations. Finally, the impact on tongue positioning of changes in gravity orientation is assessed (Sec. V). Perspectives and further developments are discussed in Sec. VI.

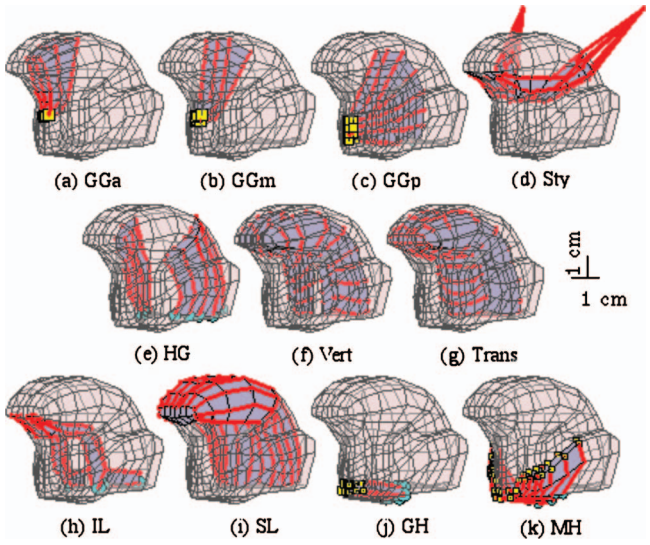


FIG. 1. Mesh representation (gray elements) of lingual and mouth floor muscles as subsets of tongue elements (global mesh) (anterior oblique view). [(a)–(c)] anterior, medium, and posterior part of the genioglossus, (d) styloglossus, (e) hyoglossus, (f) verticalis, (g) transversalis, (h) inferior longitudinalis, (i) superior longitudinalis, (j) geniohyoid, and (k) mylohyoid. The muscle fibers are represented in red. The yellow squares and the blue dots represent the muscle insertions on the mandible and the hyoid bone, respectively.

II. MODELING THE ORAL CAVITY

Modeling the oral cavity by a finite element approach requires meshing the structure of interest, specifying its mechanical properties and defining a motor control scheme. Then, the simulation of movements in response to motor commands requires solving the body motion equations. These different aspects will be described in this section.

The primary goal of our work is the development of a model which allows a better understanding of how motor control and physical aspects combine and interact to determine the characteristics of speech production signals. Hence, a high degree of realism is essential in the design of the model, not only concerning the geometrical properties but also the mechanical and control aspects.

The model described below is an improved version of the model developed by Gérard and colleagues (Gérard *et al.*, 2003, 2006). The original model was based on the Visible Human Project® data for a female subject and the work of Wilhelms-Tricarico (2000). It was then adapted to a specific male subject, PB henceforth. Major differences between the current version and those of Gérard and colleagues (Gérard *et al.*, 2003, 2006) lie in (1) the motor control scheme (muscle forces are now computed via the λ-model of the EPH), (2) the constitutive law for the tongue tissues [the law inferred by Gérard *et al.* (2005) from indentation measurements of fresh cadaver tissues was modified to match the properties of living tissues; in addition, the law now depends on the level of muscle activation], (3) the modeling of the hyoid bone (a new scheme was also developed to deal with hyoid bone mobility and to model the infrahyoid and digastric muscles). Modifications were also made to the tongue mesh, the muscle fibers, the bony insertions, and the areas of contact between the tongue and the surrounding surfaces, namely, the mandible, the hard palate, and the soft palate. The 3D vocal tract model was also coupled with an acoustic model.

A. Geometrical and anatomical structures

A precise description of the tongue anatomy will not be given here. A thorough description, which lies at the root of this work, can be found in Takemoto (2001). The tongue model represents the 3D structure of the tongue of a male subject (PB), for whom several sets of data have been collected in the laboratory in the past 15 years. This model is made of a mesh composed of hexahedral elements. The anatomical location of the major tongue muscles is specified via subsets of elements in the mesh. Figure 1 shows the implementation of the 11 groups of muscles represented in the model and known to contribute to speech production. Nine of them exert force on the tongue body itself, while the other two, depicted in the last two panels [Figs. 1(j) and 1(k)], are considered to be the major mouth floor muscles. Of course, due to the elastic properties of tongue tissues, each muscle is likely to induce strain in all the parts of the tongue and mouth floor. On rare occasion the muscle shape is somewhat unnatural because the tongue muscles were defined as a subset of elements of the global mesh. This is, for example, the case with the inferior longitudinalis (IL). However, when activated, the force generated by the IL appeared correct in amplitude and direction. The insertion of the different parts of the geniglossus on the mandible can also appear odd: in human beings, GGp emanates from the lower surface of the short tendon that reduces crowding of the fibers at the mandibular symphysis by allowing GGp to arise from below and the radial fibers to arise from above. In the model, the tendon is not represented and the origins of GGp, GGm, and GGa

are all on the mandibular symphysis. This results in a somewhat too large region of insertion on the mandible. Only a refined mesh structure would allow a better muscle definition in this area.

It is generally accepted that a muscle can possibly be divided into a number of functionally independent parts. For tongue muscles this possibility exists, but little work has been done in the past concerning this issue. Some proposals were the results of *ad hoc* choices made in order to explain measured two-dimensional (2D) or 3D tongue shapes (e.g., the most recent proposal for the styloglossus in Fang *et al.*, 2008). Some more physiologically based studies used electromyographic (EMG) signals, generally assuming that these signals reflect the underlying motor control. Among these studies, the one carried out by Miyawaki *et al.* (1975) showed evidence for different activities in different parts of the genioglossus. However, EMG activity is the result of a combination of efferent and afferent influences and it cannot be seen as a direct image of the underlying control. In addition, as emphasized by Miyawaki *et al.* (1975), if subdivisions exist in a muscle, we do not know in what manner they are voluntarily controlled (p. 101). We believe that the only reliable way to address this issue would be to look at the motor unit distribution within tongue muscles. To our knowledge, we lack information on the localization of motor unit territories in human tongue muscles. One way to know more about it could be to study the architecture of the muscles, with the underlying hypothesis that structurally separated muscle parts could be innervated by independent motor units. Slaughter *et al.* (2005) carried out such a study for the human superior longitudinalis (SL), and they found that this muscle consists of a number of in-series muscle bundles that are distributed along the front-back direction. However, they could not provide clear evidence for the fact that these muscle bundles are innervated by independent motor units. In the absence of convincing physiological evidence, and in order to limit the complexity of the model, only the genioglossus, for which a consensus seems to exist, was subdivided: three independent parts called the GGa (anterior genioglossus), the GGm (medium genioglossus), and the GGp (posterior genioglossus) were thus defined.

To mesh the hard and soft structures forming the oral cavity, data of different kinds such as CT scans, MRI data, and x-ray data, all collected for PB, were exploited. In addition to the tongue and mouth floor meshes, the model (Fig. 2) includes a surface representation of the mandible, the soft palate, the hard palate, and the pharyngeal and laryngeal walls as well as a volumetric mesh (tetrahedral elements) of the hyoid bone. A set of six pairs of springs (right and left sides), emerging from the hyoid bone, are used to represent the elastic links between this mobile bone and fixed bony structures associated with the anterior and posterior belly of the digastric, infrahyoid muscles (sternohyoid, omohyoid, and thyrohyoid muscles), as well as the hyo-epiglottic ligaments.

The relative positions of the different articulators were carefully adjusted so as to represent well PB's morphology in a seated position and at rest, just as they are described by lateral x-ray views of PB's oral cavity. The final tongue shape in the midsagittal plane at rest was also adapted so as to match the corresponding x-ray view. This induced some geometrical changes to the original shape proposed in Gérard *et al.* (2006), because the MRI data used for the original design of that model corresponded to the subject in the supine position; gravity was then shown in that case to influence tongue shape.

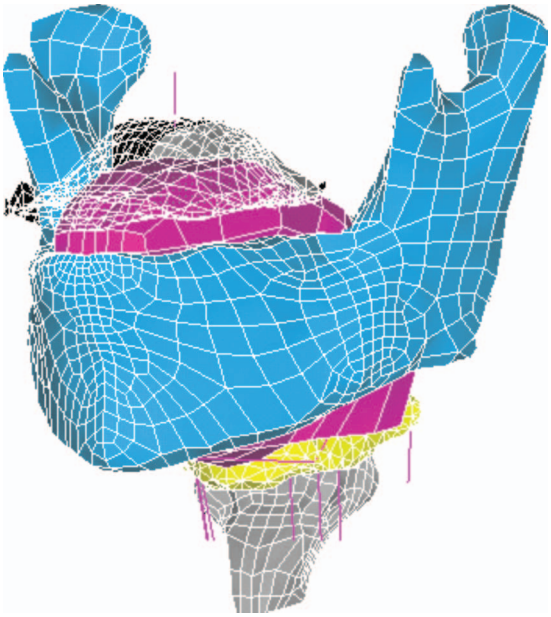


FIG. 2. Oblique anterior view of the 3D tongue mesh in the whole oral cavity for a rest position (tongue mesh in magenta, mandible in cyan, hyoid bone in yellow, translucent soft palate, pharyngeal and laryngeal walls in gray, infra- and supra-hyoid muscles represented as magenta lines).

B. Mechanical properties

The lingual tissues were modeled with a non-linear hyper-elastic constitutive law, more precisely a second order Yeoh constitutive law (Gérard *et al.*, 2005, 2006). Two different constitutive equations were introduced: one describes the passive behavior of tongue tissues and the other one models the strain/stress relation for active muscle tissues as an increasing function of muscle activation. For a particular mesh element, the passive or the active constitutive law is used according to whether this element belongs to a passive or to an active region [i.e., a region made of activated muscle(s)]. The passive constitutive law was directly derived from the non-linear law proposed by Gérard *et al.* (2005), which was derived from measurements on a fresh cadaver. However, since the stiffness of tissues measured shortly after death is known to be lower than that measured in *in vivo* tissues, the constitutive law originally proposed by Gérard *et al.* (2005) was modified.

To our knowledge, one of the most relevant *in vivo* measurements of human muscle stiffness is the one carried out by Duck (1990), who proposed a value of 6.2 kPa for the Young's modulus for a human muscle at rest and a maximum value of 110 kPa for the same muscle once contracted. The Young's modulus measured by Gérard *et al.* (2005) on a cadaver tongue at low strain is 1.15 kPa, which is significantly smaller than Duck's (1990) *in vivo* measurements.

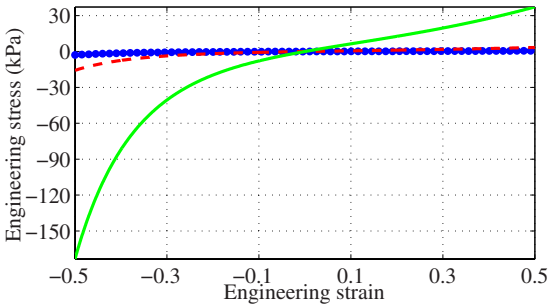


FIG. 3. (Color online) Stress/strain hyperelastic constitutive law (Yeoh second order material) for lingual tissues. The dotted curve represents the original law obtained from fresh cadaver tissues (c_{10}=192 Pa and c_{20}=90 Pa), the dashed curve represents the law used in the current model for passive tissues (c_{10}=1037 Pa and c_{20}=486 Pa), and the solid line represents the law used for the maximal activation (c_{10}=10.37 kPa and c_{20}=4.86 kPa).

This difference is not surprising, since in living subjects a basic muscle tonus exists, even at rest. Hence, it was decided to multiply both second order Yeoh law coefficients originally proposed by Gérard *et al.* (2005) by a factor of 5.4, in order to account properly for the Young's modulus at rest measured by Duck (1990). Multiplying both coefficients by the same factor allows preservation of the overall non-linear shape of the Yeoh constitutive law (Fig. 3). This new law specifies the properties of passive tongue tissues. In order to account for the stiffening associated with muscle activation as measured by Duck (1990), it was decided for the elements belonging to an activated muscle to multiply the coefficients of the Yeoh constitutive law for passive tissues by a factor that is a function of muscle activation. Thus, an activation-related constitutive law was defined for the active muscles. The multiplying factors were chosen by taking into account the fact that the contributions of the different muscles to the Young's modulus of an element combine in an additive manner. The basic idea is that an activation of a muscle leads to an increase in its Young's modulus. Given $c_{10,r}$ and $c_{20,r}$ the Yeoh parameters for tongue tissues at rest, the parameters $c_{i0}(e,t)$ $(i \in \{1,2\})$ at time t for an element e belonging to the tongue or mouth floor, are given by

$$c_{i0}(e,t) = c_{i0,r}\left(1 + \sum_{\text{muscles } m} p_1(m)\left[\sum_{\text{fibers } f \in m} A(f,t)p_2(f,e)\right]\right), \quad (1)$$

where p_1 is a positive muscle-dependent factor, $A(f,t)$ is the activation level for the macrofiber f at time t [see Eq. (2) below], and $p_2(f,e)$ is a factor equal to 1 if e belongs to m and if the fiber f runs along the edges of e, 0 otherwise.

The multiplying factor p_1 was chosen in order to maintain the stiffness value below 110 kPa, when maximal muscle activation is reached.

Since tongue tissues are considered to be quasi-incompressible, a Poisson coefficient equal to 0.499 was used. Furthermore, tongue tissue density was set to 1040 kg m^{-3}, close to water density.

Currently, only the tongue and the hyoid bone (with the springs connecting it to fixed bony structure) are modeled as movable structures and need to be mechanically characterized. The hyoid bone was considered as a rigid body and its density (2000 kg m^{-3}) was estimated based on values published in the literature (Dang and Honda, 2004). The same stiffness coefficient (220 N m^{-1}) was chosen for all the springs connecting the hyoid bone to solid structures; this value enabled us to reproduce displacements of the hyoid bone that were consistent with data published in Boë *et al.* (2006).

C. Motor control: Implementation of postural control with short latency feedback

The motor control scheme implemented is based on the λ version of the EPH (Feldman, 1986). This theory is known to be controversial in the motor control domain. The main criticisms are about the fact that this theory claims that the time variation of motor control variables does not result from any inverse kinematics or inverse dynamics processes (see, for example, Gomi and Kawato, 1996 or Hinder and Milner, 2003). However, the defenders of the EPH theory have systematically provided refutations of these criticisms that support the value of the model in research (e.g., in Gribble and Ostry, 1999 or Feldman and Latash, 2005). Our own work has also shown that speech motor control based on the EPH gives a good account of complex kinematic patterns with a 2D biomechanical model of the vocal tract (Payan and Perrier, 1997; Perrier *et al.*, 2003). From our point of view, this motor control theory seems particularly interesting for speech production because it provides the framework for a discrete characterization of continuous physical signals at a motor control level, thanks to the link that can be made between successive equilibrium points and targets; it thus allows a connection to be made between the discrete phonological units and the physical targets that underlie continuous articulatory and acoustic signals (Perrier *et al.*, 1996). In addition, the EPH integrates short latency feedback to contribute to the accuracy of speech gesture, which is for us a crucial feature for speech production control (Perrier, 2006). Hence, the approach used in our previous modeling work with the 2D biomechanical model of the vocal tract was extended to the 3D model.

1. Adjustment of feedback delay

The implementation chosen for the EPH follows the approach proposed by Laboissière *et al.* (1996) and further developed by Payan and Perrier (1997). In the model, bundles of fibers are represented by way of *macrofibers* (specified as ordered lists of mesh nodes along the edges of elements) that represent the main directions of muscle fibers in the different parts of the tongue. In the current version of the model, a unique activation threshold was defined for each muscle (three for the genioglossus, which was divided into three parts that are assumed to be separately controlled: the anterior, posterior, and medium parts). Every muscle was assumed to be controlled independently. Obviously, synergies and antagonisms exist in tongue muscles. However, there is no evidence in the literature supporting the hypothesis that these muscle coordinations are implemented in humans from birth. It is much more likely that coordinated muscle activations are the result of learning and that they could be task

TABLE I. Cross-sectional areas and corresponding force generation capacities (ρ).

	GGa	GGm	GGp	Sty	HG	Vert	Trans	IL	SL	GH	MH
Area (mm^2)	82	55	168	109	295	91	227	41	86	80	177
(ρ) (N)	18	12	37	24	65	20	50	9	19	17.5	39

specific. Our modeling approach is in line with this statement. The design of our biomechanical model gives the largest possible number of degrees of freedom to the system to be controlled and does not impose a priori hypotheses that could bias our study. It allows future work on the emergence of muscle coordinations through task specific learning. For each muscle, the motor command λ_{muscle} was determined for the longest macrofibers l_{max}; the λ value for each macrofiber of the same muscle was then determined by simply multiplying the λ_{muscle} value by the ratio of the macrofiber length at rest over l_{max}.

For a given macrofiber, the muscle activation A takes into account the difference between the macrofiber length and the motor command λ, as well as the lengthening/shortening rate.

A stretch reflex delay d, which corresponds to the propagation delay for the electrical signals to travel along the reflex arc plus the synaptic time and the integration time of these signals at the interneurons, is taken into account for fiber length and velocity intervening in the computation of A. In their model of the mandible, Laboissière *et al.* (1996) proposed a delay of 10 ms, and in their tongue/jaw model Sanguineti *et al.* (1998) suggested a delay of 15 ms. In the present model, d was set to 17 ms, based on the data of Ito *et al.* (2004). Simulations conducted for d ranging from 5 to 20 ms showed that this value had a limited impact on tongue motion; the trajectory, peak velocity, acceleration, or force levels were altered, but in a limited range so that the choice of this value did not seem to be critical within this range of variation. The sensitivity of the activation to the lengthening/shortening rate $\dot{l}$ is modulated by a damping coefficient μ, considered for the sake of simplicity as constant and identical for all the muscles. μ was chosen to be equal to 0.01 s to ensure the stability of the system, following numerous simulations

$$A(t) = [l(t-d) - \lambda(t) + \mu\dot{l}(t-d)]^{+}. \quad (2)$$

Muscle activation is associated with the firing of the motoneurons (henceforth MNs). Hence A is either positive or zero (if A is mathematically negative, it is set to zero). A zero value corresponds to the MN fire threshold; beyond this threshold, the MN depolarization becomes possible: the higher the activation A, the higher the firing frequency of MNs. As long as the activation A is zero, no force is generated. Force varies as an exponential function of the activation (see below).

2. Feedback gain: A key value for postural control stability

Active muscle force $\tilde{M}$ is given as a function of the activation $A(t)$ by the following equation:

$$\tilde{M}(t) = \max[\rho(\exp^{cA(t)} - 1), \rho], \quad (3)$$

with ρ a factor related to the muscle capacity of force generation and c a form parameter symbolizing the MN firing gradient.

The determination of the parameter ρ, which modulates the force generation capacity, is based on the assumption that, for a fusiform muscle, ρ is linked in a first approximation to the cross-sectional area of the muscle. The values are based on the work of Payan and Perrier (1997) for the tongue muscles, except for the transversalis, which was non-existent in a 2D tongue model, and with some adaptations for the verticalis, the implementation of which was slightly different. For the mouth floor muscles, ρ values were estimated from the data of van Eijden *et al.*, 1997, and were measured on the model for the transversalis. This muscle force capacity (Table I) was distributed among the different macrofibers proportionally to the volume of the surrounding elements. Given a fiber f belonging to a muscle m, its capacity of force generation ρ_{fib} is such that

$$\rho_{fib}(f) = \rho(m)\frac{\Sigma_e V(e) \times p(e,f)}{S}, \quad (4)$$

where e is an element belonging to m, $V(e)$ is the volume of e, $p(e,f)$ is a parameter equal to 1 if f is located inside the muscle, 0.5 on a muscle face (exterior surface of a muscle excluding muscle corners), and 0.25 on a muscle edge (exterior surface of a mesh, corners only). S is a normalization term, such that the ρ_{fib} values for the different fibers of m sum up to $\rho(m)$.

Parameter c is an important factor for stability issues since it determines how feedback information included in the activation influences the level of force. Original values for c found in the literature ($c = 112$ m^{-1}, Laboissière *et al.*, 1996) brought about dramatic changes in the muscular activation level for a small variation in the muscle length. This generated mechanical instabilities. Therefore, parameter c was decreased. After several trials, c was fixed to 40 m^{-1}. This value is not the only one that ensured a stable mechanical behavior of the model. A large range of values was possible. The value 40 m^{-1} was chosen because it provides a fair compromise between the level of reflex activation and stability (Buchaillard *et al.*, 2006).

The influence of muscle lengthening/shortening velocity on the force developed is also included. The model accounts for the sliding filaments theory (Huxley, 1957) by calculating the total muscle force F with the following equation (Laboissière *et al.*, 1996):

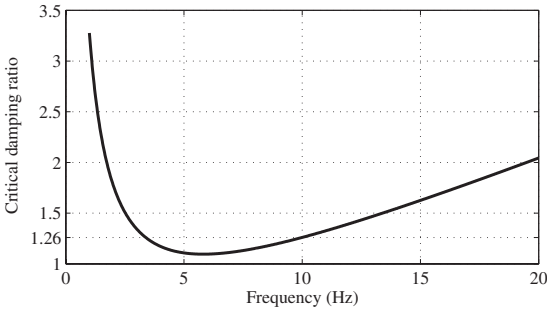


FIG. 4. Ratio damping over critical damping versus frequency for $\alpha = 40\ s^{-1}$ and $\beta = 0.03$ s (Rayleigh damping model). For a modal frequency from 3 to 10 Hz, the damping ratio is below 1.26, i.e., close to the critical damping (ratio equal to 1).

$$F(t) = \tilde{M}(t)\left(f_1 + f_2 \arctan\left(f_3 + f_4\frac{\dot{l}(t)}{r}\right) + f_5\frac{\dot{l}(t)}{r}\right), \qquad (5)$$

where $\dot{l}$ is the lengthening/shortening velocity and r is the muscle length at rest. The parameters used are based on the work of Payan and Perrier (1997) for rapid muscles, but are slightly different: $f_1 = 0.7109$, $f_2 = 0.712$, $f_3 = 0.43$, $f_4 = 0.4444$ s, and $f_5 = 0.0329$ s.

D. Lagrangian equation of motion and boundary conditions

The Lagrangian equation of motion that governs the dynamic response of the finite element system is given by

$$\mathbf{M}\ddot{\vec{q}} + \mathbf{C}\dot{\vec{q}} + \mathbf{K}\vec{q} = \vec{F}, \qquad (6)$$

where $\vec{q}$ is the nodal displacements vector, $\dot{\vec{q}}$ and $\ddot{\vec{q}}$ are its first and second derivatives, **M** is the mass matrix, **C** is the damping matrix, **K** is the stiffness matrix, and $\vec{F}$ is the load vector (the reader can refer to Bathe, 1995 for a detailed description of the finite element method).

A Rayleigh damping model was chosen for the definition of the damping matrix: $\mathbf{C} = \alpha\mathbf{M} + \beta\mathbf{K}$. α and β were set to 40 s^{-1} and 0.03 s, respectively, in order to have a damping close to the critical one in the range of modal frequency from 3 to 10 Hz (Fig. 4).

The load vector $\vec{F}$ includes the muscle forces computed for every macrofiber [Eq. (5)], the gravity and contact forces between tongue and vocal tract walls.

Two kinds of boundary conditions were introduced through the definition of no-displacement constraints to model muscular insertions and the management of contacts. Muscle insertions on the bony structures (inner anterior and lateral surface of the mandible and hyoid bone) were implemented and they match as well as possible the information about PB's anatomy that was extracted from x-ray scans. During speech production, the tongue comes into contact with the hard and soft tissues that compose the vocal tract walls. Consequently, the contacts were modeled between the tongue and the set hard palate/upper dental arch, the soft palate, and the set inner surface of the mandible/lower dental arch. The modeling of contacts is non-linear. A face-to-face detection was used to avoid the interpenetration of the surfaces in contact, which are potentially in contact with the tongue. A relatively low Coulomb friction was used, since friction is assumed to be limited due to the saliva. The contacts are managed through an augmented Lagrangian method, which corresponds to an iterative series of penalty methods.

The partial differential equation (6) was solved by the ANSYS™ finite element software package, based on a combination of Newton–Raphson and Newmark methods.

E. Acoustic modeling

A model of sound synthesis, including the determination of the 3D area function of the vocal tract, was coupled with the mechanical model.

The computation of the area function from the mesh node coordinates was achieved by using MATLAB® software. Before computing the area function, the surface of the tongue was interpolated using 35 periodic cubic splines in order to get a more accurate detection of the constriction locations in the vocal tract. This processing and its use for the computation of the area function make the implicit assumption that the spatial sampling of the tongue surface provided by the finite element mesh is sufficient to allow a correct interpolation of the tongue surface from the positions of the nodes. A set of planes, which will be referred to as cutting planes below, was computed for the vocal tract in its rest position. These cutting planes, orthogonal to the sagittal plane, are approximately perpendicular to the vocal tract midline at rest. For a given vocal tract configuration, the intersections between the cutting planes and the surface of the tongue (approximated by a set of periodic cubic splines), of the mandible, of the hyoid bone, of the hard and soft palates, and of the pharyngeal and laryngeal walls were computed. On every cutting plane, a closed contour based on these intersections and representing the shape of the vocal tract was computed and approximated by periodic cubic splines. The inner surface of each of the thus-determined closed contours was calculated. The lips, which are not part of the biomechanical model, were represented by a single cylinder, whose length and section represented lip protrusion and aperture, respectively. To determine the distance between two consecutive cutting planes, and thus to compute the length of the path from the glottis to the lips, it was decided to compute the distance between the centers of gravity of two successive surfaces. This distance approximates the average distance traveled by the acoustic wave between two consecutive cutting planes. An acoustic model (analog harmonic of the vocal tract) was used to generate the spectrum of the signal produced from the area function.[1]

III. MUSCLE ACTIVATIONS DURING FRENCH ORAL VOWEL PRODUCTION

A. Muscle activations

To study the postural control of speech sounds, the best approach would consist of roaming the motor command space of the biomechanical model in a systematic and comprehensive way, using, for example, a Monte Carlo method, in order to characterize the links between motor commands, tongue shapes, and acoustics, following the approach of Per-

rier *et al.* (2005) for their 2D model. However, such an approach is currently impossible with this 3D model, because of the running time (around 40 min for a 100 ms simulation, with ANSYS™ 11.0 and Windows XP SP2 running on a Pentium IV CPU at 3 GHz and 1 GB of RAM). Consequently, it was necessary to work with a more limited number of simulations to study muscle activations in vowel production and the sensitivity of the vowel configurations to changes in motor commands. The results presented in this section were obtained based on the 300 simulations, all carried out with a fixed mandible. These simulations resulted from a specific choice of motor commands guided primarily by studies with our model of the individual impact of each muscle on tongue shape (see below). Our objectives were to generate a very good match of the tongue shapes classically observed in the midsagittal plane for French oral vowels by means of cineradiographic data (Bothorel *et al.*, 1986). EMG studies by Miyawaki *et al.* (1975) and Baer *et al.* (1988) were also used as sources of complementary information on the main tongue and mouth floor muscles activated during vowel production. Acoustic signals were synthesized from the final vocal tract shape, and the formants were calculated.

The selection of the optimal vowels has involved a mostly qualitative evaluation of the similarity between the computed tongue shapes and the 3D tongue shapes measured for the speaker PB (CT data). A quantitative comparison of the simulated tongue shapes with the measured 3D shapes was not possible and would not have been very informative, mainly for two reasons:

(1) In our vocal tract model, the jaw is fixed. It is known that a variety of jaw positions is possible for the same sound without endangering the quality of its perception, and, in particular, producing speech with a fixed jaw does not prevent the speakers from producing satisfactory vowels with fair formants, as shown by bite block experiments (Mooshammer *et al.*, 2001); however, this articulatory perturbation has an impact on the tongue shape considered in its entirety.
(2) The model is a symmetrical one while human subjects are never symmetrical. Hence a detailed comparison of the constriction shape was not possible. This is why our simulations were essentially assessed in terms of global tongue elevation, proximity to the palate, and front/back position of the constriction in the vocal tract. However, a quantitative evaluation of the simulated and measured formant patterns was carried out.

Only the simulations obtained for the extreme vowels /i, a, u/ will be presented in this paper. The results correspond to the shape and position of the tongue at the end of the simulated movement. For single muscle activations, movement lasted 400 ms while it lasted only 200 ms for the vowels (for the three vowels, steady-state equilibrium positions were reached). In all cases, the movement started from rest position.

1. Impact of individual muscles on tongue shape

Figures 5 and 6 show the individual impact of the tongue and mouth floor muscles on the tongue shape in the midsagittal plane and in the 3D space from a front view perspective. Target motor commands were defined such that a single muscle was activated during each simulation. For the only activated muscle, the command (i.e., the threshold muscle length above which active muscle force is generated) was set either to 75% or 85% of the muscle length at rest (the smaller the percentage, the larger the activation; hence, a larger percentage was chosen for larger muscles to avoid too strong deformations). For the other muscles, the motor commands were set to a large enough value so as to prevent these muscles from generating forces; for example, a command twice as large as the muscle length at rest ensures that this muscle will remain inactive throughout a simulation. These simulations show that the role of the individual muscles in our model matches well with classic knowledge inferred from experimental data and clarify their impact on the tongue shape. The anterior genioglossus moves the tongue downward in its front part, essentially in the region close to the midsagittal plane (tongue grooving in the palatal region $\approx$6 mm). This downward movement is associated with a slight backward movement in the pharyngeal region [Fig. 5(a), $\approx$1.6 mm]. Note that the backward movement is much smaller than the one predicted by 2D (Payan and Perrier, 1997) or 2.5D (Dang and Honda, 2004) models. This can be explained by the fact that in these models the volume conservation is, in fact, implemented as a surface conservation property in the midsagittal plane. In our model, volume conservation causes the changes that are generated in one part of the tongue to be compensated not only in the other parts of the midsagittal plane but also in the whole tongue volume. Indeed, a slight enlargement of the tongue is observed in the transverse direction (up to 2.2 mm). It can also be noticed that the limited backward motion is consistent with data showing that a larger expansion may occur in the transverse plane, local to the compression, while a small expansion occur in the same plane (Stone *et al.*, 2004). The medium part of the genioglossus lowers the tongue in its dorsal region ($\approx$5 mm) and moves the apical part forward ($\approx$3 mm) and upward, while an enlargement of the tongue is observed in the transverse direction (up to 1.4 mm). The GGp enables the tongue to be pushed forward ($\approx$5.3 mm); this forward movement is associated with an elevation of the tongue ($\approx$2.4 mm) due to the apex sliding on the anterior part of the mandible [Fig. 5(c)]. However, the elevation of the tongue is less strong than what was predicted by the 2D and the 2.5D models. As for the GGa, it leads to the enlargement of the tongue in the transverse direction [$\approx$1.3 mm in the apical area and 1.6 mm in the pharyngeal area, Fig. 6(c)]. The styloglossus (Sty) causes downward ($\approx$9.6 mm) and backward ($\approx$7 mm) displacements of the tongue tip, producing an elevation of the dorsal part of the tongue and a lowering of the apical region [Figs. 5(d) and 6(d)]. No change is observed in the transverse direction. Changes in the midsagittal plane are similar to the predictions of 2D or 2.5D models. The hyoglossus generates a backward movement in the pharyngeal part ($\approx$5 mm), an apex elevation (upward displacement of $\approx$4.7 mm), and a lowering of the tongue in its dorsal part [Fig. 5(e)]. An enlargement is observed in the transverse direction in the pharyngeal part [Fig. 6(e), $\approx$5 mm]. The ver-

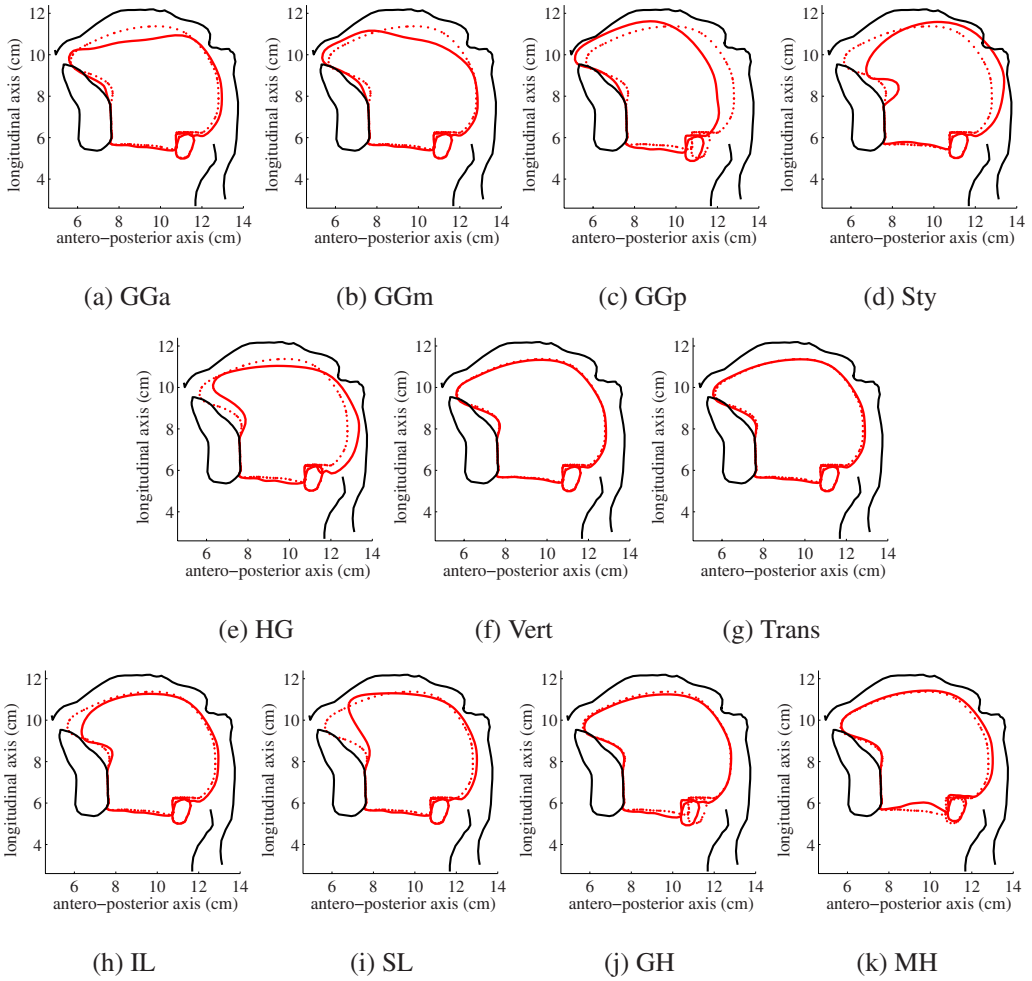


FIG. 5. (Color online) Impact of the activation of individual lingual and mouth floor muscles on tongue shape (400 ms command duration, sufficient to reach mechanical equilibrium). The contours of the articulators (tongue and hyoid bone, mandible, hard and soft palates, and pharyngeal and laryngeal walls) are given in the midsagittal plane (tongue tip on the left). For every simulation, the target motor command of the only activated muscle equals 75% of the muscle length at rest, except for the activation of the long muscles Sty, IL, and SL (85% of the muscle length at rest). The dotted contours correspond to the tongue shape in its rest position.

ticalis provokes only a very small lowering in the palatal region (below ≈0.5 mm) associated with a very slight backward movement in the pharyngeal part (below ≈0.5 mm) [Fig. 5(f)]. Its contraction also widens the tongue (≈1.4 mm in the apical area). Its impact will then be essentially indirect: by stiffening its elements in the palatal part of the tongue, it will modify the action of other muscles. The transverse muscle induces essentially a reduction in the tongue width in the transverse direction [up to 2.1 mm in the superior part of the tongue, Fig. 6(g)]. Due to the volume conservation property, this change spreads over the whole tongue in the midsagittal plane, generating at the same time a small forward movement of the apex and a small backward movement in the pharyngeal part. The inferior longitudinalis lowers the tongue tip (≈5 mm) and moves it backward (≈6 mm). A small backward movement of the tongue is also observed in the pharyngeal region (≈1.3 mm) [Fig. 5(h)]. In the transverse direction, a slight enlargement is observed in the dorsal region (≈1 mm). The activation of the superior longitudinalis mainly induces an elevation (≈12 mm) and a backward movement (≈11 mm) of the tongue tip with a slight backward movement in the pharyngeal part (≈1.7 mm) [Fig. 5(i)]. The geniohyoid essentially moves the hyoid bone forward and downward, which induces a slight lowering in the dorsal region (≈0.7 mm) [Fig. 5(j)]. Finally, the mylohyoid elevates the mouth floor in its midsagittal part (up to 4 mm) and moves the dorsal part of the tongue slightly upward [Fig. 5(k)]. The analysis of the influences of individual tongue muscles revealed possible synergies and antagonisms between muscles: GGp, Sty, and GH can act in synergy to produce an elevation of the tongue in the palatal region; in this part of the tongue they act antagonistically with the GGa and the GGm. The Sty and GGm are antagonists for the control of the vertical position of the dorsal part of the tongue. The GGp, SL, and GGm contribute to the tongue tip elevation and their action can be counteracted by that of the IL, the Sty, and the GGa. As for the control of the width of the tongue in the transverse direction, Trans tends to reduce it in the whole tongue body; GGa and GGm are the main muscles enlarging it in the palatal part, while HG contributes to its enlargement in the pharyngeal part.

2. Simulations of French vowels

In order to generate the 300 simulations used to determine the muscle activation patterns for the French vowels,

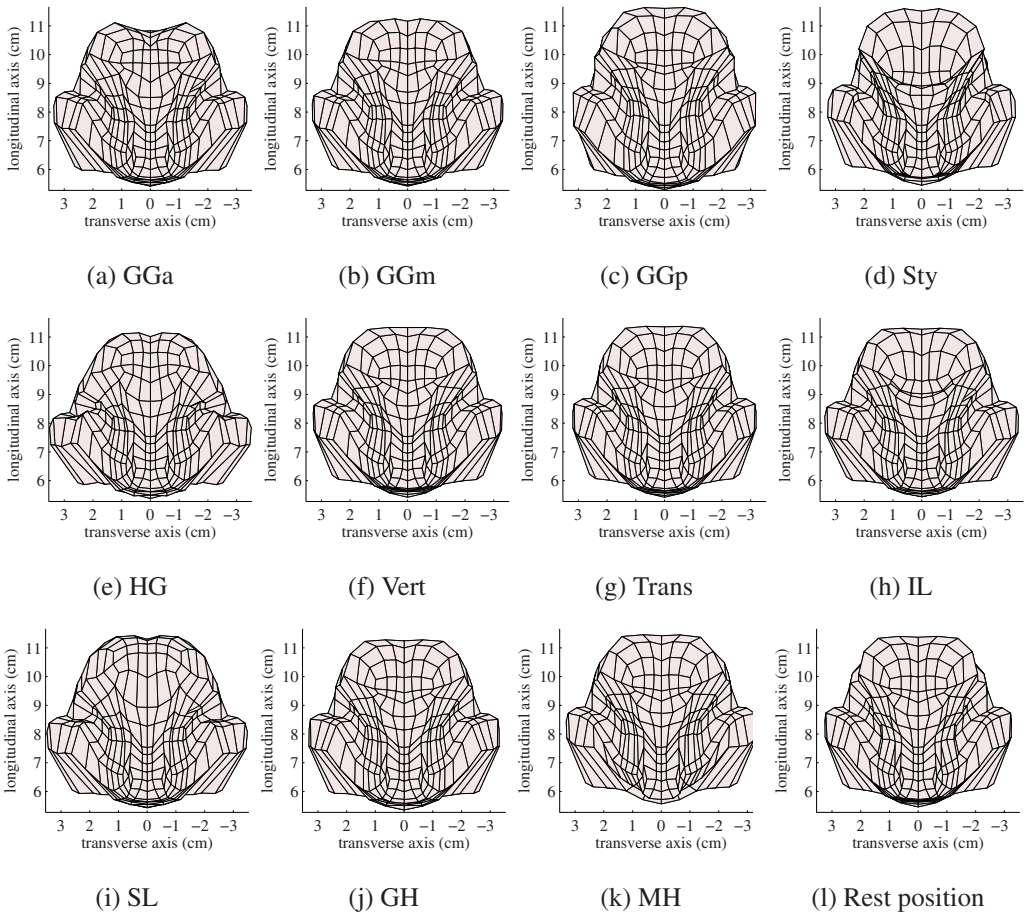


FIG. 6. (Color online) Impact of the activation of individual lingual and mouth floor muscles on tongue shape (400 ms command duration, sufficient to reach mechanical equilibrium) (frontal view). For every simulation, the target motor command of the only activated muscle equals 75% of the muscle length at rest, except for the activation of the long muscles Sty, IL, and SL (85% of the muscle length at rest). The shape of the tongue at rest is given on the bottom right.

the timing of the motor commands was as follows: at time $t=0$, the central commands were equal to the muscle length at rest; then they varied linearly for a transition time of 30 ms up to the target values. Coarse sets of motor commands were first determined for each vowel, guided by prior knowledge of the tongue shapes and by literature data. The values of the commands for the main muscles involved in the production of the vowels were then made to vary within a more or less wide range around their primary value. The range was determined according to the tongue shape sensitivity to their modification.

Within the set of 300 simulations, the best motor commands for French extreme vowels (Table II) were selected based on the obtained tongue shapes and the formant patterns. Optimal vowels were chosen in order to get the best match between the tongue shape in the midsagittal plane with PB's MRI data, and between the formants computed with the formants measured from PB's acoustic data. The tongue shapes and the formant patterns obtained for the 300 simulations were compared to the MRI data and the formant patterns collected from subject PB. For each French extreme vowel, the motor commands providing the best match of the tongue shape experimentally measured on PB in the midsagittal plane and of the corresponding formant patterns have been selected as reference motor commands (Table II). The corresponding tongue shapes are represented in Fig. 7 (oblique anterior and posterior views) and the formants are given in Table III (the lip aperture and protrusion are also indicated in this table). Table IV summarizes the force levels computed at the end of the selected simulations for every tongue and mouth floor muscle. The values indicated correspond to the algebraic sum of the force levels computed for

TABLE II. Motor commands used for the production of French cardinal vowels and for /ə/. These values are given as a percentage of the muscle length at rest. Values below 1 therefore correspond to a voluntary activation.

Vowel	GGa	GGm	GGp	Sty	HG	Vert	Trans	IL	SL	GH	MH
/i/	1.03	1.05	0.60	0.90	1.23	1.13	1.05	1.02	1.09	0.76	0.75
/ə/	0.98	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.94
/u/	1.20	1.20	0.91	0.84	1.25	1.35	0.95	0.98	1.20	0.95	0.80
/a/	0.75	1.10	1.00	1.10	0.70	0.85	1.30	1.00	1.20	1.05	1.05

TABLE III. Lip aperture l_a and protrusion l_p chosen for the determination of the vocal tract area function (based on Abry *et al.*, 1980) and the values of the first four formants for the simulation of French oral vowels (extreme cardinal vowels and /ə/). These values were computed with WINSNOORI software.

Vowel	l_a (cm^2)	l_p (cm)	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/i/	3	0.5	321	2095	2988	4028
/ə/	1.5	0.8	502	1235	2407	3612
/u/	0.3	1.5	298	723	2547	3450
/a/	4.5	0.8	667	1296	2875	3948

each macrofiber. It is not a true value of the force exerted on the tongue, but it provides a fair idea of its order of magnitude. Figure 7 reflects the traditional relationships between the French extreme vowels (anterior versus posterior, low versus high) while the formants are consistent with the classic published values and with acoustic data obtained for the speaker PB (Table V). We note a good correspondence between the formants of the acoustic data measured for PB and those obtained with the simulations. The average difference between the formants that were measured and those that were simulated is below 3.3% for the first four formants. The difference does not exceed 4.3% for the first formant (vowel /u/) and 10.2% for the second formant (vowel /i/).

Due to the redundancy of the system (some pairs of muscles interact as agonist-antagonists), the commands were also chosen such that the amount of force generated by the different muscles remains reasonable. Only the extreme cardinal vowels will be presented in detail.

a. Muscle activation pattern in vowel /i/. Figure 8 shows the tongue shapes obtained by simulation and those obtained experimentally for the speaker PB (CT data). Some discrepancies can be seen in the tongue posterior part, but the delimitation of the tongue contours in this area is less precise (the delimitation of the tongue body on CT images is a tedious and less obvious task in this part of the body, due to the presence of the hyoid bone, epiglottis, and other soft tissues) and acoustically less relevant than in the anterior part. This figure shows a good correspondence between the experimental results and the computed data, in particular, in the anterior part of the tongue, which plays an important role in the production of vowel /i/. As expected from Fig. 5, since vowel /i/ is an anterior and high vowel [Fig. 7(a)], the model predicted the GGp, GH, and MH muscles to play a fundamental role in its production. In addition to their slight impact on the tongue geometry (see Sec. III A 1), the GH and MH muscles can help stiffen the mouth floor, thanks to a significant propagation of the stress into the lingual tissues. Activated alone, the styloglossus pulls the tongue backward [Fig. 5(d)]; this movement is here counterbalanced by the strong GGp activation, while both muscles elevate the tongue in the palatal region. For the transversalis and the anterior genioglossus, the motor commands (λ commands of the EPH) are

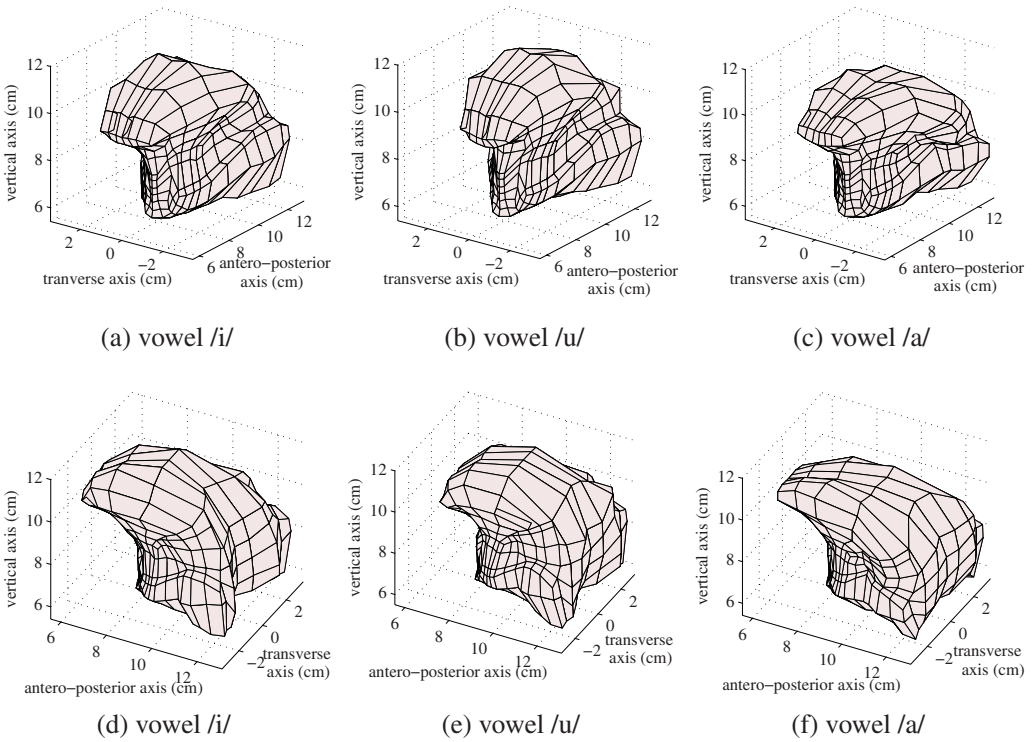


FIG. 7. (Color online) Final tongue shape for the simulation of the French cardinal vowels (first row: anterior oblique view; second row: posterior oblique view).

TABLE IV. Final force levels (in newtons) observed for every tongue and mouth floor muscle during the production of the French cardinal vowels and /ə/. The levels of force indicated correspond to the algebraic sum of the forces computed for every macrofiber. Bold cells represent voluntarily activated muscles.

Vowel	GGa	GGm	GGp	Sty	HG	Vert	Trans	IL	SL	GH	MH
/i/	0.51	0	**25.82**	**6.90**	0	0	1.61	0	0	**3.37**	**13.89**
/ə/	**0.11**	**0.13**	**0.95**	0.04	0.04	0.00	0.25	0	0.10	**0.10**	**1.62**
/u/	0	0	**6.73**	**7.42**	0	0	**1.83**	**0.47**	0	**1.02**	**6.79**
/a/	**3.34**	0	1.91	0	**8.21**	**2.31**	0	0.16	0	0	0.78

larger than the muscle lengths at rest (Table II). From a motor control perspective, these two muscles can consequently be seen as being in their rest state, and their activation is the result of reflex loops (Table IV). The transversalis reflex activation avoids an overwidening of the tongue that would otherwise result from the combination of the GGp activation (see Sec. III A 1), while ensuring a contact between the palatal arch and the lateral borders of the tongue in the alveolar region. The GGa reflex activation limits tongue elevation in the median alveolar region, thus creating the slight groove characteristic of an /i/. The voluntary activations are consistent with the EMG data of Baer *et al.* (1988), except for the Sty, for which no activity was measured by these authors for vowel /i/. With the model, the combined activation of the GGp and Sty is essential to precisely control the location of the constriction for high vowels. This co-activation is consistent with our previous findings with a 2D tongue model (Payan and Perrier, 1997).

Qualitatively the tongue shape proposed for /i/ is in good agreement with different kinds of data published in the literature. This is true for the 2D shape in the midsagittal plane, which is consistent with Bothorel *et al.* (1986) data for French speech sounds. It is also true for the 3D distribution of the contacts between the hard palate and the upper dental arch on the one side and the tongue lateral borders on the other side. These contacts are represented in Fig. 10(a). The surface of contact stretches over the whole hard palate and is also extended to the inner aspects of the molars. In addition, we note the presence of contacts between the apex and the mandible inner surface, behind the lower incisors (not shown). These observations are consistent with the EPG data of Stone and Lundberg (1996) (Fig. 9) and Yuen *et al.* (2007) for English vowels.

b. Muscle activation pattern in vowel /u/. The model produces vowel /u/, a posterior and high vowel [Fig. 7(b)], essentially with the activation of the styloglossus, the mylohyoid, and the transversalis (Table II). As with vowel /i/, the model requires the activation of the MH to stiffen the mouth floor and thus contribute to the tongue elevation, due to the complementary action of other muscles. The styloglossus allows the tongue to be pulled both backward and upward. The GGp is also active. It increases the size of the vocal tract back cavity by propelling the tongue forward and contributes to the upward movement of the tongue. The transversalis contributes to the limitation of the tongue widening, but this is not its only role. Indeed, for this vowel the model uses an active recruitment of the transversalis in order to facilitate the tongue elevation, due to the incompressibility of the lingual tissues (note, however, that the amount of force generated by the transversalis is close to that used in the production of /i/). The motor commands proposed in our model are consistent with the EMG data of Baer *et al.* (1988). Here again, the 2D tongue shape in the midsagittal plane is in good agreement with data of Bothorel *et al.* (1986). In our simulation, the tongue tip is located in the midheight of the tongue. Figure 10(b) shows the distribution of the contacts of the tongue dorsum and the tongue tip with the surrounding structure, namely, the hard and soft palates, the superior dental arch, and a part of the pharyngeal walls. The figure shows that the tongue post-dorsal surface is laterally in contact with the inner surface of the molars and, further back, with the lateral sides of the pharyngeal walls. The contacts between the tongue and hard palate observed in the simulations are consistent with the EPG data of Stone and Lundberg (1996) (Fig. 9) and Yuen *et al.* (2007). However, EPG data do not provide information on possible contacts between the tongue and velum.

c. Muscle activation pattern in vowel /a/. Vowel /a/, a posterior and low vowel, was essentially produced in the model with the activations of the HG and GGa muscles [Fig. 7(c) and Table II]. The HG pulls the tongue backward and downward but also rotates the tongue tip toward the palate

TABLE V. Values of the first four formants based on acoustic data obtained for the speaker PB. The values were averaged over ten repetitions of every one of the extreme cardinal vowels in different contexts.

Vowel	F1 (Hz)	F2 (Hz)	F3 (Hz)	F4 (Hz)
/i/	311	2308	3369	4126
/u/	285	792	2783	4055
/a/	661	1291	2657	3717

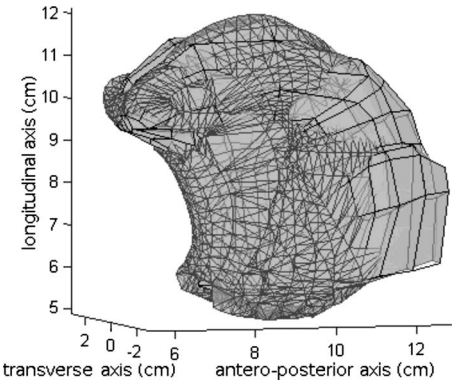


FIG. 8. Superimposition of the shape of the tongue for the speaker PB (CT data) (dense mesh) and the shape of the tongue obtained by simulation (coarse mesh) for vowel /i/.

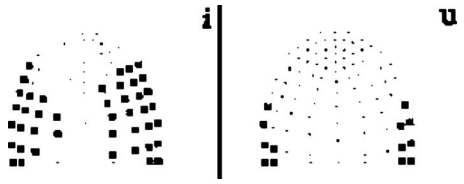


FIG. 9. EPG data for American English vowels /i/ and /u/. Reprinted from Stone, M., and Lundberg, A., J. Acoust. Soc. Am., "Three-dimensional tongue surface shapes of English consonants and vowels," **99**(6), 3728–3737, 1996. Copyright© 1996, Acoustical Society of America.

[Fig. 5(e)]. The GGa limits the apex rotation by flattening the tongue tip and maintaining it in contact with the inner surface of the mandible, thus preventing the creation of a sublingual cavity and increasing the size of the anterior cavity. The GGp activation is a reflex activation, since from the motor command point of view it is in its rest state (see Table II); the GGp limits the backward movement of the tongue and thus avoids the occlusion of the vocal tract in the laryngopharyngeal region. The motor commands are in agreement with the EMG data of Baer *et al.* (1988). The tongue shape in the midsagittal plane is in agreement with data of Bothorel *et al.* (1986). The lateral borders of the tongue are in contact with the lower dental arch over its entire length, but not with the palate. The lower surface of the tongue anterior part is partially in contact with the inner surface of the mandible. For vowel /a/, the EPG data of Stone and Lundberg (1996) and Yuen *et al.* (2007) reported an either extremely limited or non-existent contact between tongue and palate; the results obtained are consistent with their data.

B. Highlighting the role of the transverse muscle in midsagittal tongue shaping

A 3D biomechanical tongue model allows the study of the transverse muscle action during speech production. Since speech has experimentally mainly been studied in the sagittal domain, the potential role of this muscle has essentially been ignored. However, it could be of great importance in speech production, since it is the only muscle able to directly act on tongue deformations in the transverse dimension orthogonal to the sagittal plane.

The role of the transverse muscle in the midsagittal deformation of the tongue was recently observed by Gilbert *et al.* (2007) for swallowing through the analysis of diffusion-weighted MRI measurements. They found, in particular, that the recruitment of the transversalis is used to generate depressions in the tongue to facilitate the movement of the food toward the pharynx. Unfortunately, similar experimental observations do not yet exist for speech, and it is a strength of our 3D model that it offers the possibility to quantitatively assess the role of the transversalis in speech production. As a matter of fact, the simulations of vowel production reported in Sec. III A 2 highlighted the fundamental role of this muscle in the maintenance of the tongue dimension along the transverse direction and its influence on midsagittal shaping. These results have been obtained in the context of our motor control model, based on the EPH theory, which gives an account of the postural control in a particularly effective way, thanks to the integration of reflex activation in the muscle force generation mechanisms. Indeed, the model predicts that for vowel /i/ (and also for the high anterior vowels /y/ and /e/ not presented here), the transverse muscle is active, despite the fact that the motor commands for this muscle were those of the rest position or higher [see Eq. (2)]. This is the result of a reflex activation (or limited active contraction) due to the lengthening of the transverse fibers induced by the centrally activated muscles that mainly act on the tongue shape in the sagittal plane. This reflex activation limits the amplitude of the deformations in the transverse dimension and, in turn, due to the incompressibility of tongue tissues, it increases the deformations in the sagittal plane. According to the simulations, a voluntary activation of the transversalis would lead to a decrease in the tongue width that does not seem compatible with the production of high anterior vowels, unless this decrease can be compensated by the action of other muscles. Hence, the combination of a voluntary co-activation of the transversalis and of other tongue muscles could also be considered as an alternative to the proposed reflex activation of the transversalis. Such a strategy is realistic, but it would imply the activation of a larger number of muscles acting antagonistically, inducing an increase the amount of force necessary to produce high anterior vowels. Our simulations do not rule out

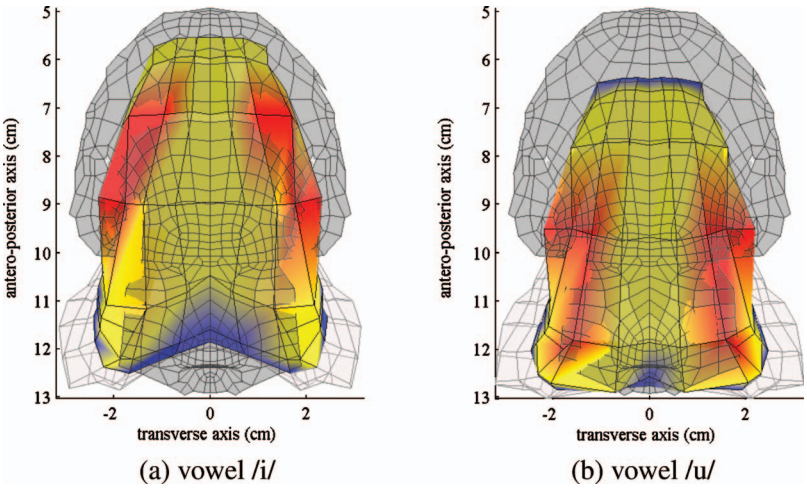


FIG. 10. Illustration of the contacts between the tongue dorsum and tongue tip, and surrounding structures of the vocal tract for vowels /i/ (a) and /u/ (b) (superior view, apex at the top). The surrounding structures, represented by translucent gray meshes, include the hard palate, the upper dental arch, the velum, and the pharyngeal walls close to the velum. The entire tongue mesh is represented, but only the tongue surface elements used for the detection of potential contacts between the tongue and the surfaces listed above are colored. Red elements represent tongue surfaces in contact with the surrounding structures, yellow elements represent tongue surfaces close to the surrounding structures, and blue elements represent tongue surfaces far from the surrounding structures.

the possibility of a voluntary activation of the transversalis. However, such a strategy does not sound like an economical way to control tongue shapes for high vowels. As already mentioned above, 2D or 2.5D models, such as those of Payan and Perrier (1997) or Dang and Honda (2004), could only account for tongue incompressibility in the sagittal plane due to a simplifying assumption assimilating volume conservation and area preservation in this very plane. In a way, this simplifying approach implicitly included the role of the transverse muscle, without formalizing it in explicit terms. We have seen in Sec. III A 1 that this hypothesis led to partially inaccurate conclusions concerning the role of muscles taken individually. Our 3D modeling approach allows these former conclusions to be corrected and emphasizes the indirect role of the transverse muscle in the shaping of the tongue midsagittally [Fig. 7(a)].

Based on simulations made with their 2D model, Perrier *et al.* (2000) concluded that the main directions of deformation for the tongue during speech production as observed for different languages [namely, the factors front and back raising of the parallel factor analysis (PARAFAC) of Harshman *et al.* (1977), see Jackson, 1988; Maeda, 1990; Nix *et al.*, 1996; Hoole, 1998, or more recently Mokhtari *et al.*, 2007] did not result from a specific speech control, but emerged naturally from the actions of the major tongue muscles (GGp, GGa, HG, and Sty). Similar conclusions could be drawn from Honda's (1996) EMG data. The results concerning the role of individual muscles in our 3D model can be used to reformulate these conclusions more accurately. The main directions of deformation could indeed emerge naturally, provided that the tongue widening along the transverse direction is strictly controlled by the reflex transversalis activation. This reflex activation, based on the use of the motor commands at rest, is not likely to be speech specific, since it allows the tongue to remain within the space determined by the dental arches, possibly in order to avoid biting problems [several observations indeed show a widening of the tongue for edentulous people (Kapur and Soman, 1964)]. Taking into account this reflex limitation of tongue width seems to be essential to understanding the precise control process of the place of articulation in the vocal tract.

IV. VARIABILITY OF MOTOR COMMANDS AND TONGUE POSITIONING ACCURACY FOR VOWEL /i/

A. Methodology

The accuracy of speech motor control is an important and still unsolved issue. Indeed, speech movements can be as short as a few tens of milliseconds, so it is traditionally suggested that cortical feedback, involving long latency loops, can only be used to monitor speech after its production and not during on-going production (see, for example, Perkell *et al.*, 2000 for details). Tongue positioning has to be very accurate though for the production of some sounds, such as fricatives and high vowels. This apparent contradiction (the absence of cortical feedback versus the accuracy requirement) suggests that speech motor control has developed into a very efficient process to ensure, in a simple way, accuracy and stability of tongue positioning. This efficient treatment and accuracy can be seen as the result of the high amount of training and experience in speaking that speakers have.

For the high vowel /i/ more specifically, it has been argued that control accuracy would come from a combination of biomechanical effects, namely, the co-contraction of the GGp and the GGa associated with tongue/palate contacts (Fujimura and Kakita, 1979). This effect is called the "saturation effect." Using a rudimentary 3D tongue model, Fujimura and Kakita (1979) showed that the tongue was stabilized during the production of /i/ when laterally pressed against the palate, due to the combined action of the GGa and GGp, which stiffened the tongue. Our 3D model, which integrates numerous improvements as compared to Fujimura and Kakita's (1979) original model (smaller mesh elements, non-linear tissue elasticity, gravity, stiffening due to activation, and accurate model of contacts), offers a powerful context to revisit this hypothesis and to better understand how the different biomechanical factors interact. With the current model, a number of simulations were realized around the reference tongue shape for /i/ to evaluate the articulatory and acoustic sensitivity of the vowel to changes provided to the motor commands. The tongue shape variations as well as the formant variations resulting from small changes in the central commands were studied for this vowel, so as to better understand the patterns of variability observed during its production. The motor commands defined previously (see Sec. III and Table II) formed the basis of this study. The motor commands of the main tongue muscles (i.e., the anterior, medial and posterior genioglossus, the styloglossus, the hyoglossus, the transversalis, the lingual inferior and superior muscles, and the mylohyoid) were independently modified. For the GGa, GGp, Sty, MH and Trans, the motor commands were modified by $\pm 2\%$, $\pm 5\%$, $\pm 8\%$, and $\pm 10\%$ around their values at target. For the GGm, HG, IL, and SL, which were not active during the production of vowel /i/ in our modeling, the motor commands were only modified by -2%, -5%, -8%, and -10%, since an increase in their values would leave them inactive. The same lip protrusion and aperture parameters as previously applied were used to generate the acoustic signals and to determine the formants associated with the different area functions. Table VI indicates the first three formants for each of the 56 simulations.

B. Results

Figure 11 shows the scatter plots in the midsagittal plane for six nodes on the tongue surface obtained from the simulations. Results are presented in the upper left panel for the variations in all muscle commands together, and in the other panels, more specifically, for the variations in the commands to three muscles that play a major role in the production of vowel /i/: the styloglossus and the anterior and posterior genioglossus. For the global results (upper left panel), 3σ ellipses characterizing the node position dispersion with a Gaussian statistical model are superimposed on the data. Considering first the influence of all muscles taken together, the following observations can be made. In the pharyngeal and velopharyngeal regions (three most posterior nodes), the major axes of the dispersion ellipses essentially correspond

TABLE VI. First, second and third formants computed after a local modification of the motor commands for vowel /i/. The formants for vowel /i/ are given in Table III. The motor commands of the vowel /i/ were modified for 9 muscles independently by ±2%, ±5%, ±8% or ±10%. The formants obtained following these modifications are given. In the λ-model, an increase in the motor commands corresponds to a decrease in the muscle activation; therefore a modification by +10% corresponds to the lowest level of activation for a given muscle, whereas a modification by −10% corresponds to the highest level of activation.

		F1 (Hz)	F2 (Hz)	F3 (Hz)
GGp	+10%	350	2084	2942
	+8%	340	2073	2931
	+5%	322	2080	2950
	+2%	326	2090	2976
	−2%	317	2105	2992
	−5%	309	2102	3014
	−8%	305	2102	3024
	−10%	302	2111	3026
GGa	+10%	273	2078	3035
	+8%	273	2078	3035
	+5%	283	2082	3012
	+2%	307	2099	3008
	−2%	333	2082	2967
	−5%	351	2071	2928
	−8%	368	2062	2903
	−10%	375	2049	2895
MH	+10%	329	2101	2977
	+8%	325	2096	2976
	+5%	322	2095	2986
	+2%	321	2095	2982
	−2%	319	2091	2983
	−5%	318	2100	2986
	−8%	317	2096	2985
	−10%	316	2096	2980
IL	−2%	321	2097	2980
	−5%	322	2094	2969
	−8%	322	2098	2964
	−10%	321	2103	2958
GGM	−2%	321	2085	2979
	−5%	322	2057	2960
	−8%	321	2019	2938
	−10%	322	2001	2924
HG	−2%	320	2094	2986
	−5%	323	2087	2973
	−8%	326	2079	2963
	−10%	331	2076	2957
Sty	+10%	272	2009	3017
	+8%	278	2018	3000
	+5%	296	2050	2994
	+2%	311	2076	2981
	−2%	334	2135	3003
	−5%	348	2167	2924
	−8%	360	2214	2814
	−10%	364	2251	2766
Trans	+10%	316	2118	3049
	+8%	317	2117	3043
	+5%	318	2112	3027
	+2%	320	2108	3004
	−2%	321	2083	2966
	−5%	321	2069	2944
	−8%	321	2056	2920
	−10%	319	2048	2913
SL	−2%	322	2094	2989
	−5%	327	2104	2977
	−8%	342	2103	2958
	−10%	346	2106	2912

to a displacement along the front-to-back direction (from the pharyngeal to the velopharyngeal position, lengths of the major axes 3.0, 3.9, and 4.4 mm, respectively, lengths of the small axes 1.5, 1.4, and 0.8 mm, angles of the major axes with the antero-posterior axis 133°, 147°, 173°). In the palatal and alveopalatal parts of the tongue (second and third nodes from the front), the ellipses have no clear direction and they tend to be more circular. In addition, the maximal variability is smaller than in the back part of the tongue (lengths of the major axes 2.9 and 3.5 mm, respectively, lengths of the small axes 2.4 and 2.1 mm). Finally, in the apical part (most anterior node), a very strong correlation is observed between elevation and forward movement. This leads to a global ellipse orientation similar to the one observed in the tongue blade region, but much stronger and clearer and with much more variation along the principal axis (length of the major axis 8.3 mm, length of the small axis 3.4 mm, and angle of the major axis 147°).

These observations are in quite good agreement with experimental data published in the literature about vowel variability. See, in particular, Perkell and Nelson, 1985; Beckman *et al.*, 1995; or Mooshammer *et al.*, 2004: the front-back orientation of the variability in the velar region and the reduced variability in the palatal and alveopalatal regions (the region of constriction for /i/) were already observed by these authors. In addition, the absence of clear orientation of the ellipses in the region of constriction was also observed in two of the three subjects studied by Mooshammer *et al.* (2004) while Perkell and Nelson (1985) and Beckman *et al.* (1995) rather observed ellipses parallel to the palatal contour in this region. The large variability in the apical part was observed by Mooshammer *et al.* (2004), but not by Perkell and Nelson (1985) and Beckman *et al.* (1995). Note, however, that this specific aspect of the displacement of the apex relative to that of the tongue body has already been observed many times by different authors, in particular, Perkell (1969).

Our model allows one to look more specifically at the biomechanical factors influencing these articulatory patterns. Looking at the variability associated with the variation in the GGa, GGp, and Sty activations separately, it can be observed that the angle of the main ellipses in the three posterior nodes is similar to the orientation of the scatter plots generated by the Sty and GGa. However, for the GGp, the largest variability is also observed in the front-back direction. The reduction in the variability in the region of constriction is observed both for the Sty and the GGp, while the GGa, in contrast,

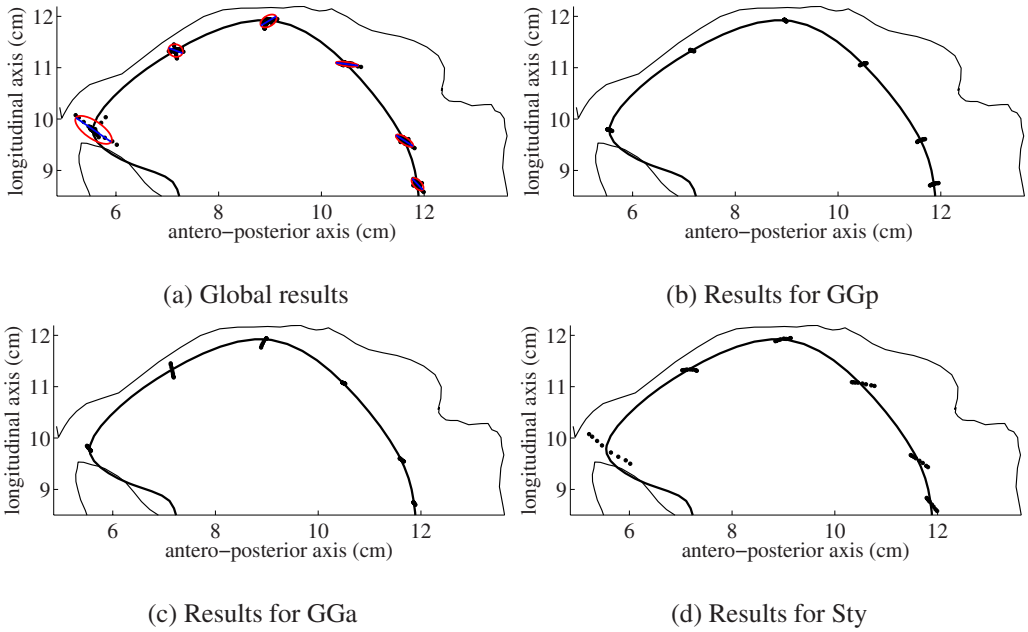


FIG. 11. (Color online) Displacement scatter plots (circles) for vowel /i/ in the midsagittal plane. Only the surface of the tongue is represented. Panel (a) summarizes the results obtained for the nine muscles whose motor commands were modified. The 3σ ellipses of dispersion are also represented, and their major axes are drawn. Panels (b)–(d) represent the dispersion obtained when modifying the motor commands of the posterior genioglossus (b), the anterior genioglossus (c), and the styloglossus (d) only.

shows the largest variability in this region. This can be interpreted in the light of the palatal contacts for vowel /i/ (see Fig. 10). The GGp and Sty act on the position of the whole tongue body, whose variability in the constriction region is limited by the palatal contacts. The GGa influences only the center of the front part of the tongue, which is not in contact with the palate. It can be noted that in our simulations the styloglossus generates the largest variability, as compared to the other muscles, except as explained above in the constriction region. This phenomenon is intrinsically linked to the approach that was used in our simulations. In the context of the λ-model, since the styloglossus macrofibers are longer than in the other muscles, a given percentage of variation generated a larger change in the commands for the styloglossus and then, in turn, larger changes in force level. This approach could have influenced the global amount of variability depicted in the upper left panel of Fig. 11, but not its relation with the node position on the tongue, neither in terms of orientation nor of amplitude, except for the tongue tip variation, which is largely dependent on the styloglossus.

As concerns the other muscles (not depicted in Fig. 11), their impacts are smaller, but some interesting observations can be mentioned. The transversalis shows a notable contribution in the variability in the velopharyngeal region (third node from the back), where an increase in its activation induces backward displacements of the tongue dorsum. The mylohyoid participates in the up-down displacements, with an amplitude of approximately 1.5 mm in the pharyngeal region.

C. The saturation effect for vowel /i/ revisited

From the simulations of the consequences of motor command variability for tongue shape for vowel /i/, interesting conclusions can be drawn for this vowel concerning the influence of the GGa and of its variability on the vocal tract shape and formants. Contrary to what could be inferred from the statistical processing of articulatory speech data (see, for example, Badin *et al.*, 2002), the central tongue groove observed for vowel /i/ in many languages does not seem to be a consequence of the combined activations of the GGp and Sty muscles. It is, in fact, obtained in our model very specifically by activating the GGa. As mentioned above, this statement is consistent with Fujimura and Kakita's (1979) hypothesis of a co-activation of the GGp and GGa in the production of /i/.

However, the variability patterns generated with our model, together with their interpretation in terms of the respective influence of each muscle, strongly suggest that there is no saturation effect, which would facilitate the accurate control of the constriction area for /i/. This observation questions Fujimura and Kakita's (1979) original hypothesis as well as the numerous follow-up contributions that have used this hypothesis to explain the control of high front vowels, in particular, those of Perkell *et al.* (2000) and Badin *et al.* (1990).

In agreement with the work of the previous authors, our model tends to confirm that the tongue is indeed stabilized in its entirety by these palatal contacts, and that this should contribute to simplifying its motor control. However, in contrast to Fujimura and Kakita's (1979) tongue model, which was quite rudimentary because of the computational limitations existing at that time, our model shows that the variability of the GGa activation leads to a variation in the alveolar groove with noticeable consequences for its formant pattern (see below). This variation is highly localized in the globally well-stabilized tongue, but it is fundamental to acoustics, because it plays on the constriction size.

The amplitudes of variation for the first three formants

were as follows: $\Delta F1 \approx 103$ Hz, $\Delta F2 \approx 250$ Hz, $\Delta F3 \approx 283$ Hz (Table VI).[2] An important part of the variability is due to the styloglossus (impact on F1, F2, and F3, but see our remark above about the force level variation for this muscle), but other muscles also have a noticeable influence, either on the first, second, or third formant. The F1 variability is due in great part to the modifications in the level of activation of the GGa and Sty, and secondarily of the GGp and SL. According to the model, the variability of F2 results mainly from the modification of the Sty and GGm motor commands, while that of F3 is due to the Sty and Trans.

The variability of F1 for /i/ has important consequences; indeed the perception of vowel /i/ is sensitive to F1 variations in French (one can easily move from /i/ to /e/). Likewise for F3, too low an F3 value moves the perception from /i/ to /y/ (Schwartz and Escudier, 1987). It can therefore be concluded that the articulatory variability generated in the simulations is too important to ensure proper perception of vowel /i/. It is necessary to reduce it. This need for an active reduction in the articulatory variability is consistent with the observations made by Mooshammer *et al.* (2004) with German speakers: they concluded from their study that the potential saturation effect related to the interaction between tongue and palate did not seem to be sufficient for their speakers to meet the perceptive requirements of the German vowel system, and that a specific control adapted to the individual palate shape of each speaker was necessary to limit the articulatory variability and its consequences for perception.

This is then quite an important result as it throws back into question a widely made assumption to explain the precise control of the vowel /i/, namely, the saturation effect.

V. IMPACT OF GRAVITY ON LINGUAL MOVEMENTS

With the increasing use of MRI systems, numerous speech data are acquired while the subject is lying on his or her back. Due to the change in the orientation of gravitational forces in relation to the head, this position is likely to alter the vocal tract shape and its control. This is why many studies have tried to compare the production of speech sounds and speech articulations (either vowels or consonants) for subjects when they are sitting, standing, or lying (Weir *et al.*, 1993; Tiede *et al.*, 2000; Shiller *et al.*, 2001; Stone *et al.*, 2007). Our model allows the impact of gravity to be tested and quantitatively assessed. With this aim in view, the pattern of activation needed to keep the tongue in its neutral position was first studied in the presence of a gravitational field in an upright and in a supine position. Then, the influence of gravity on the tongue shape during the production of vowels was evaluated together with its impact on the acoustic signal.

A. Impact of gravity in the absence of active and reflex muscle activation

First, the impact of gravity alone on the tongue shape and position was studied: the force generator was deactivated (no internal force could be generated, whether active or reflex forces). The final tongue shape is given in Fig. 12 for upright and supine positions starting from the rest position and after a 1 s movement. For a standing subject, there is a clear lowering of the tongue body, which is particularly marked in the posterior radical part of the tongue but is also visible in its apical region (approximately 1.5 mm). For a lying down subject, the gravity alone produces a strong backward displacement of the tongue body, with a displacement of the tongue tip equal to 9 mm. These results show that tongue muscle activations are required to maintain the tongue in its rest position, whether the subject is lying on his back or standing.

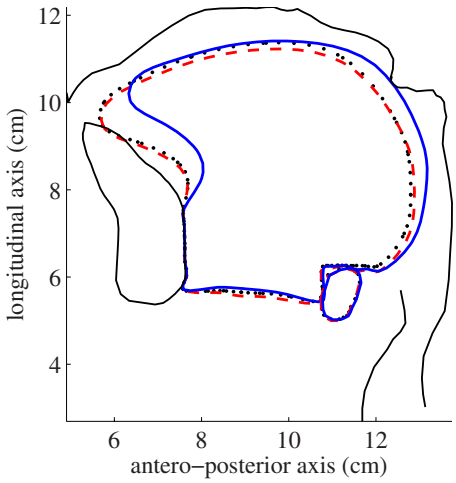


FIG. 12. (Color online) Final tongue position in the midsagittal plane for 1 s simulation under the influence of gravity alone. The neutral position of the tongue and the hyoid bone (rest position for a subject in upright position) is represented by a dotted line, the final shape for the tongue and hyoid bone for a subject in upright position by a dashed line, and for a subject in supine position by a solid line. Other solid lines correspond to the contours of the mandible, hard and soft palates, and pharyngeal and laryngeal walls.

Reflex activation obtained with motor commands equal to muscle lengths at rest is not sufficient to maintain the tongue in the rest position, as shown by Fig. 13 for a 1 s simulation. A small backward displacement of the apex (≈ 1 mm) and of the rear part of the tongue is visible in the upright position, as well as a more limited rotation of the apex in the supine position than in the absence of muscle activations (displacement of the tongue tip ≈ 3 mm). A limited voluntary activation of the GGp and GGa combined with a stronger activation of the MH associated with the reflex activation of the other tongue and mouth floor muscles can compensate for the gravity effect (commands indicated in Table II for vowel /ə/). Based on the model, the MH activation strengthens the mouth floor and limits the lowering of the tongue inferior region [Fig. 5(k)]. The GGp action prevents the backward displacement of the tongue [Fig. 5(c)] and the GGa counteracts the GGp action in the apical and dorsal areas, limiting the tongue elevation [Fig. 5(a)]. A good equilibrium between the activation of these three muscles, based on numerous simulations, leads to the stabilization of the tongue in a "neutral" upright position. Corresponding force levels computed at the end of the simulation for every tongue and mouth floor muscle are given in Table IV (vowel /ə/).

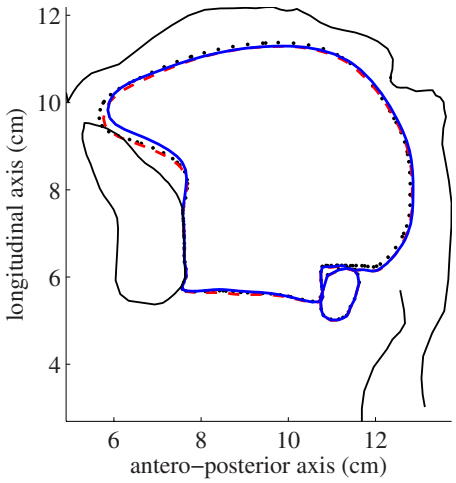


FIG. 13. (Color online) Final tongue position in the midsagittal plane for 1 s simulation under the influence of the reflex activation alone. The neutral position of the tongue and the hyoid bone (rest position for a subject in upright position) is represented by a dotted line, the final shape for the tongue and hyoid bone for a subject in upright position by a dashed line, and for a subject in supine position by a solid line. Other solid lines correspond to the contours of the mandible, hard and soft palates, and pharyngeal and laryngeal walls.

B. Impact of gravity on French oral vowels

The impact of the subject position (upright or supine) on vowel production was studied by modifying the orientation of the gravitational field. Simulations were realized for the supine position for the ten French oral vowels with the same commands and the same timing as in the upright position. The tongue shapes and positions for the supine and upright positions were compared, as well as the force levels for tongue and mouth floor muscles.

The differences in tongue shape and formant values between upright and supine positions were negligible for all vowels. However, differences were noticed in the level of forces developed by the GGp, with an increase in supine position that is variable across vowels: the peak and the final forces increased on the orders of 8% for /a/, 7% for /u/, and 1% for /i/. On the whole, this modification in the force level affects in percentage more posterior than anterior vowels. This is consistent with the fact that the production of front vowels necessitates a strong force from the GGp anyway, in comparison to which the gravitational force becomes quasinegligible. Our results are also in agreement with experimental observations, in which an increase in the GGp activity in supine position is commonly observed [see, for example, the EMG data of Niimi *et al.* (1994) and Otsuka *et al.* (2000)].

In the model, the tongue weight, which is on the order of 1 N, is small as compared to the muscular forces. Hence, feedback activation efficiently counteracts the effects of gravity orientation changes and limits tongue shape variation. This small shape variation is in contradiction to experimental values typically found in the literature. For instance, Badin *et al.* (2002) reported a more important backward displacement of the tongue for both vowels and consonants in supine position (MR images) compared to upright position (cineradiofilm images), which they attributed to the tongue weight. Shiller *et al.* (1999) found differences in the formant values between the upright and supine positions for vowels /a/ and /ɛ/. They found that when the head was in the supine orientation, the jaw was rotated away from occlusion, which led them to conclude that the nervous system did not completely compensate for changes in head orientation relative to gravity. In the current model, the fact that the model has a fixed jaw position (the same for the supine and upright orientation) could in part explain the absence of notable differences. However, it should be mentioned that recent experimental findings provide good support for our simulation results. Indeed, Stone *et al.*(2007) showed that the impact of gravity was low (or even negligible) for some speakers when vowels were pronounced in context and not in an isolated manner as has thus far been the case.

VI. CONCLUSION

A 3D finite element model of the tongue has been presented which was used to study biomechanical aspects and tongue control during vowel production. The model provides a high level of realism both in terms of compliance with anatomical and morphological characteristics of the tongue and in terms of soft tissue modeling hypotheses (geometrical and mechanical non-linearity). The tongue and mouth floor muscles were controlled using a force generator based on the EPH theory. Simulations with the model coupled with an acoustic analog of the vocal tract allowed muscle activation patterns to be proposed for the French oral vowels which were consistent with the EMG data published in the literature and which generated realistic tongue shapes, tongue/palate contact patterns, and formant values. The simultaneous analysis of these activation patterns and of the actual muscle forces generated for each vowel revealed, among other things, a systematic feedback activation of the transversalis. This suggests that this muscle is used to maintain the dimension of the tongue quasi-constant along the transverse direction orthogonal to the sagittal plane. This role is very important for the control of tongue shape in the midsagittal plane, since, due to tongue tissue incompressibility, it allows more deformation in this plane. This is consistent with the recent experimental observations made by Gilbert *et al.* (2007) for swallowing. The results obtained from the simulations have led us to conclude that the main directions of tongue deformation in the midsagittal plane [as described by the classic front and back raising factors of Harshman *et al.* (1977)] could naturally emerge from the combined action of the major tongue muscles and of the transversalis playing the role of a "size maintainer" in the transverse direction. This conclusion is in line with Perrier *et al.* (2000), who suggested that these main directions of deformation are not speech specific, but are intrinsically linked to tongue muscle arrangements.

The muscle activation patterns proposed for each French vowel served as a basis for further studies. The patterns of articulatory variability, and their associated acoustic variability, were analyzed for local changes in the central commands for vowel /i/. These results cast doubt over the idea, generally accepted since the work of Fujimura and Kakita (1979),

that a muscular saturation due to a simultaneous co-activation of the GGa and GGp muscles would facilitate the accurate control of /i/. Indeed, the tongue grooving in the constriction region was shown to be sensitive to change in the GGa activation with a significant impact on the F1 formant value. The impact of gravity was also considered. Simulations showed the importance of low-level feedback in the postural control for the rest position, as well as the impact of the head orientation on the tongue shape and position. These results are at odds with data published in the literature for isolated sound production, but they find support in the recent work of Stone *et al.* (2007) on the production of vowels in context.

Further work will be required to significantly reduce the computation time, and thus increase the number of simulations and refine the results. Studies have also been undertaken to assess the contribution of this model to medical applications, in particular, the surgical planning of tongue exeresis, with lingual tissue resection and reconstruction processes. First results have proved to be promising and show the potential of such a model (Buchaillard *et al.*, 2007). The results obtained for the planning of tongue surgeries and the comparison with patients' data should also provide particularly interesting information about the compensation processes and the motor control mechanisms.

ACKNOWLEDGMENTS

The authors wish to thank Ian Stavness for helpful comments and suggestions. This project was supported in part by the EMERGENCE Program of the Région Rhône-Alpes and by the P2R Program funded by the CNRS and the French Foreign Office (POPAART Project).

[1]Program written by Pierre Badin (ICP/GIPSA-Lab).

[2]It should be noted that these values were obtained with tongue motions starting from a resting state. Modifying this starting state would have an impact on the formant values, but due to the simulation durations, sufficient to reach an equilibrium position, and to the model of motor control, the variations of the formant values should remain limited.

Abry, C., Boë, L.-J., Corsi, P., Descout, R., Gentil, M., and Graillot, P. (**1980**). *Labialité et phonétique: Données fondamentales et études expérimentales sur la géométrie et la motricité labiales (Labiality and phonetics: Fundamental data and experimental studies on lip geometry and mobility)* (Publications de l'Université des langues et lettres de Grenoble, France).

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., and Savariaux, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images," J. Phonetics **30**, 533–553.

Badin, P., Perrier, P., Boë, L.-J., and Abry, C. (**1990**). "Vocalic nomograms: Acoustic and articulatory considerations upon formant convergence," J. Acoust. Soc. Am. **87**, 1290–1300.

Baer, T., Alfonso, P. J., and Honda, K. (**1988**). "Electromyography of the tongue muscles during vowels in /əpVp/ environment," Ann. Bull. RILP **22**, 7–19.

Bathe, K.-J. (**1995**). *Finite Element Procedures*, 2nd ed. (Prentice Hall, Englewood Cliffs, NJ).

Beckman, M. E., Jung, T.-P., Lee, S.-L., de Jong, K., Krishnamurthy, A. K., Ahalt, S. C., Cohen, K. B., and Collins, M. J. (**1995**). "Variability in the production of quantal vowels revisited," J. Acoust. Soc. Am. **97**, 471–490.

Boë, L.-J., Granat, J., Autesserre, D., Perrier, P., and Peyre, É. (**2006**). "Variation et prédiction de la position de l'os hyoïde de l'Homme moderne (Variation and prediction of the hyoid bone position for modern man)," Biom. Hum. Anthropol. **24**, 257–272.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J.-P. (**1986**). *Cinéradiographie des voyelles et des consonnes du français (Cineradiography of vowels and consonants in French)* (Institut de Phonétique, Université Marc Bloch, Strasbourg, France).

Buchaillard, S., Brix, M., Perrier, P., and Payan, Y. (**2007**). "Simulations of the consequences of tongue surgery on tongue mobility: Implications for speech production in post-surgery conditions," Int. J. Med. Robot. Comp. **3**, 252–261.

Buchaillard, S., Perrier, P., and Payan, Y. (**2006**). "A 3D biomechanical vocal tract model to study speech production control: How to take into account the gravity?," in Proceedings of the Seventh International Seminar on Speech Production, Ubatuba, Brazil, pp. 403–410.

Dang, J., and Honda, K. (**2004**). "Construction and control of a physiological articulatory model," J. Acoust. Soc. Am. **115**, 853–870.

Duck, F. A. (**1990**). *Physical Properties of Tissues: A Comprehensive Reference Book* (Academic, London).

Fang, Q., Fujita, S., Lu, X., and Dang, J. (**2008**). "A model based investigation of activation patterns of the tongue muscles for vowel production," in Proceedings of InterSpeech 2008, Brisbane, Australia, pp. 2298–2301.

Feldman, A. G. (**1986**). "Once more on the equilibrium-point hypothesis (λ model) for motor control," J. Motor Behav. **18**, 17–54.

Feldman, A. G., and Latash, M. L. (**2005**). "Testing hypotheses and the advancement of science: Recent attempts to falsify the equilibrium point hypothesis," Exp. Brain Res. **161**, 91–103.

Fujimura, O., and Kakita, Y. (**1979**). "Remarks on quantitative description of lingual articulation," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, San Diego), pp. 17–24.

Gérard, J.-M., Ohayon, J., Luboz, V., Perrier, P., and Payan, Y. (**2005**). "Non-linear elastic properties of the lingual and facial tissues assessed by indentation technique. Application to the biomechanics of speech production," Med. Eng. Phys. **27**, 884–892.

Gérard, J.-M., Perrier, P., and Payan, Y. (**2006**). "3D biomechanical tongue modelling to study speech production," in *Speech Production: Models, Phonetic Processes, and Techniques*, edited by J. Harrington and M. Tabain (Psychology, New York), pp. 85–102.

Gérard, J.-M., Wilhelms-Tricarico, R., Perrier, P., and Payan, Y. (**2003**). "A 3D dynamical biomechanical tongue model to study speech motor control," Recent Res. Dev. Biomech. **1**, 49–64.

Gilbert, R. J., Napadow, V. J., Gaige, T. A., and Wedeen, V. J. (**2007**). "Anatomical basis of lingual hydrostatic deformation," J. Exp. Biol. **210**, 4069–4082.

Gomi, H., and Kawato, M. (**1996**). "Equilibrium-point control hypothesis examined by measured arm-stiffness during multi-joint movement," Science **272**, 117–120.

Gribble, P. L., and Ostry, D. J. (**1999**). "Compensation for interaction torques during single- and multijoint limb movement," J. Neurophysiol. **82**, 2310–2326.

Harshman, R., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693–713.

Hashimoto, K., and Suga, S. (**1986**). "Estimation of the muscular tensions of the human tongue by using a three-dimensional model of the tongue," J. Phys. Soc. Jpn. **7**, 39–46.

Hinder, M. R., and Milner, T. E. (**2003**). "The case for an internal dynamics model versus equilibrium point control in human movement," J. Physiol. (London) **549**, 953–963.

Honda, K. (**1996**). "The organization of tongue articulation for vowels," J. Phonetics **24**, 39–52.

Hoole, P. (**1998**). "Modelling tongue configuration in German vowel production," in Proceedings of the Fifth International Conference on Spoken Language and Processing, Sidney, Australia, Vol. **5**, pp. 1867–1870.

Huxley, A. F. (**1957**). "Muscle structure and theories of contraction," Prog. Biophys. Biophys. Chem. **7**, 255–318.

Ito, T., Murano, E. Z., and Gomi, H. (**2004**). "Fast force-generation dynamics of human articulatory muscles," J. Appl. Physiol. **96**, 2318–2324.

Jackson, M. T. (**1988**). "Analysis of tongue positions: Language-specific and cross-linguistic models," J. Acoust. Soc. Am. **84**, 124–143.

Kakita, Y., Fujimura, O., and Honda, K. (**1985**). "Computation of mapping from muscular contraction patterns to formant patterns in vowel space," in *Phonetic Linguistics*, edited by V. A. Fromkin (Academic, Orlando, FL), pp. 133–144.

Kapur, K. K., and Soman, S. D. (**1964**). "Masticatory performance and efficiency in denture wearers," J. Prosthet. Dent. **14**, 687–694.

Laboissière, R., Ostry, D. J., and Feldman, A. G. (**1996**). "The control of multi-muscle systems: Human jaw and hyoid movements," Biol. Cybern. **74**, 373–384.

Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence

from the analysis and synthesis of vocal-tract shapes using an articulatory model," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrect), pp. 131–149.
Miyawaki, K., Hirose, H., Ushijima, T., and Sawashima, M. (**1975**). "A preliminary report on the electromyographic study of the activity of lingual muscles," Ann. Bull. RILP **9**, 91–106.
Mokhtari, P., Kitamura, T., Takemoto, H., and Honda, K. (**2007**). "Principal components of vocal-tract area functions and inversion of vowels by linear regression of cepstrum coefficients," J. Phonetics **35**, 20–39.
Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., and Pape, D. (**2004**). "An EMMA and EPG study on token-to-token variability," AIPUK **36**, 47–63.
Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., and Payan, Y. (**2001**). "The control of token-to-token variability: An experimental and modeling study," in Proceedings of the Fourth International Speech Motor Conference, Nijmegen, The Netherlands.
Niimi, S., Kumada, M., and Niitsu, M. (**1994**). "Functions of tongue-related muscles during production of the five Japanese vowels," Ann. Bull. RILP **28**, 33–40.
Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross-linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707–3717.
Otsuka, R., Ono, T., Ishiwata, Y., and Kuroda, T. (**2000**). "Respiratory-related genioglossus electromyographic activity in response to head rotation and changes in body position," Angle Orthod. **70**, 63–69.
Payan, Y., and Perrier, P. (**1997**). "Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the equilibrium point hypothesis," Speech Commun. **22**, 185–205.
Perkell, J. S. (**1969**). *Physiology of Speech Production: Results and Implication of a Quantitative Cineradiographic Study* (MIT, Cambridge, MA).
Perkell, J. S. (**1974**). "A physiologically oriented model of tongue activity in speech production," Ph.D. thesis, Massachusetts Institute of Technology, Boston, MA.
Perkell, J. S. (**1996**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically oriented modeling," J. Phonetics **24**, 3–22.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233–272.
Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.
Perrier, P. (**2006**). "About speech motor control complexity," in *Speech Production: Models, Phonetic Processes, and Techniques*, edited by J. Harrington and M. Tabain (Psychology, New York), pp. 13–26.
Perrier, P., Ma, L., and Payan, Y. (**2005**). "Modeling the production of VCV sequences via the inversion of a biomechanical model of the tongue," in Proceedings of the Ninth European Conference on Speech Communication and Technology (Interspeech'2005), Lisbon, Portugal, pp. 1041–1044.
Perrier, P., Ostry, D. J., and Laboissière, R. (**1996**). "The equilibrium point hypothesis and its application to speech motor control," J. Speech Hear. Res. **39**, 365–378.
Perrier, P., Payan, Y., Zandipour, M., and Perkell, J. S. (**2003**). "Influence of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study," J. Acoust. Soc. Am. **114**, 1582–1599.
Perrier, P., Perkell, J., Payan, Y., Zandipour, M., Guenther, F., and Khalighi, A. (**2000**). "Degrees of freedom of tongue movements in speech may be constrained by biomechanics," in Proceedings of the Sixth International Conference on Spoken Language Processing, ICSLP'2000, Beijing, China.
Sanguineti, V., Laboss ière, R., and Ostry, D. J. (**1998**). "A dynamic biomechanical model for neural control of speech production," J. Acoust. Soc. Am. **103**, 1615–1627.
Schwartz, J. L., and Escudier, P. (**1987**). *The Psychophysics of Speech Perception* (Martinus Nijhoff, Dordrecht), pp. 284–292.
Shiller, D. M., Ostry, D. J., and Gribble, P. L. (**1999**). "Effects of gravitational load on jaw movements in speech," J. Neurosci. **19**, 9073–9080.
Shiller, D. M., Ostry, D. J., Gribble, P. L., and Laboissière, R. (**2001**). "Compensation for the effects of head acceleration on jaw movement in speech," J. Neurosci. **21**, 6447–6456.
Slaughter, K., Li, H., and Sokoloff, A. J. (**2005**). "Neuromuscular organization of the superior longitudinalis muscle in the human tongue. I. Motor endplate morphology and muscle fiber architecture," Cells Tissues Organs **181**, 51–64.
Stone, M., Epstein, M. A., and Iskarous, K. (**2004**). "Functional segments in tongue movement," Clin. Linguist. Phonetics **18**, 507–521.
Stone, M., and Lundberg, A. (**1996**). "Three-dimensional tongue surface shapes of English consonants and vowels," J. Acoust. Soc. Am. **99**, 3728–3737.
Stone, M., Stock, G., Bunin, K., Kumar, K., Epstein, M., Kambhamettu, C., Li, M., Parthasarathy, V., and Prince, J. (**2007**). "Comparison of speech production in upright and supine position," J. Acoust. Soc. Am. **122**, 532–541.
Takemoto, H. (**2001**). "Morphological analysis of the human tongue musculature for three-dimensional modelling," J. Speech Lang. Hear. Res. **44**, 95–107.
Tiede, M. K., Masaki, S., and Vatikiotis-Bateson, E. (**2000**). "Contrasts in speech articulation observed in sitting and supine conditions," in Proceedings of the Fifth International Seminar on Speech Production, pp. 25–28.
van Eijden, T. M. G. J., Korfage, J. A. M., and Brugman, P. (**1997**). "Architecture of the human jaw-closing and jaw-opening muscles," Anat. Rec. **248**, 464–474.
Weir, A. D., McCutcheon, M. J., and Flege, J. E. (**1993**). "A comparison of formant frequencies for vowels pronounced in the supine and upright positions," in Proceedings of the 12th Southern Biomedical Engineering Conference, pp. 188–190.
Wilhelms-Tricarico, R. (**1995**). "Physiological modeling of speech production: Methods for modeling soft-tissue articulators," J. Acoust. Soc. Am. **97**, 3085–3098.
Wilhelms-Tricarico, R. (**2000**). "Development of a tongue and mouth floor model for normalization and biomechanical modelling," in Proceedings of the Fifth Speech Production Seminar and CREST Workshop on Models of Speech Production, Kloster Seeon, Bavaria, pp. 141–148.
Yuen, I., Lee, A., and Gibbon, F. (**2007**). "Lingual contact in selected English vowels and its acoustic consequence," in Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Talker-to-listener distance effects on speech production and perception

Harold A. Cheyne[a)]
Sensimetrics Corporation, 14 Summer Street, Suite 403, Malden, Massachusetts 02148

Kaustubh Kalgaonkar and Mark Clements
Center for Signal and Image Processing, Georgia Institute of Technology, Atlanta, Georgia 30332

Patrick Zurek
Sensimetrics Corporation, 48 Grove Street, Suite 305, Somerville, Massachusetts 02144

(Received 29 February 2008; revised 24 July 2009; accepted 25 July 2009)

Simulating talker-to-listener distance (TLD) in virtual audio environments requires mimicking natural changes in vocal effort. Studies have identified several acoustic parameters manipulated by talkers when varying vocal effort. However, no systematic study has investigated vocal effort variations due to TLD, under natural conditions, and their perceptual consequences. This work examined the feasibility of varying the vocal effort cues for TLD in synthesized speech and real speech by (a) recording and analyzing single word tokens spoken at 1 m $\leq$ TLD $\leq$ 32 m, (b) creating synthetic and modified speech tokens that vary in one or more acoustic parameters associated with vocal effort, and (c) conducting perceptual tests on the reference, synthetic, and modified tokens to identify salient cues for TLD perception. Measured changes in fundamental frequency, intensity, and formant frequencies of the reference tokens across TLD were similar to other reports in the literature. Perceptual experiments that asked listeners to estimate TLD showed that TLD estimation is most accurate with real speech; however, large standard deviations in the responses suggest that reliable judgments can only be made for gross changes in TLD.
 [DOI: 10.1121/1.3205400]

PACS number(s): 43.70.Fq, 43.70.Mn, 43.71.Bp, 43.72.Ar [ADP] Pages: 2052–2060

I. INTRODUCTION

Current virtual audio environments can effectively simulate the direction to a sound source relative to the listener but have difficulty simulating distance. Using distance cues such as overall level and direct-to-reverberant ratio requires the listener to have prior knowledge of the intensity of the source and/or the reverberance of the acoustic environment. One class of sounds with which listeners are very familiar is speech from a conversant. Listeners know from experience that a talker adjusts his or her vocal effort to compensate for the acoustic loss due to the distance separating them. A listener can use the learned acoustic correlates of vocal effort, together with the received sound level, to estimate the talker-to-listener distance (TLD). In order to exploit this prior knowledge of vocal effort cues for improving the simulation of talker distance in virtual audio environments, it would be necessary to have a means of imposing those cues onto either synthesized or processed speech.

Several recent investigations have sought to characterize the role of particular acoustic parameters in the perception of vocal effort as it relates to distance. These data were used as the basis for formulating the hypotheses described below. One set of experiments performed by Brungart and Scott[1] tested the role of talker speech intensity (production level) vs intensity at the listener (playback level) in a distance identification listening task. Talkers were recorded in an anechoic room with a microphone 1 m from their mouth. They were instructed to produce single words and short phrases at selected sound pressure level (SPL) thresholds in 6 dB intervals. Then, the speech was played back over headphones to a listener in a large open field, where the listener had several distance markers serving as visual cues. They found that perceived TLD doubled with every 8 dB increase in production level for levels greater than or equal to 66 dB SPL at 1 m. Brungart *et al.*[2] then investigated how fundamental frequency (f0) affects the perception of distance by pitch-shifting speech such that the "up-shifted" or "down-shifted" speech had a mean f0 that matched unmodified speech produced at a 6 dB higher or lower level, respectively. The listeners again performed a distance identification task with up-shifted, unmodified, and down-shifted speech. Their results indicated that down-shifted speech provides at least half of the perceived TLD change when compared to unmodified speech above 78 dB SPL at 1 m, while up-shifted speech provides less than half of the perceived TLD change when compared to the unmodified speech.

Liénard and Di Benedetto[3] recorded isolated French vowels at three different indoor TLDs: 0.4, 1.5, and 6 m. Vocal effort at each distance was first established by a qualitative mutual comprehension exercise, in which the listener (experimenter) and the talker engaged in a brief dialog. Their analysis concentrated on the spectral properties of the vowels that changed with vocal effort as a function of TLD. The data

[a)]Author to whom correspondence should be addressed. Electronic mail: harold@sens.com

suggested that increases in fundamental frequency (f0), first formant frequency (F1), and first through third formant amplitudes (A1, A2, and A3) occur as the TLD increases.

Eriksson and Traunmüller[4] also used isolated vowels as stimuli, with the changes in vocal effort being varied by the distance between talker and experimenter and being evaluated by the experimenters' judgments. Similar to Brungart's work, they conducted distance identification listening tasks, but with two separate tasks for the listeners: (a) estimate the talker-to-receiver (communication) distance, where the listener in the perceptual experiment is not the receiver, and (b) estimate the talker-to-listener (listening) distance, where the recorded talker speaks to the listener in the perceptual experiment. Their results suggested that listener's memory is more robust in estimating the communication distance, which they believe is related to the talker's vocal effort than in estimating the listening distance, which they believe is related to the overall acoustic losses due to distance. They also noted increases in f0 and F1 and decreases in spectral tilt (the relation of the formant amplitudes), which corroborated the results of Liénard and Di Benedetto.[3]

Traunmüller and Eriksson[5] investigated segmental and suprasegmental changes that occur with increasing vocal effort as their subjects spoke one sentence (in Swedish) over increasing TLDs in the set {0.3, 1.5, 7.5, 37.5, 187.5 m}. They reported results from 12 adults (6 female, 6 male) and 8 children (4 female, 4 male) showing increases in intensity, fundamental frequency, F1, and vowel durations as TLD increases. They sampled the largest range of TLDs of the studies reviewed, but not with much resolution.

Tassa and Liénard[6] used speech modification techniques to transform speech produced with one level of vocal effort to another. Demonstrations, but no data, are available at their website, suggesting that intensity, pitch, and segment durations are much more salient cues for vocal effort than formant frequencies or spectral tilt.

Other studies that manipulated vocal effort, but did not directly study the effect of TLD on vocal effort, have used masking noise to elicit changes in subjects' vocal effort[7,8] or asked the subjects to achieve quantitative[9] or qualitative[10] target levels (e.g., using a sound level meter or requesting "maximum vocal effort").

This work presents a systematic, comprehensive study of the effect of TLD on vocal effort by combining aspects of different past studies with logical extensions of their methods. The goals of this study are to quantify changes in vocal effort that occur in actual conditions with varying TLD (hereafter termed "production TLD") and to better understand how such changes in vocal effort affect the perception of TLD (hereafter termed "perceived TLD"). Specifically, this work included

(1) conducting the collection of the speech corpus in a large, open field, with actual talkers and listeners, rather than using pre-recorded speech; this better captures the natural changes that occur in vocal effort with variations in production TLD than some previous studies' methods;
(2) sampling the production TLD space with more resolution than previous studies (e.g., Liénard and Di Benedetto[3]) while also sampling a large range of production TLD like Brungart and Scott;[1]
(3) establishing comprehension at each production TLD through a reading task requiring a qualitative response from an actual listener, similarly to Liénard and Di Benedetto;[3]
(4) at each production TLD, having the talkers speak a set of isolated words that contain a wide range of vowels and consonants;
(5) maintaining the natural changes in vocal effort due to production TLD through a two-alternative, forced-choice comprehension task for each word spoken; and
(6) for perceived TLD, asking listeners to estimate the communication distance and the listening distance in separate tasks, similarly to Eriksson and Traunmüller.[4]

This study used a formant-based speech synthesizer and describes modifications to a standard vocoder algorithm used to produce synthetic speech and modified real speech with varied vocal effort intended to mimic changes that occur in production TLD. Last, results from preliminary perceptual studies of the real, modified, and synthesized speech are given.

II. METHODS

A. Reference recordings

The reference recordings were conducted in a large, open farm field. Four participants, males (P1, P2) and females (P3, P4), each performed in talker-listener pairs: P1-P2, P2-P3, P3-P4, and P4-P1. At each trial, one talker spoke three randomized lists of 24 words at a predetermined random production TLD. Three subsequent trials would then elapse before that participant would be the talker again to minimize vocal fatigue. None of the participants had prior voice training, but participants P1, P2, and P3 knew each other through work, and P2 and P4 knew each other casually. This familiarity among the participants did not appear to influence their effectiveness in communication, as the method for eliciting appropriate vocal effort described below resulted in only three comprehension mistakes, all at the 16 m production TLD: two during the third repetition for talker P3 and listener P4, and one during the second repetition of talker P1 and listener P2.

To establish communication at each distance between a talker-listener pair, a preamble of five sentences, taken from the Connected Speech Test,[11] was spoken by the talker. The listener then provided qualitative feedback (e.g., "I can't hear you"), and additional preambles were spoken as needed until comprehension was established. Speech tokens were then collected using 12 pairs of single words from the Diagnostic Rhyme Test (DRT), repeated three times by the four talkers at 11 production TLDs in the set {1, 1.4, 2, 2.8, 4, 5.6, 8, 11.2, 16, 22.4, 32 m}, for a total of 3168 words. Talkers were given three randomized lists of the 24 words on paper at each production TLD. Listeners circled the word they heard from a list of pairs matched to the talkers' lists, with the experimenter watching the listeners' responses. Using talker-

TABLE I. The 12 word-pair subset of the DRT used as reference speech tokens.

beat–meat	daunt–taunt	rue–you
boast–ghost	dill–gill	sag–shag
boss–moss	fan–pan	sank–thank
caught–taught	mall–shawl	wall–yawl

listener pairs promoted natural changes in vocal effort with production TLD through both subjective feedback to the talker (e.g., "You're talking too quietly.") and objective feedback via the forced-choice word discrimination task for the listener. For the three instances in which the listener marked the incorrect word in a pair, the experimenter asked the talker to repeat that word once at the end of the current 24-word list. The corpus of DRT words (see Table I) was selected to maximize the sampling of consonants and vowels while minimizing the number of tokens.

A Marantz PMD-600 digital recorder captured the signals (Fs=44.1 kHz) from two Sennheiser MKE-2 microphones during the reference recordings. One microphone was worn on the talker's head and held at a fixed 15 cm from the talker's mouth, while the other microphone was fixed on a stand 0.5 m away from the listener, approximately at the listener's ear height, along the line between the talker and listener.

Acoustic analysis of the head-mounted microphone signal involved locating the 2048-point (46.4 ms) window that contained the maximum energy during the vocalic portion of each DRT word. Over that window, the means of following parameters were calculated—intensity, fundamental frequency (f0), first through third formant frequencies (F1–F3) and amplitudes (A1–A3), and first and second harmonic amplitudes (H1 and H2). The quantity H1-H2 was taken as a correlate of open quotient: the percentage of time that the vocal folds are open during one fundamental period, after Hanson and Chuang.[12] A linear regression was applied to the changes in each parameter across production TLD as averaged over all words and talkers, except for f0 and F1 for which individual talker regressions were calculated.

Four hypotheses, listed below, were tested through the analysis of the reference recordings. Each hypothesis suggests an *increase* in a parameter with vocal effort as a function of production TLD, followed by an estimate of the rate of change:

(1) intensity, by 8 dB per doubling of production TLD;
(2) f0, by 40 Hz per doubling of production TLD;
(3) F1, by 25 Hz per doubling of production TLD; and
(4) A3-A1 (spectral tilt), by 2 dB per doubling of production TLD.

The first hypothesis presumes a larger increase than was reported from previous work because those studies did not involve objective comprehension tests between talkers and listeners, which could conceivably require talkers to increase intensity beyond what is needed for sustained vowels (Liénard and Di Benedetto) or a fixed sentence (Traunmüller and Eriksson). As such, the first hypothesis is based on the perceptual results of Brungart and Scott, while recognizing that their results reflect listeners' perceptions rather than talkers' productions. The second, third, and fourth hypotheses were dervied from Liénard and Di Benedetto's reporting of the average rate of increase with level for (a) f0 (5.1 Hz/dB), (b) F1 (3.5 Hz/dB), (c) A1 (1.1 dB/dB), and (d) A3 (1.3 dB/dB).

B. Speech modification and synthesis framework

1. High-level synthesis

High-level synthesis (HLsyn), a quasi-articulatory speech synthesis system developed at Sensimetrics Corporation,[13,14] is driven by a 13-component vector updated every 5 ms. The HLsyn concept is called "quasi-articulatory" because it is based on the observation that many parameters needed to control a formant synthesizer are not independent but are constrained by the anatomy and physiology of the human speech production system. HLsyn incorporates many of these constraints in a higher-level control system layered on top of a formant synthesizer (KLSYN88, developed by Klatt and Klatt[15]) that generates the output speech. Thirteen parameters control HLsyn and are transformed into the more than 40 KLSYN88 parameters through a set of mapping relations.

The 13 components are the fundamental frequency, the first four formants, and eight more parameters related to the physiology of speech production [e.g., the area of the velopharyngeal port (a_n), the area of the opening at the lips (a_l), the area of the tongue blade constriction (a_b), etc.]. The subglottal pressure parameter (p_s) can effect changes in f0 and overall intensity, both of which have already been identified as correlates of vocal effort.[1–4,15,16]

2. Synthetic speech

Synthetic speech tokens were generated using the HLsyn speech synthesizer (Sensimetrics Corp.). The acoustic parameters f0, intensity, A1, and spectral tilt (TL) were manipulated in the synthetic speech tokens to investigate their effect on the perceived TLD. HLsyn implements changes in TL as a low-pass filter applied to the voicing source, with TL controlling the filter's magnitude at 3 kHz relative to 300 Hz. The HLsyn stimuli modeled only one talker and were not intended to sound similar to any of the participants. One set of synthetic speech tokens only varied in f0, with the intent of identifying the relative importance of changing f0 to the other parameter changes. Similar changes in A1 and TL were not tested perceptually because those changes resulted in speech that sounded unnatural.

3. MELP vocoding

The mixed excitation linear prediction (MELP) coder is a purely parametric method that lends itself to high quality analysis/synthesis in the absence of quantization. It is the basis for the U.S. DOD standard FS-1017 (MIL-STD-3005)[17] and the recently adopted NATO standard (NATO STANAG 4591). Its novelty is based on three innovations: (a) a highly accurate pitch extraction algorithm that has subsample precision, (b) a periodic excitation waveform that has both pulse shaping and a natural jitter, and (c) an

overall excitation pattern that allows mixing periodic and non-periodic characteristics in a frequency-dependent fashion. Its parametric design is sufficiently modular so as to allow changing the excitation, the vocal tract filtering, or both. The speech modifications are carried out in two stages. First, the incoming speech is parametrized; modifications to the linear predictive coding (LPC) polynomial and excitation parameters are made to change the formant frequencies and amplitudes, spectral tilt, and pitch. In the second stage, speech is synthesized using these new parameters.

4. Modified speech using a MELP vocoder

Modified speech tokens were produced using a modified version of the MELP vocoder described above—altered to allow manipulation of f0, intensity, A1, and spectral tilt. Changing the f0 of the MELP-vocoded speech is straightforward as f0 is a parameter of the vocoder that can be directly modified. However, the standard MELP vocoding algorithm does not permit increasing f0 above 200 Hz. This limitation is discussed more in Sec. III B. The reference recordings made at TLD=1 m were processed by the vocoder to simulate production TLDs of 4, 8, 16, and 32 m.

Although the parameters of the MELP vocoder allowed independent control of the vocal source and filter, the formant amplitudes could not be directly modified from the LPC polynomial $A(f)$. To overcome this limitation, the formant bandwidths were modified to indirectly change their amplitudes. To modify the formant bandwidths, each formant was modeled as a two pole digital resonator with a pair of complex conjugate poles located near the unit circle,

$$p_{1,2} = re^{\pm j\omega_0}. \tag{1}$$

The amplitude of the transfer function of such a digital resonator is given by Eq. (2), after Proakis and Manolakis,[18]

$$|H(\omega_0)| = \frac{1}{(1-r)\sqrt{1+r^2-2r\cos(2\omega_0)}}. \tag{2}$$

To manipulate the amplitude $|H(\omega_0)|$ independently of the pole frequency, the pole radius r can be changed. For $r \approx 1$, changing r will also affect the bandwidth of the resonator B_{ω_0} according to

$$B_{\omega_0} \approx 2(1-r). \tag{3}$$

Oppenheim and Schafer[19] showed that in the LP spectrum, a formant bandwidth B_k can be approximated with Eq. (4) below, which relates the group delay to the bandwidths for single pole systems

$$B_k \approx \frac{-1}{\pi}\left[\ln\left(\frac{\partial \angle A(f)}{\partial f}\right) - \ln\left(2\pi + \frac{\partial \angle A(f)}{\partial f}\right)\right]\Bigg|_{F_k}, \tag{4}$$

where $A(f)$ is the LP spectrum given by

$$A(f) = 1 - \sum_{k=1}^{P} a_k e^{-j2\pi fk}, \tag{5}$$

the derivative of $A(f)$ is given by

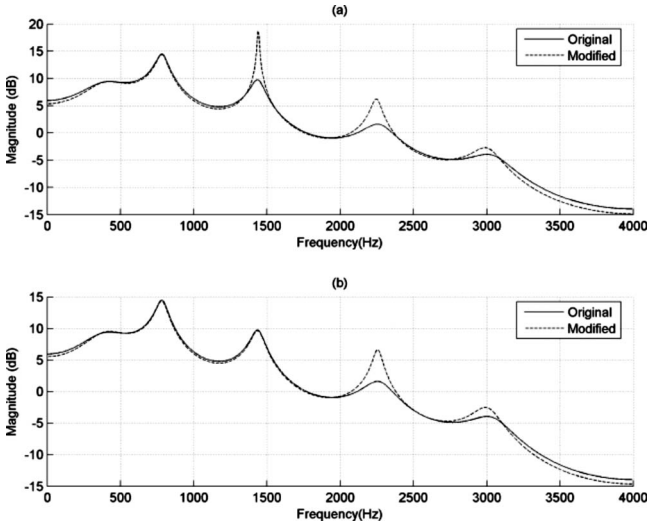


FIG. 1. Spectra for original (solid) and modified (dashed) LPC polynomials, showing increases in (a) A2, A3, and A4 by 8, 4, and 2 dB, respectively, and (b) A3 and A4 by 4 and 2 dB.

$$\frac{\partial A(f)}{\partial f} = j2\pi \sum_{k=1}^{P} a_k e^{-j2\pi fk}, \tag{6}$$

and the derivative of $\angle A(f)$ is given by

$$\frac{\partial \angle A(f)}{\partial f} = \mathrm{Im}\left\{\frac{\frac{\partial A(f)}{\partial f}}{A(f)}\right\}. \tag{7}$$

Setting Eq. (3) equal to Eq. (4) provides a method for approximating the formant pole radius r without explicit root solving, as expressed in

$$r \approx 1 + \frac{1}{2\pi}\left[\ln\left(\frac{\partial \angle A(f)}{\partial f}\right) - \ln\left(2\pi + \frac{\partial \angle A(f)}{\partial f}\right)\right]\Bigg|_{F_k}. \tag{8}$$

Thus, the algorithm to alter the amplitude of a formant in a LPC spectrum can be stated as follows:

(1) Approximate the radius of the formant r using Eq. (8) above.
(2) Choose a new radius $\tilde{r}=r+\Delta r$ to model the desired formant amplitude $|\tilde{H}(\omega_0)|$ using Eq. (2).
(3) Calculate the new formant bandwidth using Eq. (3).
(4) Use the LSP-based formant bandwidth modification method described by Morris and Clements[20] to obtain the new LPC coefficients.

Spectral tilt and gain were changed by manipulating multiple formant amplitudes simultaneously. Figure 1 shows example spectra using the above modification method.

C. Perceptual testing

The perceptual testing presented listeners with randomized speech tokens from one of the three sets of tokens: the reference recordings, synthetic speech, or modified speech. The listeners heard the tokens in a quiet office environment (ambient noise below 60 dBA) over headphones (Sennheiser

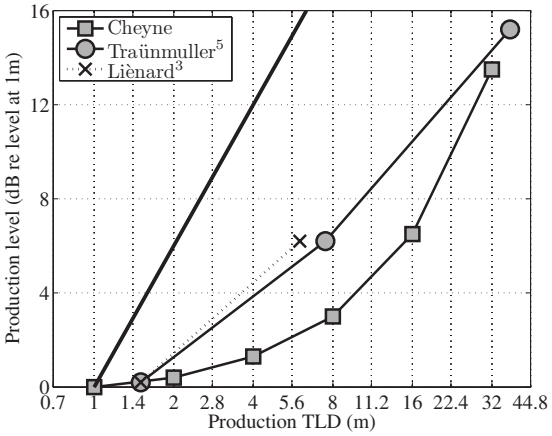


FIG. 2. Production level (dB) relative to the level at 1 m as a function of production TLD (m): mean data from this study, Liénard and Di Benedetto (Ref. 3), and Traunmüller and Eriksson (Ref. 5). The legend is limited to first authors to conserve space.

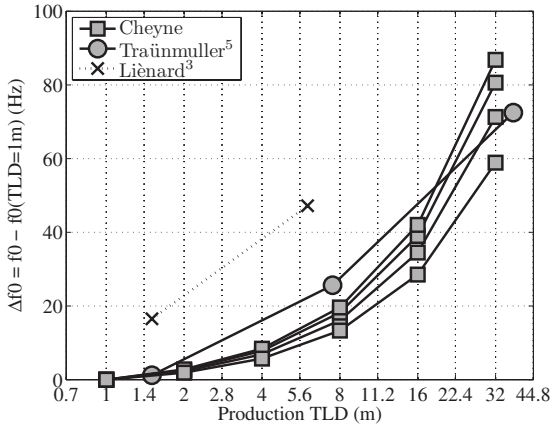


FIG. 3. Change in fundamental frequency (f0) relative to f0 at production TLD=1 m, as a function of production TLD (m): mean data from this study, Liénard and Di Benedetto (Ref. 3), and Traunmüller and Eriksson (Ref. 5). The legend is limited to first authors to conserve space.

HD201) with two corrections applied to the tokens: a head-related transfer function[21] to produce a binaural signal from the single channel original token and a frequency compensation function to flatten the headphone response to ±3 dB over 125–11025 Hz. The listeners were asked to identify the distance between the talker and the conversant by sliding a bar on a computer screen that represented the range of 1–32 m. Visual anchors in the form of scaled pictures of a talker at 1 and 32 m were displayed next to the corresponding ends of the slider.

Two contexts were presented to the listeners in separate experiments: first, where the listener is the conversant (the "two-party" case), a situation in which the speech signals contain the full set of distance-related acoustic cues; second, where a third party is the conversant at a variable distance while the listener is at a fixed distance of 1 m from the talker (the "three-party" case), a situation that conveys vocal effort cues but not the level variation due to distance. In the two-party case, the listeners heard the token at its original recorded intensity, whereas in the three-party case the token intensity was attenuated to simulate the 1 m fixed distance. The two questions were asked separately; that is, after one subject had listened to an entire set of reference tokens and judged the two-party distance, they listened to the set again without attenuation due to distance and were asked to estimate the three-party distance. For audio samples of the speech tokens, visit the website www.sens.com/distance_cues/samples.

III. RESULTS AND DISCUSSION

A. Reference recordings

Of the parameters listed above in Sec. II A, those found to vary systematically with production TLD were intensity, f0, and spectral tilt (A3–A1).

The change in intensity vs production TLD is shown in Fig. 2, with a comparison of this study's data to that of Liénard and Di Benedetto[3] and Traunmüller and Eriksson.[5] The data were normalized to production TLD=1 m for the comparison because the intensity was not calibrated to absolute SPL for those other studies. The absolute intensity in this study ranged from 82.4 to 96.7 dB SPL at 15 cm, which is equivalent to 65.9 to 80.2 dB SPL at 1 m assuming a spherical source. Overall, the data from the three studies display similar increases in intensity with production TLD that is shallower than would be required to simply compensate for the 6 dB per TLD doubling acoustic loss (shown as the solid dark line in Fig. 2). Furthermore, except for the 7 dB increase in the 16–32 m doubling of production TLD, the increases were much less than the hypothesized 8 dB per doubling, which was based on perceptual data from Brungart and Scott[1] showing that an 8 dB increase in production level results in a doubling of perceived TLD. Differences in methodology likely explain this discrepancy, as this study's reference recordings used actual talker and listener pairs with the goal of comprehension at each production TLD, and the talkers' maximum SPL was about 80 dB at 1 m. In contrast, Brungart and Scott recorded talkers at different target intensities (production levels, not production TLDs) ranging from 35 to 95 dB at 1 m and then played the recorded speech to listeners in a field with markers along the 0.25–64 m range. Their expanded SPL range, particularly above 80 dB SPL at 1 m, contributed to their finding of the 8 dB increase in production level per perceived TLD doubling.

Fundamental frequency changes with production level, averaged over all DRT words, are displayed in Fig. 3 for the individual talkers in this study, again compared to the mean data from previous studies.[3,5] For easier comparison, the data have been normalized to the f0 at production TLD=1 m. The increases seen in f0 with production TLD from this study were not as large as expected based on the results of Brungart *et al.*[2] and Liénard and Di Benedetto,[3] which led to the second hypothesis. Even the largest increases in f0 were only about one-half of what was hypothesized. However, this study's individual data are very similar to those of Traunmüller and Eriksson,[5] suggesting that using the perceptual data of Brungart and Scott[1] to extrapolate f0 increases outside the range used by Liénard and Di Benedetto[3] overestimated the actual increases in f0 with production TLD. Comparing this study's results to changes in vocal effort elicited

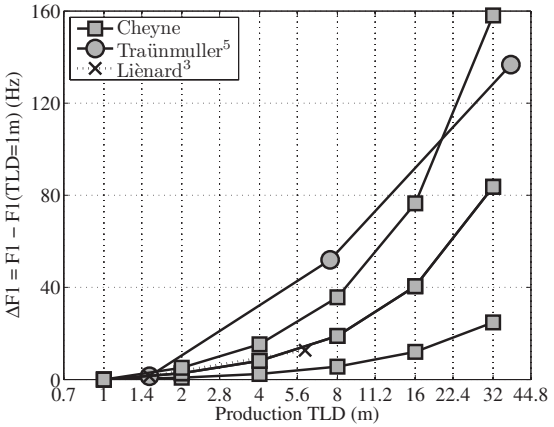


FIG. 4. Change in first formant frequency (F1) relative to F1 at production TLD=1 m as a function of production TLD (m): mean data from this study, Liénard and Di Benedetto (Ref. 3), and Traunmüller and Eriksson (Ref. 5). The legend is limited to first authors to conserve space.

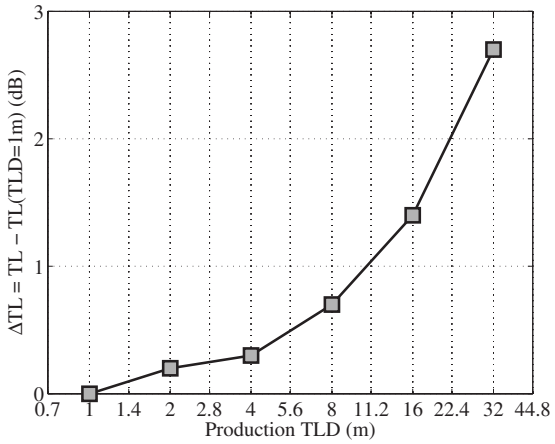


FIG. 5. Change in spectral tilt (TL) relative to TL at production TLD =1 m as a function of production TLD (m): mean data from this study.

by other experimental conditions, Van Summers *et al.*[8] and Huber *et al.*[9] both found more gradual increases in f0 with speech intensity when using masking noise and intensity targets, respectively. Rostolland[10] studied comfortable and shouted voice—with an intensity difference of 20 dB—and reported increases in f0 that are larger than the rate of change with intensity in this study. In contrast, Bond *et al.*[7] also studied speech in masking noise and found increases in f0 with intensity similar to those for P1 and P4.

First formant frequency (F1) changes with production TLD were not as consistent across talkers as the f0 changes, as shown in Fig. 4. Again, the data are normalized to a production TLD of 1 m for ease of comparison with other investigators' data. Note that the data for P2 and P3 were so nearly identical that their square markers overlap as the middle data set from this study in Fig. 4. The maximum increase in F1 for a doubling of production TLD ranged from 12 Hz (P1) to 82 Hz (P4), bracketing the hypothesized 25 Hz (based on the work of Eriksson and Traunmüller,[4] Lienard and DiBenedetto,[3] and Traunmüller and Eriksson[5]). However, increases in F1 exceeding 25 Hz per production TLD doubling only occurred at the larger production TLDs: for the 5.6–11.2 m and 8–16 m doublings of talker P4 and for the 11.2–22.4 m and 16–32 m doublings for talkers P2, P3, and P4. For a comparison with other methods of eliciting changes in vocal effort, Van Summers *et al.*[8] reported changes in F1 in two male participants' speech in quiet and under increasing masking noise, while Huber *et al.*[9] reported changes in F1 with intensity for sustained /a/ vowels. One of the participants in the study of Van Summers *et al.*[8] showed increases in F1 with increasing vocal effort resulting from the increasing masking noise, while the other showed essentially no change in F1. The female participants in the study of Huber *et al.*[9] showed an average increase in F1 that was very similar to the increases shown by P1, P2, and P3 in this study. In contrast, the males exhibited an average decrease in F1 with increasing intensity.

The spectral tilt (TL=A3-A1) did increase with production TLD but not at the rate hypothesized above. Figure 5 shows the mean change in TL over production TLD, again normalized to a production TLD of 1 m. The hypothesized 2 dB increase per distance doubling, based on the formant amplitude changes noted by Liénard and DiBenedetto,[3] was only nearly approached at the final TLD doubling from 16 to 32 m. Perhaps their findings reflect different habitual degrees of TL in French speech as compared to American English. Compared to changes in TL with masking noise from the data of Van Summers *et al.*,[8] P1–P4 increased TL more gradually with intensity, suggesting that the task difference—increasing vocal effort to overcome noise rather than distance—may lead the talkers to employ different strategies acoustically.

B. MELP-vocoded speech tokens

The modified MELP vocoder processed some of the reference recordings from a production TLD of 1 m to produce tokens that mimicked the changes in f0 and intensity that occurred for production TLDs of 4, 8, 16, and 32 m. Tokens intended to mimic a production TLD of 1 m were obtained by passing the reference tokens through the vocoder without modification.

In the MELP standard used for this study,[17] f0 is restricted to fall within a fairly narrow range, with a maximum of 200 Hz. This excludes some f0 values measured from the reference recordings, particularly at a production TLD of 32 m. Modifying the MELP framework to expand the f0 range is non-trivial, as the framing, short-term predictors, long-term predictors, harmonic voiced-unvoiced assignments, and pulse dispersion would all be affected. As such, only f0 values that were below 200 Hz in the reference recordings were used for producing MELP-modified tokens. With that limitation, only the words "dill," "ghost," "mall," "meat," "shag," "taught," and "you" were modified by the MELP algorithm as they were spoken by all talkers with f0 below 200 Hz. For male speakers, f0 was increased by 30 Hz to model a production TLD of 16 m and by 65 Hz for 32 m. For female speakers, f0 was increased by 20 and 40 Hz for production TLDs of 16 and 32 m, respectively.

Although the MELP controls can be used to modify formant bandwidths, the bandwidths of all poles must be modified to effect the desired change in amplitude of even one

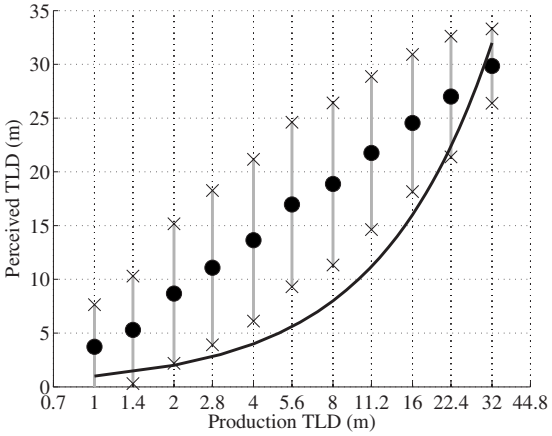


FIG. 6. Perceived TLD (m) vs production TLD (m) data for the two-party case with reference tokens. Circles are means, bounded lines are ±1 s.d., black curve is 1:1 correspondence.

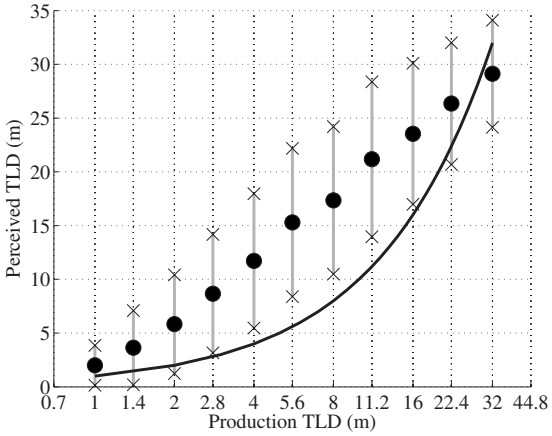


FIG. 8. Perceived TLD (m) vs production TLD (m) data for the two-party case with HLsyn tokens. Circles are means, bounded lines are ±1 s.d., black curve is 1:1 correspondence.

formant. This likely follows from the anatomical changes used to alter one formant's bandwidth or amplitude, which will affect the other formants (e.g., stiffening the vocal tract walls). Further, spectral tilt also affects the formant amplitudes. For the set of MELP tokens with modified spectral tilt and gain, the spectral tilt was increased by 2 dB/oct to synthesize vocal effort for a production TLD of 16 m and 4 dB/oct for 32 m. The spectral gain was increased by 8 and 15 dB for synthesizing the vocal effort for production TLDs of 16 and 32 m, respectively. Unfortunately, some formant and f0 modifications, particularly for the higher f0 examples, resulted in modified speech that sounded unnatural, possibly confounding the perceptual results.

C. Perceptual testing

Figures 6 and 7 show the results of the perceptual tests on the reference recording stimuli. Nine listeners participated in the reference stimuli perceptual tests, with six (L1–L6) participating in the two-party (Fig. 6) case and six (L1, L3, L4, L5, L8, and L9) participating in the three-party (Fig. 7) case. The filled circles show the mean values at each actual distance, and the lines bounded by crosses show ±1 standard deviation. The means and standard deviations are calculated across all talkers, DRT tokens, and listeners. In general, the listeners' perceived TLD increased with the production TLD, but not at a 1:1 rate as shown by the solid black curve. Ten *t*-tests, conducted only on adjacent production TLD data (e.g., 4 m vs 5.6 m), revealed significant differences at the $p=0.05$ level for every adjacent pair, not necessarily an intuitive result because of the large overlap in the adjacent distributions, but likely due to the large number of listener responses gathered—288 at each production TLD. For the two-party case, the mean perceived TLD exceeds the production TLD by as much as a factor of 4 (e.g., between production TLDs of 1 and 4 m), and that difference only drops below a factor of 2 for production TLDs of 16 m and above. Aside from the endpoints, the standard deviation (range of 5.0–7.7 m) remains relatively constant but is rather large throughout the range.

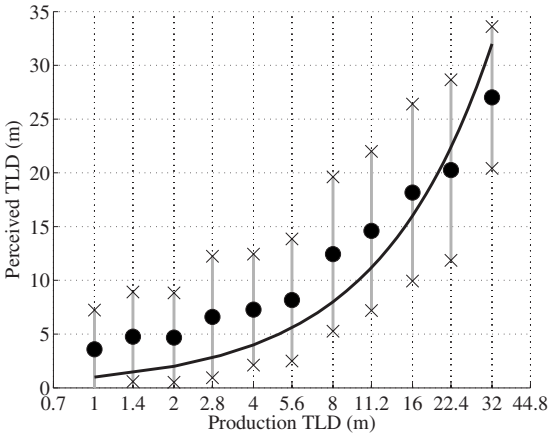


FIG. 7. Perceived TLD (m) vs production TLD (m) data for the three-party case with reference tokens. Circles are means, bounded lines are ±1 s.d., black curve is 1:1 correspondence.

For the three-party case, there is less of a tendency for overestimating distance, and the standard deviation is similar to the two-party case (range of 4.3–7.3 m, excepting endpoints). However, there is more confusion (overlap) among adjacent distances, especially in the 1–2 m range and the 2.8–5.6 m range. Qualitatively, the overlap in adjacent distributions appears to group the responses into five quanta: 1–2, 2.8–5.6, 8–11.2, 16–22.4, and 32 m. Ten *t*-tests on only the adjacent response distributions revealed significant differences at the $p=0.05$ level for all pairings except 1.4 m with 2 and 2.8 m with 4 m, lending support to the grouping observation.

Figure 8 shows the perceptual results from four listeners (L1, L3, L7, and L9) for the HLsyn stimuli that vary in f0 only for the two-party distance estimates. The plot format follows that of Fig. 6. Again, the general trend shows increasing perceived TLD with increasing production TLD. Note the similarity of the HLsyn token results shown in Fig. 8 and the reference recording results of Fig. 6, suggesting that attenuation due to production TLD and changes in f0 are salient cues, an observation supported by the work of Brungart *et al.* and Brungart and Scott. Again, ten *t*-tests conducted only on adjacent production TLD response distributions resulted in significant differences at the $p=0.05$ level

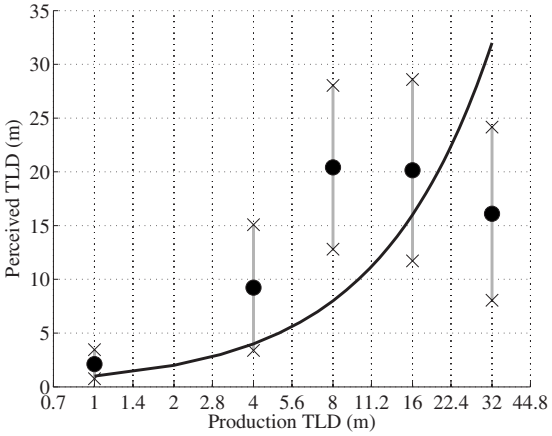


FIG. 9. Perceived TLD (m) vs production TLD (m) data for the two-party case with MELP-modified tokens. Circles are means, bounded lines are ±1 s.d., black curve is 1:1 correspondence.

for every pairing, just as with the reference stimuli. The perceived TLD exceeds the production TLD by as much as a factor of 2 (e.g., at 1 m), except for the 32 m actual distance.

The three-party case was not tested for the HLsyn stimuli due to listeners' reports of the synthesized stimuli sounding unnatural, with the unnaturalness increasing with perceived TLD. The consistency of these reports of unnaturalness suggested the potential of listeners categorizing their responses based on some feature of the synthesized speech that was not intended to evoke a perceptual change in TLD, which would lead to confusing results. Thus, the three-party tests were postponed until more natural HLsyn tokens could be generated, and the two-party data must be interpreted with this caveat.

Initial results of one listener's (L1) perceived TLD for the MELP-modified speech are shown in Figs. 9 and 10 for the two-party and three-party cases, respectively. The stimuli for these tests were generated as described in Sec. III B. The responses in Fig. 9 demonstrate the feasibility of the system for the TLDs of 1, 4, and 8 m in that the mean perceived TLD increases with the production TLD, although the overestimation is similar to what is seen in Figs. 6 and 8. Four *t*-tests, only between adjacent production TLD response distributions, revealed significant differences at the $p=0.05$ level for every pairing except the 8–16 m pairing. The response confusion between 8 and 16 m and the decrease in perceived TLD for the 32 m production TLD likely arose from the f0 limitation of the current MELP vocoder algorithm (see Sec. III B). Listener L1 reported that some of the MELP stimuli sounded unnatural, which could also contribute to the response confusion. Last, the authors decided that the f0 limitation and unnatural quality of some tokens needed to be addressed before additional listener data are collected for the MELP stimuli to avoid listeners using unintended cues to categorize their responses similarly to the HLsyn data.

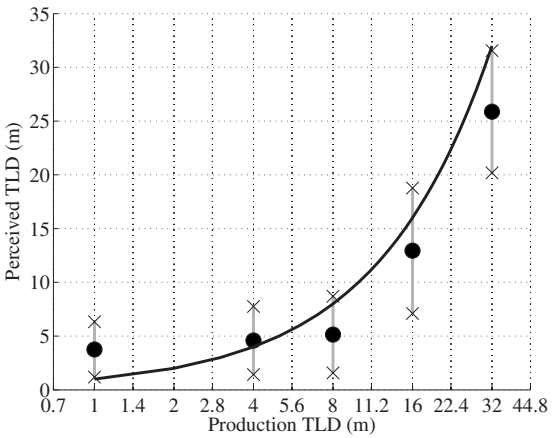


FIG. 10. Perceived TLD (m) vs production TLD (m) data for the three-party case with MELP-modified tokens. Circles are means, bounded lines are ±1 s.d., black curve is 1:1 correspondence.

Figure 10 shows the three-party results from listener L1. Note that here the response confusion appears reversed relative to that in Fig. 9: There is confusion among the 1, 4, and 8 m but distinction among the 8, 16, and 32 m production TLDs. This observation is supported by four *t*-tests on the response distributions of adjacent production TLDs, which showed significant differences at the $p=0.05$ level only for the 8–16 and 16–32 m pairings. This suggests that (for this listener at least) although the vocal effort changes in f0 may not have been realistic for the tokens at the larger production TLDs, in the absence of attenuation due to production TLD the MELP-modified vocal effort changes were sufficient for discrimination. Thus, even without the range of f0 change needed to mimic the actual changes in vocal effort, these initial results support the feasibility of the approach.

IV. CONCLUSIONS

The reference recordings obtained during this work represent the largest corpus of real speech that varies in vocal effort as it relates to production TLD. The analysis of these recordings for changes in f0, intensity, and formant amplitudes as a function of production TLD provides an initial look at what aspects of speech vary systematically with vocal effort, as vocal effort is modified to compensate for changes in production TLD. Perceptual testing of synthetic and modified speech tokens revealed that in both the two-party and three-party experimental paradigms, perceived TLD overestimates production TLD, but to a lesser extent when only the vocal effort cues are available.

Algorithms for changing vocal effort, and thus perceived TLD, would have numerous military and commercial applications. Potential military applications include (a) virtual training environments with enhanced acoustic cues for consistency with visual cues of nearby or distant talkers; (b) command-and-control operations using virtual audio displays where distance perception may enhance the separation of talkers when added to directional cues; (c) remote battlefield environments, such as several operators remotely controlling unmanned aircraft or SWORDS robots, where a virtual acoustic space incorporating distance and direction cues is necessary to maintain awareness of their machines' relative locations when the operators' physical separation from each other does not mimic that of the machines; and (d) an aid for unfamiliar route navigation, in which a virtual voice can speak to the person(s) navigating the route (e.g., "Over

here!") and the voice contains perceptual cues relating to the listener-to-target distance as well as direction. A related potential application of an algorithm to modify vocal effort would be to simulate a change in emotion, or specifically urgency. For a command-and-control operation, a command issued as a shout nearby a listener would naturally carry more urgency or import to the listener than the same command spoken conversationally.

Potential commercial applications include (a) enhanced realism in virtual audio spaces, such as motion pictures and video games, to provide auditory cues for TLD that are consistent with visual cues, (b) providing speaker separation in teleconferencing on a monaural channel; and (c) aids for the visually impaired, in which a virtual voice can be perceived at the location of a distant object of interest. Last, these algorithms could suggest solutions or partial solutions to the inverse problem: Given a speech signal picked up at a listener's location, what is the distance to the talker? This distance-estimation problem may also have potential military and commercial applications (e.g., remote machine awareness of its distance to talkers).

V. FUTURE WORK

Several potential manipulations in speech that occur with production TLD need further investigation. First, segmental duration may increase with production TLD and may vary with the type of segment. For example, vowel durations may increase more rapidly than non-sonorant consonant durations. Second, increases in f0 may not be simply a shift across the entire voiced segment but may involve a larger range of f0—perhaps starting and ending at similar f0 values but increasing the excursion during the word. Third, changes in formant amplitudes may also not be constant across the vocalic portions of the words. Fourth, the vocal tract resonances during fricatives may change with production TLD. Fifth, the intensity of the non-sonorant phonemes (e.g., fricatives) may change differently from the sonorant phonemes (e.g., vowels). While not an exhaustive list of the potential changes that may occur with speech over production TLD, these five suggestions are the most probable next steps for investigation.

Work with the MELP vocoder algorithm revealed its limitations for this application. First, the bandwidth of the MELP is restricted to 4 kHz, which limits the quality of the modified speech particularly for female speakers. Manipulation of the formant amplitudes to mimic those that occur naturally requires a method whereby all target formant amplitude values can simultaneously be altered through multivariable optimization to obtain both the required formant amplitude and spectral tilt changes with production TLD.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Nathaniel Durlach in planning the experimental procedures of this project. This work was supported by USAFOSR Contract No. FA9550-05-C-0124.

[1] D. S. Brungart and K. R. Scott, "The effects of production and presentation level on the auditory distance perception of speech," J. Acoust. Soc. Am. **110**, 425–440 (2001).

[2] D. S. Brungart, A. J. Kordik, K. Das, and A. K. Shaw, "The effects of f0 manipulation on the perceived distance of speech," in Proceedings of the International Conference on Spoken Language Processing, Denver, CO (2002), pp. 1641–1644.

[3] J.-S. Liénard and M.-G. Di Benedetto, "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422 (1999).

[4] A. Eriksson and H. Traunmüller, "Perception of vocal effort and distance from the speaker on the basis of vowel utterances," Percept. Psychophys. **64**, 131–139 (2002).

[5] H. Traunmüller and A. Eriksson, "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451 (2000).

[6] A. Tassa and J. S. Liénard, "A new approach to the evaluation of vocal effort by the PSOLA method," The European Student Journal of Language and Speech 00.01 (2000), http://www.essex.ac.uk/web-sls/papers/00-01/00-01.html (Last viewed February 29, 2008).

[7] Z. S. Bond, T. J. Moore, and B. Gable, "Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask," J. Acoust. Soc. Am. **85**, 907–912 (1989).

[8] W. Van Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes, "Effects of noise on speech production: Acoustic and perceptual analyses," J. Acoust. Soc. Am. **84**, 917–928 (1988).

[9] J. E. Huber, E. T. Stathopoulos, G. M. Curione, T. A. Ash, and K. Johnson, "Formants of children, women, and men: The effects of vocal intensity variation," J. Acoust. Soc. Am. **106**, 1532–1542 (1999).

[10] D. Rostolland, "Acoustic features of shouted voice," Acustica **50**, 118–125 (1982).

[11] R. M. Cox, G. C. Alexander, and C. A. Gilmore, "Development of the connected speech test (CST)," Ear Hear. **8**, 119S–126S (1987).

[12] H. M. Hanson and E. S. Chuang, "Glottal characteristics of male speakers: Acoustic correlates and comparison with female data," J. Acoust. Soc. Am. **106**, 1064–1077 (1999).

[13] H. M. Hanson and K. N. Stevens, "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn," J. Acoust. Soc. Am. **112**, 1158–1182 (2002).

[14] K. N. Stevens and C. A. Bickley, "Constraints among parameters simplify control of Klatt formant synthesizer," J. Phonetics **19**, 161–174 (1991).

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

[16] D. S. Brungart, "A speech-based auditory display," Proceedings of the AES 109th Convention, Los Angeles, CA (2000), pp. 1–11.

[17] T. P. Barnwell III, E. B. George, A. V. McCree, K. K. Truong, and V. R. Viswanathan, "A 2.4 kbit/s MELP coder candidate for the new U.S. federal standard," Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (1996).

[18] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications* (Prentice-Hall, Englewood Cliffs, NJ, 1996).

[19] R. W. Oppenheim and A. V. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

[20] R. Morris and M. Clements, "Modification of formants in the line spectrum domain," IEEE Signal Process. Lett. **9**, 19–21 (2002).

[21] http://sound.media.mit.edu/KEMAR.html (Last viewed August 21, 2008).

Perceived loudness of speech based on the characteristics of glottal excitation source

Guruprasad Seshadri[a)]
Dept. of Computer Science and Engineering, Indian Institute of Technology Madras, Chennai 600036, India

B. Yegnanarayana
International Institute of Information Technology, Hyderabad 500032, Andhra Pradesh, India

(Received 6 June 2008; revised 10 July 2009; accepted 13 July 2009)

The impulse-like characteristic of glottal excitation in speech production is an important factor in the perception of loudness of speech signals. This characteristic is attributed to the abruptness of the closing phase in the glottal cycle. In this paper, an acoustic feature, called strength of excitation, is proposed to represent the impulse-like nature of excitation. The strength of excitation is derived from the linear prediction residual of speech signals, where the residual can be considered as an estimate of the source of excitation. Since the loudness of speech is perceived over one or more utterances of speech, it is hypothesized that the distribution of strength of excitation is indicative of the perceived loudness of speech. The distribution of strength of excitation is shown to distinguish between soft and loud utterances of speakers. The distribution can also help in discriminating between the loudness of two speakers. The loudness measure obtained using the distribution of the strength of excitation is in agreement with the subjective judgment of loudness of speech.
 [DOI: 10.1121/1.3203668]

PACS number(s): 43.71.Gv, 43.66.Cb [DOS] Pages: 2061–2071

I. INTRODUCTION

Perception of loudness of sound in human beings is defined as the magnitude of auditory sensation, which depends on the acoustic characteristics of the sound (Fletcher and Munson, 1933). Loudness of a sound is related to the distribution of spectral energy of the sound (Fletcher and Munson, 1933; Fletcher and Munson, 1937). Temporal properties of sounds such as duration, and impulsive or rhythmic nature, also affect the perceived loudness (Zwicker, 1977; Zwicker and Fastl, 1999). The problem of measurement and calculation of loudness of sounds has been studied extensively. A method for calculating the loudness of a complex tone from its frequency spectrum was proposed by Fletcher and Munson (1933). For sounds with a large number of spectral components, the loudness of one component depends on the masking effects of the other components, particularly when the components are closely spaced (Fletcher and Munson, 1937). Methods for computing the total loudness of a sound from the values of loudness due to the constituent frequency bands were suggested by Beranek *et al.* (1951) and Stevens (1956, 1961). Similarly, Zwicker and Fastl (1999) described a procedure to compute the loudness levels of pure tones at different frequencies, and to construct equal loudness contours. In the case of pure tones or noise that has a uniform distribution of energy in a given band of frequencies, Zwicker and Fastl (1999) proposed the measurement of loudness as a function of frequency separation of two pure tones, and also as a function of bandwidth of the noise.

In applying the above methods for the calculation of loudness of *speech signals*, two issues need to be taken into account. First, the measurement of loudness in those methods was based on the response of the auditory perception mechanism to sounds which were pure tones, combination of pure tones, or noises of different bandwidths. But speech sounds cannot be approximated by such signals. In particular, speech sounds cannot be modeled well by pure tones (Warren, 1973), since the short-time spectrum of speech signal has a gross envelope with a few prominent peaks (formants) around which significant energy is concentrated, and a fine structure corresponding to the fundamental frequency and its harmonics (Fant, 1960). Second, loudness is a perceptual attribute which cannot be described merely by the amount of acoustic energy or its spectral distribution. For instance, a soft voice is perceived as soft, even if the level of speech from loudspeaker is increased. Similarly, a loud voice is perceived as loud, even in the presence of some amount of ambient noise. Thus, apart from the amplitude or energy of the speech signal, the excitation source and the vocal tract system characteristics in the signal can also affect the perception of loudness (Rothenberg, 1983). Hence, there is a need to examine the perception of loudness of speech based on the production characteristics of speech signals.

Since loudness of speech is a perceptual attribute, an exact definition of loudness is elusive. On a perceptual scale, loudness of speech varies from a weak/soft voice, to a normal/modal voice, and further to a loud voice, which can extend up to shouting. This aspect of loudness is determined, to a great extent, by the physiological characteristics of speech production mechanism of the speaker. The perceived loudness depends on the nature of the speech sound, due to loading of the vocal tract system on the vocal source during

[a)]Author to whom correspondence should be addressed. Electronic mail: gurus@cse.iitm.ernet.in

the production. Loudness is also affected by the behavioral characteristics of the speaker, such as emotional state of the speaker. The behavioral characteristics can cause variations in rhythm and rate of speech. They can also result in stress on particular syllables of words, and the stress may vary for different words in a sentence/phrase. Accentuation of stressed syllables, which causes a change in the pitch pattern of the stressed syllables relative to the nonstressed syllables, is also a factor in the perception of loudness. Thus, loudness is a perceptual attribute that is governed by both physiological and the behavioral characteristics of the speaker, and is perceived over a duration of one or more utterances of speech. This paper attempts to provide a quantitative measure of the loudness of speech. The proposed measure is governed by the characteristics of the source of excitation of the speech signal, and is derived from an estimate of the source of excitation.

During production of speech, the identity of the speech sounds is governed mostly by the configuration of the vocal tract system. The size and shape of the vocal tract are dictated by the positions of the articulators. For the same configuration of the vocal tract system, loudness can be varied by varying the characteristics of glottal excitation. The characteristics of glottal excitation in speech are (a) impulse-like nature during glottal closure and (b) quasiperiodic nature of the impulse-like excitation in voiced sounds. In this paper, the impulse-like excitation is represented, both by the amplitude and the "strength" of the impulse-like excitation. The amplitude is estimated at the instant of the glottal closure, whereas the strength is based on the spread of the impulse-like excitation around the instant of the glottal closure. The notion of strength of excitation is explained as follows: Greater strength is associated with an excitation when a given amount of energy is concentrated in a short duration of time, than when the same energy is spread over a longer duration of time. For a given vocal tract system, an ideal impulse excitation can be said to have maximum strength, whereas white noise excitation of the same energy as the impulse, but spread over time, has the least strength. This paper proposes methods for estimating the amplitude and strength of excitation, and relates the perception of loudness of speech to the strength of excitation.

II. TRADITIONAL MEASURES OF LOUDNESS OF SPEECH

Measures of loudness of speech have been proposed based on the physiological characteristics of speech production and on the acoustic characteristics of speech signal. Perceptual judgments of loudness, indicated by labels such as "soft," "normal," and "loud," serve as a reference for comparison with the physiological and acoustic measures of loudness. Measurements of sound pressure level (Lane *et al.*, 1961; Ladefoged and McKinney, 1963; Allen, 1971; Orlikoff, 1991; Sulter and Wit, 1996; Holmberg *et al.*, 1988) and subglottal pressure level (Ladefoged and McKinney, 1963; Allen, 1971; Sundberg *et al.*, 2005) were observed to be strongly correlated with the perceptual judgments of loudness. Features derived from the measurements of glottal waveform have been studied for their effect on the perceived loudness of speech. Glottal volume velocity (Monsen and Engebretson, 1977), intraoral air pressure, oral airflow, sound pressure (Holmberg *et al.*, 1988), and electroglottographic (EGG) signals (Orlikoff, 1991) were obtained for male and female subjects for different levels of loudness such as soft, normal, and loud. In Holmberg *et al.* (1988), the subglottal air pressure and glottal airflow were derived from the measurements of the intraoral air pressure, oral airflow, and sound pressure. The maximum airflow declination rate (MFDR), defined as the maximum amplitude of the negative peak in the first derivative of the glottal volume velocity, was observed to be significantly lower in soft voices than in normal and loud voices (Holmberg *et al.*, 1988; Sundberg *et al.*, 2005). For each loudness condition, MFDR was observed to be highly correlated with the sound pressure level (Holmberg *et al.*, 1988). The MFDR parameter was found to increase linearly with subglottal pressure (Sundberg *et al.*, 2005). Soft voice had a more symmetrical waveform of the glottal volume velocity in general (in closing and opening phases in the glottal cycle), compared to loud voice (Monsen and Engebretson, 1977; Orlikoff, 1991). The closing portion of the glottal waveform was more abrupt for loud voices than that for normal and soft voices (Monsen and Engebretson, 1977). The less abrupt closure of the glottis in soft voices was also observed to be responsible for less energy in the high-frequency regions relative to the energy in the low-frequency regions (Holmberg *et al.*, 1988). The slope of the EGG signal in the closing phase in the glottal cycle was observed to be proportional to the amplitude of the acoustic speech signal (Orlikoff, 1991). These observations based on physical measurements of the glottal waveform provide motivation for deriving similar features from the glottal waveform estimated by inverse filtering the acoustic speech signal.

Measures of loudness of speech derived from the acoustic speech signal are based primarily on the characteristics of vibrations of the vocal folds. Glottal flow waveform estimated by inverse filtering the acoustic speech signal is parametrized to obtain such measures. Of these measures, the MFDR showed a significant increase from soft to loud levels (Sulter and Wit, 1996; Gauffin and Sundberg, 1989). Also, a strong correlation was observed between MFDR and spectral tilt (Gauffin and Sundberg, 1989). A strong correlation was also observed between the abruptness of the closing phase in the glottal cycle and the spectral tilt (Gauffin and Sundberg, 1989), where the former is related to the decrease in the rate of flow during the final part of the closing phase. Spectrum of the glottal source showed lesser roll-off for loud voice, compared to normal voice (Cairns and Hansen, 1994). Doval *et al.* (2006) observed that the maximum value of glottal excitation controlled the mid-to-high-frequency spectral slope in spectrum of the glottal flow waveform. Other parameters include open quotient (OQ) (i.e., proportion of the duration of the cycle for which the glottis is open), closed quotient (CQ) (i.e., proportion of the duration of the cycle for which the glottis is closed), closing quotient (ClQ) (i.e., proportion of the duration of the closing phase in each cycle), and speed quotient (SQ) (i.e., the time taken for the vocal folds to open, divided by the time taken for them to close). While the OQ decreased from soft to loud voices (Dromey

et al., 1992), the CQ was observed to increase from soft to loud voices (Sulter and Wit, 1996). The SQ showed an increase from soft to normal voices, and a decrease from normal to loud voices (Dromey *et al.*, 1992; Sulter and Wit, 1996). The ClQ was observed to be lowest for normal voice, and increased for both soft and loud voices (Sulter and Wit, 1996; Bäckström *et al.*, 2002). Cummings and Clements (1995) observed that the closing slope of the glottal waveform was significantly higher for loud voice compared to normal and soft voices. Also, the closing duration was significantly smaller for loud voice compared to normal and soft voices. By contrast, the opening slope and the opening duration did not show specific trends or significant differences among soft, normal, and loud voices. Bäckström *et al.* (2002) defined a parameter called amplitude quotient (AQ) as the ratio of the maximum amplitude of the glottal flow and the negative peak of the differentiated glottal flow. Normalized AQ, defined as the AQ normalized by the period of vibration, was observed to decrease with increase in vocal intensity (represented by sound pressure level). Alku *et al.* (2006) observed the variation in AQ as a function of MFDR, by varying the vocal intensity from "very soft" to "extremely loud." The AQ-MFDR curve showed a rapidly decreasing trend in the soft-normal range, followed by convergence toward a horizontal line for higher levels of loudness.

Configuration of the vocal tract may also undergo changes during the production of loud voices. Schulman (1989) observed that the patterns of movement of lips and jaws in loud speech (measured by displacement, velocity and relative timing associated with the movement) were amplified compared to normal speech. Spectral features derived from the speech signal, such as spectral tilt, changes in the formant frequencies and their bandwidths, and richness of the short-time spectrum as indicated by harmonicity of the spectrum, have been proposed as measures of loudness of speech. Ternström *et al.* (2006) defined a feature called "spectrum balance," as the energy in the high-frequency band (2–6 kHz) relative to that in the low-frequency band (0.1–1 kHz). This feature, when averaged across several segments of similar vowels, increased from soft to loud voice, but the rate of increase slowed down, or even stopped altogether, at very high levels of loudness. Very loud speech is mostly accompanied by a relative increase in the low-frequency energy, in the form of a sharper spectral peak at the first formant. Sundberg and Nordenberg (2006) defined alpha measure as the ratio of spectrum intensity above and below 1000 Hz, which was observed to increase linearly with sound energy level, corresponding to the increasing levels of loudness. Cairns and Hansen (1994) observed significant shifts in the formant frequencies and their bandwidths for loud voice, compared to normal voice. Gramming and Sundberg (1988) observed that the fundamental frequency was the strongest spectral component in soft voice, while it was typically a harmonic of the fundamental in loud voice. Moreover, the spectrum in loud voice was harmonically richer (as measured by the number of harmonics of the fundamental in a given frequency band, and their spectral intensities), compared to soft voice which had very few harmonics. For the vowel /a/, the first formant frequency was generally observed to be lower in the soft voice, compared to that in the loud voice.

Production of loudness in speech is also associated with vocal effort of the speaker. While the term vocal effort is not defined, Pickett (1956) described the range of vocal effort from "weakest voiced whisper" to "loudest possible shout." Allen (1971), and Glave and Rietveld (1975) observed that an increase in the vocal effort resulted in a corresponding increase in the perceived loudness. Glave and Rietveld (1975) observed that, between the sounds produced with effort and those produced without effort, a constant difference existed in the perceived loudness and also in the loudness calculated based on Zwicker's model. Traunmüller and Eriksson (2000) defined vocal effort in terms of the distance from the speaker as estimated by a group of listeners for a given utterance, in the context of communication over a range of distances. In general, increased vocal effort results in an increase in the energy level, the spectral emphasis (an acoustic feature reflecting the relative intensity in the higher frequency bands), the fundamental frequency, and the first formant (Traunmüller and Eriksson, 2000). Liénard and Benedetto (1999) observed that the fundamental frequency and the first formant were highly correlated with the vocal effort, while the second and third formants did not vary significantly. Also, the spectral emphasis and the amplitudes of the first three formants increased significantly with increase in the vocal effort.

The fundamental frequency of glottal vibration also reflects the variations in loudness. Studies by Harris and Weiss (1964), Lieberman *et al.* (1969), and Monsen and Engebretson (1977) have shown that, in general, there is an increase in the fundamental frequency for loud speech when compared to soft and normal speech. Holmberg *et al.* (1988) too observed that, in general, loud voice was produced with a higher fundamental frequency than that of the normal voice, whereas the fundamental frequency in a soft voice was either higher or lower compared to that in a normal voice. Alku *et al.* (2002) argued that speakers, while producing loud voice, increased the fundamental frequency to increase the number of glottal closures per unit time. This increased rapid fluctuations in the speech pressure waveform, thereby increasing the vocal intensity. Loud speech is also accompanied by an increase in the durations of vowels, diphthongs, and words (Cairns and Hansen, 1994; Traunmüller and Eriksson, 2000). However, not all increase in the fundamental frequency can be associated with an increase in loudness. For instance, speakers can keep the pitch steady and yet produce varying degrees of vocal loudness (Sundberg *et al.*, 2005). The patterns of vibrations of the vocal folds also reflect other features such as rhythm and rate of speech, and the accentuation of stressed syllables (Johnstone and Scherer, 1999; Ladd *et al.*, 1994). Thus, change in the fundamental frequency is an effect of the change in loudness rather than a cause of it.

Some of the acoustic features described above have also been used to characterize the labels of voice quality (Laver, 1994), such as creakiness, breathiness (Klatt and Klatt, 1990; Childers and Lee, 1991), falsetto (Childers and Lee, 1991),

hoarseness, and roughness (Eskenazi *et al.*, 1990). These labels of voice quality have been defined based on the configurations of the laryngeal system (Laver, 1994). Similarly, labels of voice quality such as tense and lax are described according to the degree of muscular tension in the laryngeal and supralaryngeal systems, while nasality is described based on the articulatory settings (Laver, 1994). By contrast, loudness cannot be characterized on a physiological basis alone. Moreover, a degree of perceived loudness can be associated with all the above labels of voice quality. For instance, tense voice sounds intrinsically louder than lax voice. Also, loudness is not an exclusive feature, in the sense that each voice quality can be realized with varying degrees of loudness. Thus, loudness can be viewed as an underlying feature that can be varied independently of the voice quality.

In summary, features of the glottal vibration play an important role in the production of vocal loudness. Two features of the glottal vibration are significant, namely, the amplitude of the negative peak of the differentiated glottal flow and the abruptness of the closing phase in the glottal cycle. These features are reflected in other acoustic measures such as sound energy level, spectral tilt, harmonic richness of the short-time spectrum, and, to an extent, in the sharpness of the formant peaks. However, the short duration of the impulse-like excitation in time is not captured well in the spectrum. Moreover, the estimation of the features of the glottal wave is dependent on the method of parametrization of the glottal flow waveform and the accuracy of the parameters. In view of this, an *acoustic* feature called *strength of excitation* is proposed in this paper, which can be derived from the inverse filtered signal. The motivation for deriving such a feature stems from the abruptness of the glottal closure, as illustrated in Sec. III. Computation of the proposed feature of strength of excitation is described in Sec. IV. Studies described in Sec. V show that the distribution of strength of excitation is related to the perception of loudness.

III. SIGNIFICANCE OF ABRUPTNESS OF GLOTTAL CLOSURE

A. Speech material

Speech utterances of a male speaker, spoken with varying levels of loudness such as soft, normal, and loud, were chosen from VOQUAL'03 database (d'Alessandro and Scherer, 2003). The following sentence was uttered by the speaker five times in each level of loudness: She has left for a great party today. The speech signals and the corresponding electroglottograph signals were sampled at 44.1 kHz, and both the signals were synchronized in time.

B. Measure of abruptness of glottal closure

Figure 1 shows segments of speech signals within one pitch period, and the differentiated EGG (DEGG) signals, corresponding to soft, normal, and loud utterances. The segments are shown for the vowel /a/ in the word "party." It is observed from the DEGG signals in Figs. 1(b), 1(d), and 1(f) that the abruptness of the glottal closure increases from soft to loud utterances. The abruptness of the glottal closure is reflected in the rate of decay of the DEGG signal from

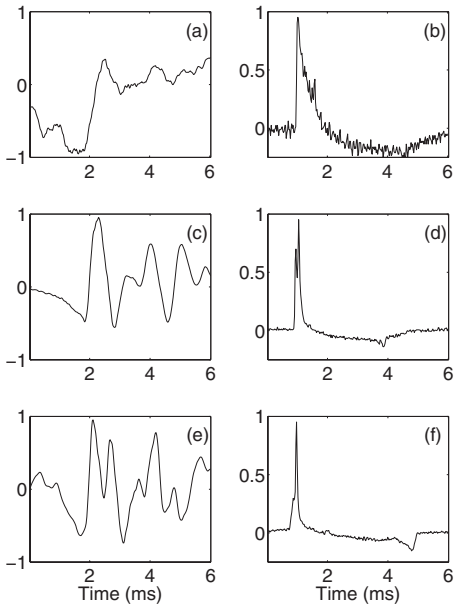


FIG. 1. Illustration of the abruptness of the glottal closure for soft, normal, and loud utterances. Speech segments within one pitch period are shown in (a), (c), and (e), which belong to soft, normal, and loud utterances, respectively. The segments correspond to the vowel /a/ in the word party in the sentence: She has left for a great party today. (b), (d), and (f) show the DEGG signals corresponding to (a), (c), and (e), respectively.

around the instant of the glottal closure. It is necessary to quantify the abruptness of the glottal closure to study its relationship with perceived loudness. First, the instants of glottal closure are estimated from the speech signal using a method described by Smits and Yegnanarayana (1995), which is based on the properties of minimum phase signals and group delay functions. A segment of 1 ms following the instant of glottal closure is considered from the DEGG signal. This segment is normalized by dividing the samples in the segment by the amplitude of the largest sample. The segment is approximated by a decaying exponential of the form $g(t)=e^{-t/\tau}$. Here the parameter τ denotes the time constant, and t denotes time. Let the samples of the segment of DEGG be denoted by $x[i]$, $i=0,1,2,\ldots,N-1$. Let t_i denote the time instant corresponding to the ith sample. It is assumed that $t_0=0$. Then, the parameter τ is estimated using the method of least squares as follows:

$$\tau^* = \arg\min_{\tau} \sum_{i=0}^{N-1} \|x[i] - e^{-t_i/\tau}\|^2. \quad (1)$$

The time constant τ^* indicates the abruptness of the glottal closure. An abrupt closure of the glottis corresponds to a faster decay of the exponential, resulting in a smaller value of the time constant τ^*. A relatively gradual closure of the glottis corresponds to a slower decay of the exponential, resulting in a larger value of the time constant τ^*.

C. Results

The values of the time constant τ^* are computed from the EGG signals of the soft, normal, and loud utterances. The distribution of the parameter τ^* is shown in Fig. 2 for the three levels of loudness. The distribution shows significant

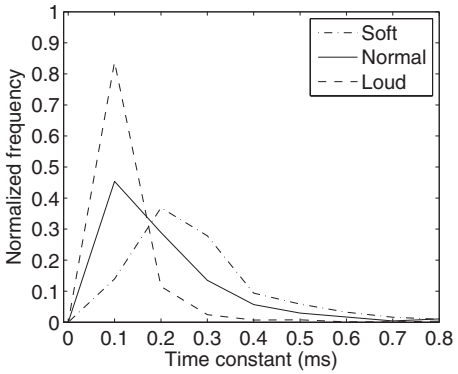


FIG. 2. Distribution of the time constant τ^* for soft, normal, and loud utterances.

variation between the soft and the loud utterances. The distribution of τ^* for the normal utterances overlaps considerably with those of both soft and the loud utterances, indicating that natural speech consists of segments of varying levels of loudness. In practice, only the speech signals are available, and not the EGG signals. Hence it is necessary to derive a measure of abruptness of glottal closure from the acoustic speech signal.

IV. MEASURES OF AMPLITUDE AND STRENGTH OF EXCITATION

To derive the features of amplitude and strength of excitation from the source of excitation of speech signal, a representation of the source of excitation is discussed in Sec. IV A. Methods for estimating the amplitude and the strength of excitation are described in Secs. IV B and IV C, respectively.

A. Representation of source of excitation

In order to characterize the impulse-like nature of excitation, an estimate of the source of excitation needs to be derived from the speech signal. Linear prediction (LP) residual can be used to approximate the source of excitation (Makhoul, 1975). LP residual is obtained by passing the speech signal through the inverse filter estimated during the LP analysis. Figures 3(a) and 3(b) show a segment of speech signal and its LP residual, respectively. LP analysis was performed on overlapped segments of speech signal (size of frame=25 ms, frame shift=5 ms, LP order=10, and sampling frequency=8 kHz). The prediction error in each glottal cycle is usually large around the instant where impulse-like excitation takes place. This happens around the instant of glottal closure for each glottal cycle due to abruptness of the closure. This is manifested as large amplitude fluctuations (both positive and negative) in the LP residual. The detection of these regions of large error in the LP residual is difficult because of the amplitude values with either polarity occurring around the instants of glottal closure. This difficulty can be overcome by using the Hilbert envelope of the LP residual (Ananthapadmanabha and Yegnanarayana, 1979). The Hilbert envelope $r[n]$ of the LP residual $e[n]$ is given by

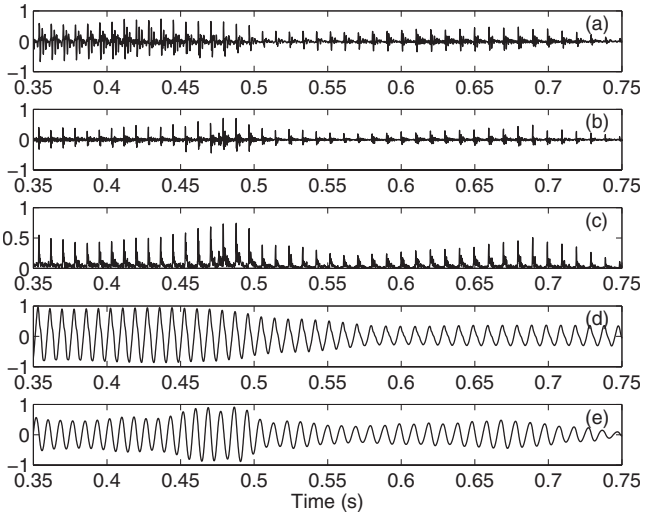


FIG. 3. (a) A segment of speech signal, (b) its LP residual, (c) Hilbert envelope of the LP residual, (d) filtered signal $y_s[n]$ derived from the speech signal, and (e) filtered signal $y_r[n]$ derived from the Hilbert envelope of the LP residual.

$$r[n] = \sqrt{e^2[n] + e_H^2[n]}, \tag{2}$$

where $e_H[n]$ denotes the Hilbert transform of $e[n]$. The Hilbert transform $e_H[n]$ of the signal $e[n]$ is given by

$$e_H[n] = \mathrm{IFT}(E_H(\omega)), \tag{3}$$

where IFT denotes the inverse Fourier transform, and $E_H(\omega)$ is given by (Oppenheim and Schafer, 1975)

$$E_H(\omega) = \begin{cases} +jE(\omega), & \omega \le 0 \\ -jE(\omega), & \omega > 0. \end{cases} \tag{4}$$

Here $E(\omega)$ denotes the Fourier transform of the signal $e[n]$. The impulse-like nature of excitation can be observed clearly from the Hilbert envelope of the LP residual, as shown in Fig. 3(c). The amplitude of the excitation can be estimated by detecting the instants of the glottal closure, and then measuring the peaks in the Hilbert envelope of LP residual around the instants. Another approach for estimating the amplitude of the excitation is proposed in Sec. IV B.

B. Estimation of amplitude of impulse-like excitation

1. Computation of filtered signal

The impulse-like excitation is due to abruptness of the glottal closure in each cycle. The characteristics of the sequence of impulse-like excitations are reflected across all the frequencies in the speech signal including 0 Hz (hereafter referred to as zero frequency). Filtering the speech signal through a resonator located at zero frequency helps in emphasizing the characteristics of excitation (Murty and Yegnanarayana, 2008). The system function of such a resonator is given by

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{5}$$

where $a_1 = -2$ and $a_2 = 1$. The above resonator de-emphasizes the characteristics of the vocal tract, since the resonances of the latter are located at much higher frequencies than the

zero frequency. A cascade of two such resonators, given by the system function $G(z)=H(z)H(z)$, is used to reduce the effect of all the resonances of the vocal tract. Let $s[n]$ denote the input speech signal. Then the output $x_s[n]$ of the cascade of resonators is given by

$$x_s[n] = s[n] \star g[n], \tag{6}$$

where $g[n]$ is the impulse response of the system function $G(z)$ and $\star$ denotes convolution operation. The output $x_s[n]$ contains mainly the features of glottal vibrations. Filtering the signal $s[n]$ through the cascade of resonators causes the output $x_s[n]$ to grow as a polynomial function of time. This trend in $x_s[n]$ is removed by subtracting the average of sample values over a window of 10 ms (approximately 0.5–1.5 times the estimated pitch period). The resulting trend-removed signal $y_s[n]$ is given by (Murty and Yegnanarayana, 2008)

$$y_s[n] = x_s[n] - \frac{1}{2N+1} \sum_{k=-N}^{N} x_s[n+k], \tag{7}$$

where $2N+1$ is the size (in samples) of the window. The signal $y_s[n]$ is called the *filtered signal*, an example of which is shown in Fig. 3(d) for the segment of voiced speech in Fig. 3(a).

2. Slope of positive-to-negative zero crossings

Murty and Yegnanarayana (2008) observed that the locations of positive-to-negative zero crossings (PNZCs) of the filtered signal $y_s[n]$ provide an accurate estimate of the instants of glottal closure. It is observed that the filtered signal $y_s[n]$ is free of the characteristics of the vocal tract system. The filtered signal can also be derived from the Hilbert envelope $r[n]$ of the LP residual, instead of the speech signal $s[n]$. Let $y_r[n]$ denote the filtered signal derived from $r[n]$. Figure 3(e) shows $y_r[n]$ for the segment of voiced speech in Fig. 3(a). We note from Figs. 3(d) and 3(e) that the locations of the PNZCs derived from $y_r[n]$ are nearly the same as those derived from $y_s[n]$.

A strong peak in the Hilbert envelope $r[n]$ has a corresponding PNZC in the filtered signal $y_r[n]$. It is observed from Fig. 3(e) that the slope of $y_r[n]$ at a PNZC is proportional to the amplitude of the corresponding peak in the Hilbert envelope $r[n]$. The slope of $y_r[n]$ at a PNZC is estimated by considering a region of 0.125 ms on either side of the PNZC, by assuming $y_r[n]$ to be linear in the vicinity of each PNZC. To observe the relationship between the slope of $y_r[n]$ at a PNZC and the amplitude of the corresponding peak in $r[n]$, speech signals collected from 50 female and 50 male speakers of TIMIT database (Garofalo *et al.*, 1993) were processed. For each speaker, ten spoken utterances were used, whose durations ranged from 2 to 5 s. Only voiced segments were processed. The scatter plots in Fig. 4 illustrate the linear dependence of the amplitude of the peak of $r[n]$ and the slope of the corresponding PNZC in $y_r[n]$. Both these quantities, which form an ordered pair in the scatter plots, are associated with an instant of glottal closure. Thus the number of points in the scatter plot shown in Fig. 4(a) [Fig. 4(b)] denotes the number of glottal closures in 500 utterances [50 female (male) speakers $\times$ 10 utterances per speaker]. For instance, if we assume that each of the 500 analyzed utterances contains 1 s of voiced speech, with an average pitch period of 4 ms (8 ms) for the female (male) speakers, then the number of glottal closures amounts to 125 000 (62 500). The actual number of points in the scatter plot shown in Fig. 4(a) is 156 264, while that in the scatter plot shown in Fig. 4(b) is 69 359. The ordinate in Fig. 4 shows only the magnitude of the slope of $y_r[n]$ at a PNZC, although the slope itself is negative. The values of the correlation coefficient computed from the sets of points in Figs. 4(a) and 4(b) are 0.92 and 0.94, respectively. The values of the correlation coefficient computed for different speakers ranged from 0.90 to 0.98. Note the approximate *linear* relation between the amplitude of excitation and the slope at PNZC, even though the gross slopes of the lines are different for female and male speakers due to differences in their average pitch periods.

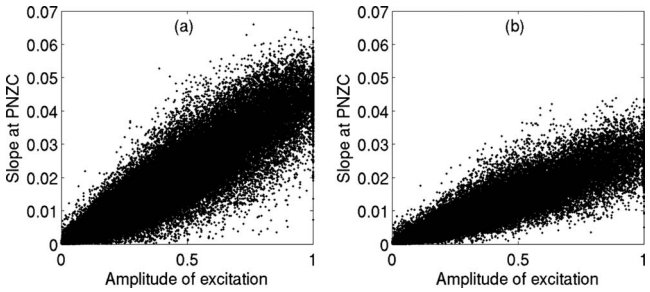


FIG. 4. Scatter plots to illustrate the linear dependence of the amplitude of excitation and the slope of PNZC in the filtered signal $y_r[n]$ for (a) 50 female speakers and (b) 50 male speakers.

C. A measure of strength of excitation

Figure 5 shows segments of voiced speech, chosen from the utterances of soft, normal, and loud voices. The impulse-like excitation, as observed from the LP residuals of the

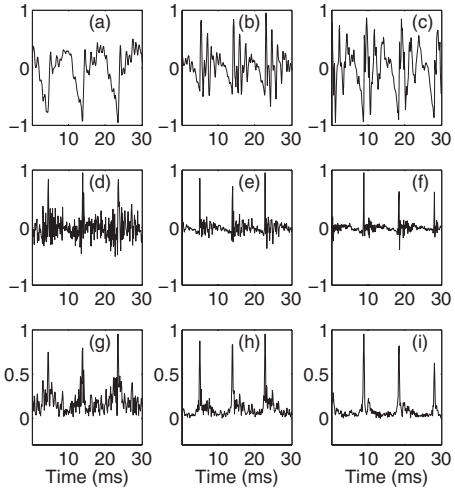


FIG. 5. Illustration of the nature of excitation in soft, normal, and loud utterances. Speech segments in (a)–(c) belong to soft, normal, and loud utterances, respectively. The segments correspond to the vowel /a/ in the word party in the sentence: She has left for a great party today. (d)–(f) show the LP residual for the signals in (a)–(c), respectively, while the figures (g)–(i) show the Hilbert envelopes of the corresponding LP residuals.

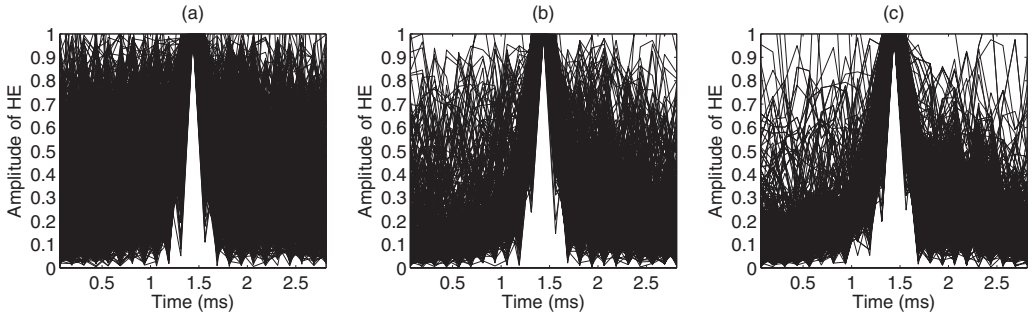


FIG. 6. Segments of Hilbert envelope of LP residual in the vicinity of impulse-like excitations for (a) soft, (b) normal, and (c) loud utterances.

speech segments, is more spread out in time for the soft utterances [Fig. 5(d)], compared to normal [Fig. 5(e)] and loud [Fig. 5(f)] utterances. The impulse-like nature of the glottal excitation can be observed clearly from the Hilbert envelope of the LP residual, shown in Figs. 5(g)–5(i) for soft, normal, and loud utterances, respectively. A measure of the strength of excitation can be derived from a short segment (1–3 ms) around the instants of impulse-like excitation. Figure 6(a) shows superimposed segments of the Hilbert envelopes around these instants derived from soft utterances. Each segment has a duration of 3 ms, and the location of the peak in the Hilbert envelope of the LP residual is at the center of the segment. Each segment is normalized by dividing the samples of the segment by the largest amplitude in the segment. All the segments are derived from the voiced regions of five soft utterances of a male speaker in the VOQUAL'03 database and are superimposed, as shown in Fig. 6(a). Similar plots are obtained for normal and loud utterances, as shown in Figs. 6(b) and 6(c), respectively. These plots show that for soft utterances, the Hilbert envelope of the LP residual is spread out more uniformly on either side of the instant of impulse-like excitation. This indicates that the impulses around the instants of impulse-like excitation are not sharp in these cases. The impulses are much sharper for loud utterances than for soft or even normal utterances.

To represent the sharpness in the Hilbert envelope of the LP residual, a feature called strength of excitation is defined as $\eta=\sigma/\mu$, where μ denotes the mean of the samples of the Hilbert envelope of the LP residual in a segment around the instant of impulse-like excitation, and σ denotes the standard deviation of the samples. For a segment of length N consisting of an ideal impulse (in discrete-time domain) of amplitude a, $\eta=\sqrt{N}$. For a segment of length N consisting of samples of equal amplitude $a/\sqrt{N}$, $\eta=0$. The segment in this case has the same energy as that of the ideal impulse of amplitude a. This case represents the maximum deviation from an ideal impulse. Thus the value of η lies between 0 and $\sqrt{N}$ for any segment, irrespective of the amplitudes of the samples in the segment. A higher value of η indicates greater strength of excitation. In general, a segment having impulse-like characteristics in excitation, as in the case of a loud voice, has a smaller value of μ and a larger value of σ, resulting in a larger value of η. By contrast, a soft voice with greater spread around the center has a larger value of μ and a smaller value of σ, resulting in a smaller value of η.

The strength η of excitation is computed for soft, normal, and loud utterances of the male speaker in the VOQUAL'03 database. Figure 7 shows the distribution of η for the three types of utterances. The plot indicates that for soft voice, the distribution has greater concentration of lower values of η, whereas for loud voice, the distribution is concentrated around larger values of η. Normal and loud utterances can also be distinguished by comparison, since the proportion of larger values of η is higher in loud utterances compared to that in the normal utterances. Also, the discrimination between soft and loud utterances in the distribution of η is comparable to that based on the distribution of the time constant parameter derived from DEGG signals (Fig. 2). Therefore, the distribution of the strength of excitation can be used as a measure of the perceived loudness of a given speech signal. The distribution of η can be used to identify soft and loud segments in the speech of a given speaker. The distribution can also help in inferring some gross speaker-specific characteristics, as discussed in the next section.

V. EVALUATION OF THE EFFECTIVENESS OF STRENGTH OF EXCITATION

In this section, the ability of the distribution of the strength (η) of excitation to distinguish between the levels of loudness within individual speakers is examined. The distribution is also used for comparing the loudness of speech from two different speakers. It is examined whether the distribution of η is in agreement with the subjective judgment of loudness of speech.

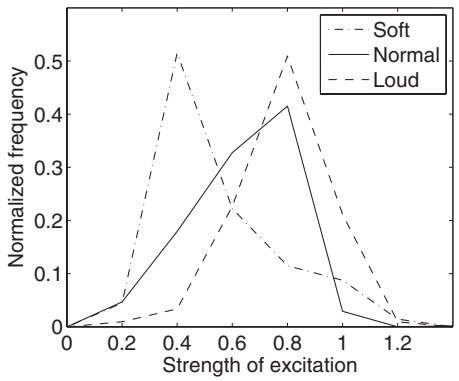


FIG. 7. Distribution of the strength (η) of excitation for soft, normal, and loud utterances.

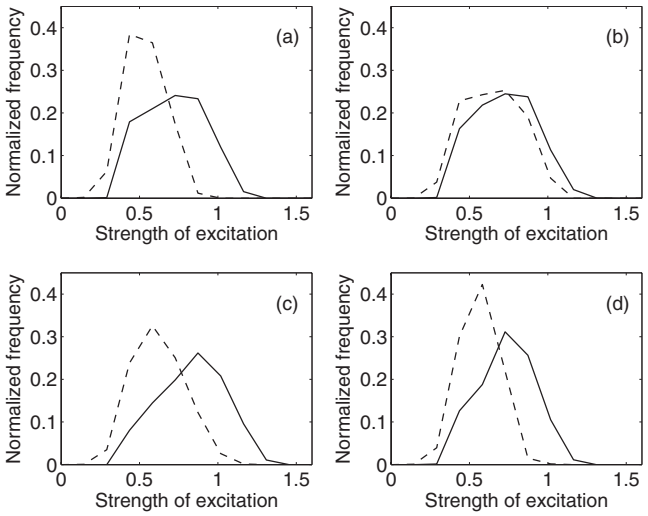


FIG. 8. Distribution of the strength (η) of excitation for four speakers. In each case, the broken and the solid lines correspond to soft and loud utterances, respectively. (a) and (b) correspond to two female speakers, while (c) and (d) correspond to two male speakers.

A. Speaker-specific nature of loudness

1. Speech material

Speech signals were collected from 44 speakers (13 female speakers and 31 male speakers) in two levels of loudness, namely, soft and loud. The speakers were undergraduate and graduate students, aged between 17 and 26 years. All the speakers spoke Indian English, and the native language of each speaker was one among Telugu, Hindi, Kannada, Tamil, Marathi, and Oriya. The speakers were guided to listen to the soft and the loud utterances of the VOQUAL'03 database, so as to help them produce the two levels of loudness while maintaining naturalness of their speech. Each speaker uttered 20 sentences in soft level and in loud level. The durations of these sentences ranged from 1 to 5 s. The speech signals were sampled at 8 kHz.

2. Results

The collected speech signals were analyzed, and the distribution of the strength (η) of excitation was computed for the soft and loud utterances of each speaker. Figure 8 shows the distribution of η for two female and two male speakers, chosen at random from among the 44 speakers for illustration. It is observed that the distribution of η does discriminate between the soft and loud utterances of the speakers. The degree of discrimination, or separation between the distributions of η for soft and loud utterances, is speaker dependent. For instance, the separation between the distributions is less in Fig. 8(b), compared to that in Figs. 8(a), 8(c), and 8(d). Some speaker-specific characteristics can be inferred from these figures. For instance, the speaker in Fig. 8(b) is not able to produce utterances which are significantly louder than the soft utterances. The distribution of η may also indicate the range of loudness that can be produced by a speaker. The plots in Fig. 8, derived from the speech of four speakers, are indicative of the general trend. Any other set of four speakers is equally suitable for illustration.

The distribution of η for a given loudness level of a speaker can be approximated by a Gaussian probability density function. A measure of distance between two distributions is the Kullback–Leibler (KL) divergence (Kullback, 1968). When both the distributions are described by univariate Gaussian probability density functions, the KL divergence is given by (Cover and Thomas, 1991)

$$d_{KL}(\mathcal{A},\mathcal{B}) = \frac{1}{2}\left\{\frac{\sigma_{\mathcal{A}}^2}{\sigma_{\mathcal{B}}^2} + \frac{\sigma_{\mathcal{B}}^2}{\sigma_{\mathcal{A}}^2}\right\} - 1 + \frac{1}{2}\{\mu_{\mathcal{A}} - \mu_{\mathcal{B}}\}^2\left\{\frac{1}{\sigma_{\mathcal{A}}^2} + \frac{1}{\sigma_{\mathcal{B}}^2}\right\}, \tag{8}$$

where $\mu_{\mathcal{A}}$ and $\sigma_{\mathcal{A}}$ denote the mean and the standard deviation, respectively, of the samples in set $\mathcal{A}$, while $\mu_{\mathcal{B}}$ and $\sigma_{\mathcal{B}}$ denote the corresponding quantities for the samples in set $\mathcal{B}$. Also computed is $|\mu_{\mathcal{A}}-\mu_{\mathcal{B}}|$, which is the absolute value of the difference of the mean values $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$. In this study, the samples in sets $\mathcal{A}$ and $\mathcal{B}$ are the values of the strength (η) of excitation. Let us consider the following two cases: (a) When the values of η in both $\mathcal{A}$ and $\mathcal{B}$ are derived from the soft utterances of a speaker, $d_{KL}(\mathcal{A},\mathcal{B})$ and $|\mu_{\mathcal{A}}-\mu_{\mathcal{B}}|$ are small. (b) A similar behavior is expected when the values of η in both $\mathcal{A}$ and $\mathcal{B}$ are derived from the loud utterances of the speaker. If we denote soft and loud as two classes of loudness, then the above two cases represent intra-class comparisons. By contrast, inter-class comparisons are those where the values of η in $\mathcal{A}$ and $\mathcal{B}$ are derived from the soft (loud) and the loud (soft) utterances, respectively, of a speaker. Both $d_{KL}(\mathcal{A},\mathcal{B})$ and $|\mu_{\mathcal{A}}-\mu_{\mathcal{B}}|$ are expected to be larger in the case of inter-class comparisons than in the case of intra-class comparisons. The ordered pair $[|\mu_{\mathcal{A}}-\mu_{\mathcal{B}}|, d_{KL}(\mathcal{A},\mathcal{B})]$ is used to distinguish between soft and loud utterances of a speaker, as described below.

Let $\mathcal{S}$ denote the set of values of η of a given speaker, derived from the 20 utterances collected in soft voice. Let $\mathcal{S}_1$, $\mathcal{S}_2$, and $\mathcal{S}_3$ denote three distinct subsets of $\mathcal{S}$, such that the values of η in each subset are derived from six utterances in the soft voice. For the same speaker, let $\mathcal{L}$, $\mathcal{L}_1$, $\mathcal{L}_2$, and $\mathcal{L}_3$ denote the corresponding sets derived from the loud utterances. For each speaker, the following ordered pairs are computed: (a) $(|\mu_{\mathcal{S}i}-\mu_{\mathcal{L}j}|, d_{KL}(\mathcal{S}_i,\mathcal{L}_j))$, for $i=1,2,3$, and $j=1,2,3$; (b) $(|\mu_{\mathcal{S}}-\mu_{\mathcal{L}}|, d_{KL}(\mathcal{S},\mathcal{L}))$; (c) $(|\mu_{\mathcal{S}i}-\mu_{\mathcal{S}j}|, d_{KL}(\mathcal{S}_i,\mathcal{S}_j))$ for $i=1,2,3$, $j=1,2,3$, and $i\neq j$; and (d) $(|\mu_{\mathcal{L}i}-\mu_{\mathcal{L}j}|, d_{KL}(\mathcal{L}_i,\mathcal{L}_j))$, for $i=1,2,3$, $j=1,2,3$, and $i\neq j$. It was observed that $\mu_{\mathcal{L}}>\mu_{\mathcal{S}}$, and $\mu_{\mathcal{L}i}>\mu_{\mathcal{S}j}$, for $i=1,2,3$, and $j=1,2,3$, for all the speakers. The ordered pairs in (a) and (b) denote inter-class comparisons within a speaker, while those in (c) and (d) denote intra-class comparisons. Each ordered pair can be plotted as a point in a two-dimensional plane. For each speaker, there are ten points due to inter-class comparisons and six points due to intra-class comparisons [since $d_{KL}(\mathcal{A},\mathcal{B})=d_{KL}(\mathcal{B},\mathcal{A})$]. Figures 9(a) and 9(c) show the intra-class points for the recorded 13 female and 31 male speakers respectively, while Figs. 9(b) and 9(d) show the inter-class points. For both female and male speakers, the intra-class points are clustered closer to the origin compared to the inter-class points which are farther from the origin and have a greater spread. Thus, the distribution of strength of excitation does help in distinguish-

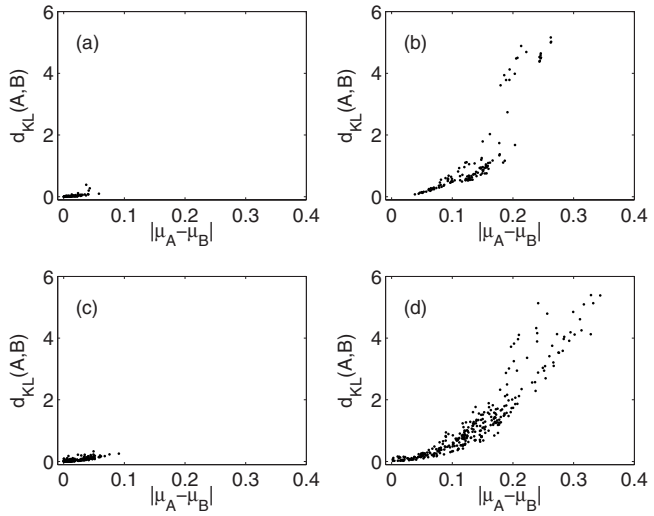


FIG. 9. Illustration of the variation of speaker-specific loudness. (a) and (b) show the results of intra-class comparisons and inter-class comparisons, respectively, for 13 female speakers. (c) and (d) show the results of intra-class comparisons and inter-class comparisons, respectively, for 31 male speakers.

ing between the loudness levels of a speaker. The distribution of η may also be useful in identifying those segments of speech signal of a speaker which are relatively soft. Such segments can then be processed, if necessary, to improve the loudness of the signal.

B. Comparison of loudness across speakers

Given two speakers, human listeners can judge in most cases, if the speech of one speaker is louder relative to the speech of the other speaker. This is particularly so if both the speakers belong to the same gender. This section describes an experiment to compare the subjective judgment of loudness and an objective measure of loudness.

The effect of prosodic factors such as changes in duration and pitch, on the perception of loudness in spontaneous speech, can be significant. For instance, speakers tend to raise their pitch while producing loud speech. Hence, when the utterances used for subjective listening consist of read speech corresponding to the same sentence, the prosodic variations across different speakers may be comparable. The subjective judgment of loudness is likely to be influenced more by the excitation characteristics of the voices than by the prosodic factors.

1. Speech material and subjective listening

Twenty-five subjects participated in the listening test to judge the relative loudness of several pairs of utterances. These 25 subjects belonged to the set of 44 subjects who participated in the experiment described in Sec. V A 1. For listening, speech signals corresponding to the utterances of the same text were selected from 20 speakers (6 female and 14 male speakers) of TIMIT database (Garofalo *et al.*, 1993). The average duration of the utterances is about 3 s. The signals were sampled at 16 kHz. The data were organized into pairs of utterances where each pair belonged to one of the following three types: (a) *XY*, when the pair consists of two different speakers; (b) *YX*, when the order of the speakers in the pair is reversed; and (c) *XX*, when the pair is a repetition of an utterance of the same speaker. For listening tests, 40 pairs of utterances were used, with 30 pairs of type *XY*, 5 pairs of type *YX*, and 5 pairs of type *XX*. The pairs of type *YX* and *XX* were used to check the consistency in the judgment of the subjects. Of the 30 pairs of type *XY*, 10 pairs consist of female speakers, and 20 pairs consist of male speakers. The utterances in each pair were normalized so that the energy of the signal was same in both the utterances.

The 40 pairs of utterances were presented in a random order. The subjects did not know the identity of the speakers. The subjects were asked to mark *A* or *B*, depending on whether the first or the second utterance in the pair was judged louder. They were asked to mark *C* if they observed no perceptible difference in loudness between the two utterances in a pair. For all the five pairs of type *XX*, all the subjects marked *C*. If a subject's decision on louder voice in the pairs of type *XY* and *YX* was not consistent, then the subject was regarded as inconsistent. If a subject was found inconsistent in two or more of the five such cases, then the decisions made by that subject were ignored. Out of the 25 subjects, four subjects were found to be inconsistent. Hence the decisions by the remaining 21 subjects were considered for evaluation. Out of these 21 subjects, the decision by the majority of the subjects on a pair of utterances of type *XY* was taken as the correct one. Here, the term majority denotes that at least 11 subjects out of the 21 subjects have voted in favor of one particular speaker, in a pair of utterances of type *XY*.

The subjective tests gave a clear decision on louder voice consistently, only for 21 out of the 30 pairs of the type *XY*. For the remaining nine pairs, there was no clear decision on the louder voice. This observation correlated with the objective measure described in Sec. V B 2.

2. Objective measurement of loudness

The loudness of two speakers in a pair is compared using the distributions of the strength (η) of excitation. For each speaker, ten utterances (including the utterance used in the listening test) were used to derive the distribution of η. The durations of the utterances varied from 2 to 4 s. All speech signals were downsampled to 8 kHz for processing. Let $\mathcal{A}$ and $\mathcal{B}$ denote the sets of values of η for a pair of speakers. The ordered pair $(|\mu_{\mathcal{A}}-\mu_{\mathcal{B}}|, d_{KL}(\mathcal{A},\mathcal{B}))$ is computed, as described in Sec. V A 2. There are 30 such ordered pairs corresponding to the 30 pairs of utterances of type *XY*, and these ordered pairs are plotted as points in a two-dimensional plane, as shown in Fig. 10. In Fig. 10(a), the points marked by "○" correspond to the nine pairs of speakers, for whom a clear decision could not be made by the listeners in the subjective test. Eight of these nine points lie close to the origin, indicating lack of discrimination between the distributions of η for the speakers in these pairs. Note that the subjective tests were conducted using only one utterance per speaker, whereas the distribution of η is obtained using ten utterances per speaker. The points marked by "+" in Fig. 10(a) denote the 21 pairs of speakers where one speaker was rated as louder in the subjective test. Here, the

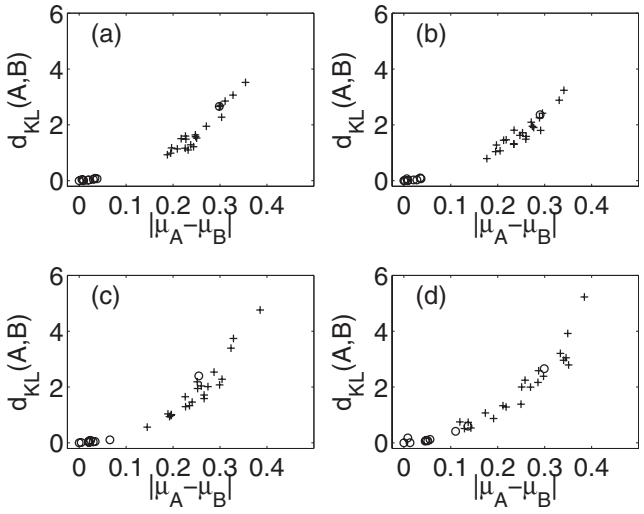


FIG. 10. In each plot, the points marked by + denote those pairs for which one speaker in each pair was judged as louder, based on the subjective listening. The points marked by o denote those pairs for which neither speaker in a pair was decisively voted as louder. For each speaker, the distribution of strength of excitation was obtained by processing (a) ten utterances, (b) six utterances, (c) three utterances, and (d) one utterance.

points are located farther away from the origin compared to the points marked by ○. Thus, discrimination between the loudness of two speakers based on the distributions of the strength of excitation is in agreement with the subjective judgment of loudness.

Figures 10(a)–10(d) correspond to the cases where the distribution of η for each speaker is obtained using ten, six, three, and one utterances, respectively. It is evident that the discrimination between the clusters of + and ○ points is reduced when the amount of speech data is reduced. Thus, 10–15 s of speech material may be required per speaker to obtain a reliable estimate of the distribution of the strength of excitation. The reliability of subjective judgment of loudness may also improve with increase in the duration of speech material used for listening.

VI. CONCLUSION

This paper presents a measure of perceived loudness in the form of distribution of a feature called strength of excitation. The strength of excitation represents the impulse-like nature of excitation in speech production. Observation of the electroglottograph signals indicates that the abruptness of glottal closure during the production of voiced speech plays an important role in the perception of loudness. The abruptness of the glottal closure lends the impulse-like characteristic to the excitation. Two features of the impulse-like excitation are investigated, namely, the amplitude and the strength of excitation. These features are derived from the Hilbert envelope of the LP residual of speech signal. The method proposed for estimating the amplitude of excitation is based on filtering the Hilbert envelope of LP residual through a zero frequency resonator. The strength of excitation is derived from a short segment of the Hilbert envelope of the LP residual of speech signal, around the instant of impulse-like excitation. The feature of the strength of excitation is that it is independent of the period of glottal vibration, and does not require parametrization of the glottal flow derivative. Experiments show that the distribution of the strength of excitation is strongly related to perceived loudness. The ability of the distribution of the strength of excitation to distinguish between soft and loud utterances of individual speakers is demonstrated using speech signals collected from a set of 44 speakers. Also, discrimination between the loudness of two speakers obtained based on the subjective judgment is in agreement with the discrimination between the distributions of the strength of excitation of the two speakers. This is illustrated on a set of 30 pairs of utterances, spoken by 20 speakers. Thus the distribution of the strength of excitation is useful for comparison of loudness of speakers. The significance of the amount of speech material required for reliable estimation of the distribution is also discussed. Since loudness varies over different segments of speech signal, it is more appropriately described by the *distribution* of the strength of excitation, than by the *strength* itself.

The proposed feature highlights the significance of the nature of excitation in the perception of loudness. The feature of the strength of excitation can help in measuring the loudness level of a speaker's voice on a quantitative basis. The proposed feature can be used to automatically identify the segments or regions of speech signal with relatively less loudness. Such segments may then be processed to enhance their loudness, if necessary. The reliability of the distribution of the strength of excitation improves with the amount of speech data. This is mainly due to variation in loudness over different segments of speech signal. Perceived loudness could be different for different sounds, due to loading of the vocal tract system on the vocal source during the production of speech sounds. The perception of loudness in human listeners is also influenced by prosodic factors, such as variations in pitch and duration, which are manifested over longer durations of speech. Hence, the assessment of loudness by human listeners is likely to improve with the duration of speech data. Prosodic factors, such as stress on particular syllables of words and accentuation of stressed syllables, also affect the perception of loudness. The influence of such factors on the perception of loudness needs to be studied and quantified, in order to obtain a more comprehensive measure of loudness of speech.

Alku, P., Airas, M., Björkner, E., and Sundberg, J. (**2006**). "An amplitude quotient based method to analyze changes in the shape of the glottal pulse in the regulation of vocal intensity," J. Acoust. Soc. Am. **120**, 1052–1062.
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.
Allen, G. D. (**1971**). "Acoustic level and vocal effort as cues for the loudness of speech," J. Acoust. Soc. Am. **49**, 1831–1841.
Ananthapadmanabha, T. V., and Yegnanarayana, B. (**1979**). "Epoch extraction from linear prediction residual for identification of closed glottis interval," IEEE Trans. Acoust., Speech, Signal Process. **27**, 309–319.
Bäckström, T., Alku, P., and Vilkman, E. (**2002**). "Time-domain parameterization of the closing phase of glottal airflow waveform from voices over a large intensity range," IEEE Trans. Speech Audio Process. **10**, 186–192.
Beranek, L. L., Marshall, J. L., Cudworth, A. L., and Peterson, A. P. G. (**1951**). "Calculation and measurement of the loudness of sounds," J. Acoust. Soc. Am. **23**, 261–269.
Cairns, D. A., and Hansen, J. H. L. (**1994**). "Nonlinear analysis and classification of speech under stressed conditions," J. Acoust. Soc. Am. **96**, 3392–3400.

Childers, D. G., and Lee, C. K. (**1991**). "Voice quality factors: Analysis, synthesis, and perception," J. Acoust. Soc. Am. **90**, 2394–2410.
Cover, T. M., and Thomas, J. A. (**1991**). *Elements of Information Theory* (Wiley, New York).
Cummings, K. E., and Clements, M. A. (**1995**). "Analysis of the glottal excitation of emotionally styled and stressed speech," J. Acoust. Soc. Am. **98**, 88–98.
d'Alessandro, C., and Scherer, K. R. (**2003**). "Voice quality: Functions, analysis and synthesis (VOQUAL'03)," ISCA Tutorial and Research Workshop, Geneva, Switzerland, http://archives.limsi.fr/VOQUAL/voicematerial.html (Last viewed 6/7/2008).
Doval, B., d'Alessadro, C., and Henrich, N. (**2006**). "The spectrum of glottal flow models," Acta. Acust. Acust. **92**, 1026–1046.
Dromey, C., Stathopoulos, E. T., and Sapienza, C. M. (**1992**). "Glottal airflow and electroglottographic measures of vocal function at multiple intensities," J. Voice **6**, 44–54.
Eskenazi, L., Childers, D. G., and Hicks, D. M. (**1990**). "Acoustic correlates of vocal quality," J. Speech Hear. Res. **33**, 298–306.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague, The Netherlands).
Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.
Fletcher, H., and Munson, W. A. (**1937**). "Relation between loudness and masking," J. Acoust. Soc. Am. **9**, 1–10.
Garofalo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1993**). "The DARPA TIMIT acoustic-phonetic continuous speech corpus cdrom," Linguistic Data Consortium, Philadelphia, PA.
Gauffin, J., and Sundberg, J. (**1989**). "Spectral correlates of glottal voice source waveform characteristics," J. Speech Hear. Res. **32**, 556–565.
Glave, R. D., and Rietveld, A. C. M. (**1975**). "Is the effort dependence of speech loudness explicable on the basis of acoustical cues?," J. Acoust. Soc. Am. **58**, 875–879.
Gramming, P., and Sundberg, J. (**1988**). "Spectrum factors relevant to phonetogram measurement," J. Acoust. Soc. Am. **83**, 2352–2360.
Harris, C. M., and Weiss, M. R. (**1964**). "Effects of speaking condition on pitch," J. Acoust. Soc. Am. **36**, 933–936.
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.
Johnstone, T., and Scherer, K. R. (**1999**). "The effects of emotions on voice quality," in Proceedings of the 14th International Conference on Phonetic Sciences, San Francisco, pp. 2029–2032.
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.
Kullback, S. (**1968**). *Information Theory and Statistics* (Dover, Mineola, NY).
Ladd, D. R., Verhoeven, J., and Jacobs, K. (**1994**). "Influence of adjacent pitch accents on each other's perceived prominence: Two contradictory effects," J. Phonetics **22**, 87–99.
Ladefoged, P., and McKinney, N. P. (**1963**). "Loudness, sound pressure, and subglottal pressure in speech," J. Acoust. Soc. Am. **35**, 454–460.
Lane, H. L., Catania, A. C., and Stevens, S. S. (**1961**). "Voice level: Autophonic scale, perceived loudness, and effects of sidetone," J. Acoust. Soc. Am. **33**, 160–167.
Laver, J. (**1994**). *Principles of Phonetics*, Cambridge Textbooks in Linguistics (Cambridge University Press, Cambridge).
Lieberman, P., Knudson, R., and Mead, J. (**1969**). "Determination of the rate of change of fundamental frequency with respect to subglottal air pressure during sustained phonation," J. Acoust. Soc. Am. **45**, 1537–1543.
Liénard, J.-S., and Benedetto, M.-G. D. (**1999**). "Effect of vocal effort on spectral properties of vowels," J. Acoust. Soc. Am. **106**, 411–422.
Makhoul, J. (**1975**). "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580.
Monsen, R. B., and Engebretson, A. M. (**1977**). "Study of variations in the male and female glottal wave," J. Acoust. Soc. Am. **62**, 981–993.
Murty, K. S. R., and Yegnanarayana, B. (**2008**). "Epoch extraction from speech signals," IEEE Trans. Audio, Speech, Lang. Process. **16**, 1602–1613.
Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Orlikoff, R. F. (**1991**). "Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects," J. Speech Hear. Res. **34**, 1066–1072.
Pickett, J. M. (**1956**). "Effects of vocal force on the intelligibility of speech sounds," J. Acoust. Soc. Am. **28**, 902–905.
Rothenberg, M. (**1983**). "The effect of flow dependence on source-tract acoustic interaction," J. Acoust. Soc. Am. **73**, S72–S72.
Schulman, R. (**1989**). "Articulatory dynamics of loud and normal speech," J. Acoust. Soc. Am. **85**, 295–312.
Smits, R., and Yegnanarayana, B. (**1995**). "Determination of instants of significant excitation in speech using group delay functions," IEEE Trans. Speech Audio Process. **3**, 325–333.
Stevens, S. S. (**1956**). "Calculation of the loudness of complex noise," J. Acoust. Soc. Am. **28**, 807–832.
Stevens, S. S. (**1961**). "Procedure for calculating loudness: Mark VI," J. Acoust. Soc. Am. **33**, 1577–1585.
Sulter, A. M., and Wit, H. P. (**1996**). "Glottal volume velocity waveform characteristics in subjects with and without vocal training, related to gender, sound intensity, fundamental frequency, and age," J. Acoust. Soc. Am. **100**, 3360–3373.
Sundberg, J., and Nordenberg, M. (**2006**). "Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech," J. Acoust. Soc. Am. **120**, 453–457.
Sundberg, J., Fahlstedt, E., and Morell, A. (**2005**). "Effects on the glottal voice source of vocal loudness variation in untrained female and male voices," J. Acoust. Soc. Am. **117**, 879–885.
Ternström, S., Bohman, M., and Södersten, M. (**2006**). "Loud speech over noise: Some spectral attributes, with gender differences," J. Acoust. Soc. Am. **119**, 1648–1665.
Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.
Warren, R. M. (**1973**). "Anomalous loudness function for speech," J. Acoust. Soc. Am. **54**, 390–396.
Zwicker, E. (**1977**). "Procedure for calculating loudness of temporally variable sounds," J. Acoust. Soc. Am. **62**, 675–682.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences Vol. **22**, 2nd ed. (Springer, Berlin).

Fractal ladder models and power law wave equations

James F. Kelly[a)]
Department of Applied Mathematics, Naval Postgraduate School, Monterey, California 93943

Robert J. McGough
Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824

(Received 9 December 2008; revised 20 July 2009; accepted 21 July 2009)

The ultrasonic attenuation coefficient in mammalian tissue is approximated by a frequency-dependent power law for frequencies less than 100 MHz. To describe this power law behavior in soft tissue, a hierarchical fractal network model is proposed. The viscoelastic and self-similar properties of tissue are captured by a constitutive equation based on a lumped parameter infinite-ladder topology involving alternating springs and dashpots. In the low-frequency limit, this ladder network yields a stress-strain constitutive equation with a time-fractional derivative. By combining this constitutive equation with linearized conservation principles and an adiabatic equation of state, a fractional partial differential equation that describes power law attenuation is derived. The resulting attenuation coefficient is a power law with exponent ranging between 1 and 2, while the phase velocity is in agreement with the Kramers–Kronig relations. The fractal ladder model is compared to published attenuation coefficient data, thus providing equivalent lumped parameters. © *2009 Acoustical Society of America.* [DOI: 10.1121/1.3204304]

PACS number(s): 43.80.Cs, 43.20.Hq, 43.35.Bf [TDM] Pages: 2072–2081

I. INTRODUCTION

The attenuation coefficient in human and mammalian tissue in the ultrasonic range has a power law dependence on frequency.[1–3] The power law exponent typically ranges between 1 and 1.7 for most tissue.[2,4] Moreover, the power law exponent has been experimentally correlated with the pathological state of tissue.[2,5,6] Classical theories for ultrasonic absorption, such as thermo-viscosity,[7] and Biot's porous media theories[8] predict a frequency-squared dependence in the low-frequency limit, while classical relaxation predicts an attenuation coefficient with a resonant peak at the relaxation frequency of the material. However, neither of these behaviors is observed in soft tissue. Multiple-relaxation mechanism models[9] predict power law behavior over a narrow frequency band by empirically choosing the proper weights and relaxation frequencies, yet these models also fail to explain power law behavior over large frequency bands.

Other phenomenological models for ultrasonic attenuation in biological media have also been proposed. Frequency-domain descriptions include the linear phase model,[10] the Hilbert dispersive model,[11] the material impulse response model,[12] and power law models.[13–15] These methods evaluate a frequency-dependent transfer function and then perform a numerical inverse Fourier transform. Attenuation and dispersion have also been modeled directly in the time-domain via partial differential equations (PDEs) and fractional partial differential equations (FPDEs). PDE formulations incorporate loss via integer-ordered derivatives,[16,17] whereas FPDEs add loss to the wave equation with a time-fractional derivative,[18,19] a space-fractional derivative,[20,21] or the combination of an integer-ordered spatial derivative and a time-fractional derivative.[22,23] Nonlinear dissipative propagation has also been described with fractional spatial derivatives via generalizations of Burgers equation,[24] as well as transient elastic wave propagation in porous[25] and viscoelastic media.[26] These FPDE models build on previous applications of fractional calculus to diffusion processes,[27,28] relaxation processes,[29] viscoelasticity,[30,31] and seismology.[22] A third approach utilizes doublet mechanics,[32] whereby discrete microstructures are incorporated into the wave equation.

Within the viscoelastic and biomechanics communities, lumped parameter networks, such as the Maxwell and Voigt models, are commonly employed[33,34] to model the mechanical response of cells[34–36] and bulk tissue[37,38] to an applied force. Lumped parameter networks have also been extended to include infinite-ladder networks consisting of alternating elastic and viscous elements[33,39–41] and fractal tree networks,[42] which generate time-fractional rheological constitutive equations[30] for polymers. As discussed in Refs. 39, 33, and 40–42, the time-fractional derivative in the constitutive equation captures the (1) elastic, (2) viscous, and (3) self-similar properties described by these infinite networks. To date, however, fractal ladder networks have not been applied to dispersion in soft tissue. Moreover, time-fractional derivatives have been linked with diffusion[27] and relaxation[28] on fractal structures. In these studies, the order of the fractional derivative is a function of the fractal dimension of the underlying geometry. Within the ultrasonics community, however, this quantitative relationship between the fractal nature of tissue and FPDE has not been explored quantitatively.

This paper describes a model for the dissipative properties of soft tissue that employs hierarchical fractal networks.

[a)]Author to whom correspondence should be addressed. Electronic mail: jfkelly@nps.edu

Section II provides biological motivation and a review of linear stress-strain relationships in viscoelastic materials. A ladder model is then proposed by considering fractal networks of springs and dashpots.[33,40–42] Section III derives a fractional constitutive equation from the ladder model using tools from fractional calculus. On the basis of this constitutive equation, a FPDE utilizing a combination of integer-ordered spatial derivatives and a time-fractional derivative is derived from basic conservation laws in Sec. IV for linear macro-homogeneous media. This FPDE, which was originally proposed within the seismology community[22] and later considered in the biomedical acoustics community,[23] yields both a power law attenuation coefficient and a phase velocity predicted by the Kramers–Kronig relationships. In Sec. V, the ladder model is matched to published data and analyzed in terms of previous biomechanical and fractal models, followed by the conclusion in Sec. VI.

II. FRACTAL LADDER MODEL

This section introduces a lumped parameter, fractal ladder network to model the stress-strain relationship in biological media. From this fractal ladder network, a time-fractional derivative constitutive equation is derived, thus providing a physical basis for time-fractional FPDE such as the models proposed by Caputo[22] for seismic wave propagation in the earth and Wismer[23] for ultrasonic wave propagation in tissue.

A linear constitutive equation postulates a functional relationship between the time-dependent stress tensor $T_{ij}(t)$ and time-dependent strain tensor $\epsilon_{ij}(t)$ via a differential, integral, or integro-differential relationship that satisfies the principle of superposition. Familiar examples of constitutive equations, such as Hooke's law for an elastic solid and Newton's law for a viscous fluid, fail to predict the behavior of many viscoelastic solids; therefore, generalized viscoelastic models involving fractional derivatives and integrals have been developed.[30,31] This section proposes a constitutive equation for biological media using a fractal ladder network as a lumped parameter model. A qualitative biological model is first postulated, followed by the basic theory of viscoelasticity for a linear, non-Newtonian fluid.

A. Biological motivation

A mechanical model for the loss mechanism in mammalian biological tissue is motivated in this section. This model satisfies a power law attenuation coefficient $\alpha(\omega)$ of the form

$$\alpha(\omega) = \alpha_0|\omega|^y \tag{1}$$

over an appropriate frequency band, where ω is angular frequency and $1 < y \leq 2$ is the power law exponent. Soft tissue consists of hierarchical arrangements of elastic and fluid-like components. These tissues are highly heterogeneous and composed of over a hundred distinct cell types.[43] Each tissue consists of aggregates of cells suspended by a fluid-like extra-cellular matrix (ECM). The ECM is often modeled as an aqueous solution of viscoelastic polymers, which possess both solid and fluid-like properties. Individual cells are modeled as elastic membranes containing fluid-like cytoplasm.[34] Within the cytoplasm are distributed organelles, such as the nucleus, endoplasmic reticulum, and lysosomes, which in turn have an elastic membrane containing with a fluid-like interior.[43]

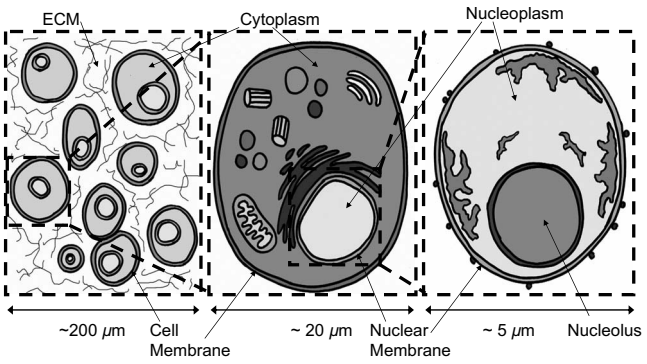


FIG. 1. Schematic showing tissue structure at three different spatial scales (tissue, cellular, and sub-cellular). The first panel ($\sim$200 μm) displays an ensemble of mammalian cells, each bounded by an elastic membrane, suspended in a viscoelastic ECM. The second panel ($\sim$20 μm) displays an individual cell at a higher level of magnification. The third panel ($\sim$5 μm) displays the cell nucleus, consisting of a double membrane, a fluid-like nucleoplasm, and an elastic nucleolus in the interior, which contains chromatin (Ref. 43). Although the specific biological structures vary at each successive spatial scale, the essential features are the same: fluid substrates containing elastic compartments. This self-similar pattern forms the basis for the fractal structure shown in Fig. 2.

This hierarchical arrangement is displayed at several scales in Fig. 1. Panel (a), which is on the scale of 200 μm, contains an ensemble of mammalian cells, each bounded by an elastic membrane that is suspended in a viscoelastic ECM. Both the ECM and cytoplasm consist of complex polymers (e.g., collagen) dissolved in a viscous fluid. The resulting structure is therefore a viscoelastic material. In panel (b), which is on the scale of 20 μm, an individual cell is shown at a higher level of magnification. Inside the elastic membrane is the cytoplasm, which has viscoelastic properties that are similar to the ECM. Panel (c) displays the cell nucleus on the scale of 5 μm, consisting of a double membrane, a fluid-like nucleoplasm, and an elastic nucleolus in the interior, which contains chromatin.[43] Based on the structure shown in Fig. 1, both the viscoelastic properties and self-similar, or fractal, properties of tissue are evident, which motivates a simplifying hierarchical mathematical model. Applying this biological picture, tissue may be visualized as a recursive arrangement of fluid substrates containing elastic membranes. Similar models, known as liquid drop models, have been proposed within the biomechanics community to describe the deformation of eukaryotic cells.[34,35] These liquid drop models typically model the cell membrane as a cortical layer with a characteristic surface tension, whereas the cytoplasm in the cell interior is modeled as a viscous, incompressible fluid with a characteristic coefficient of viscosity. The cell nucleus may also be included as an additional elastic component embedded within the viscous fluid.

In the following fractal model for the viscoelastic properties of tissue, an infinite number of nested elastic membranes, each containing a viscous compressible fluid, is proposed. By defining an infinite number of layers, larger structures (e.g., ensembles of cells) and smaller structures (e.g., cell nuclei) may be included within the lumped param-

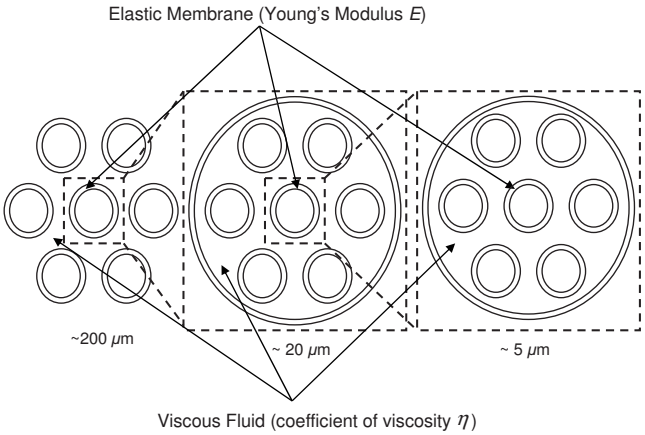


FIG. 2. Layered fractal model for biological tissue based on the schematic shown in Fig. 1. The first panel displays an infinite number of thin elastic membranes with Young's modulus E alternating with viscous compartments that have coefficients of viscosity η. The second panel zooms in on the first panel, thus showing the self-similar layered structure.

eter framework. By allowing the number of structural components to extend indefinitely, the self-similarity of biological media is revealed. This topology is depicted in Fig. 2, where the alternating elastic and viscous components are visualized as a self-similar hexagonal packing of spheres within spheres. Each of the three panels in Fig. 2 corresponds to the three panels in Fig. 1. That is, the left panel of Fig. 2 models the tissue level, the center panel models the cellular level, and the right panel models the sub-cellular level. Comparing the three panels of Figs. 1 and 2, the self-similar nature of this fractal structure is immediately evident. Hence, this fractal model captures the three essential features of the biological picture shown in Fig. 1: (1) elastic membranes, (2) fluid compartments, and (3) self-similarity over a range of spatial scales.

In order to capture these three salient properties of biological media, a fractal network of springs and dashpots is proposed in Fig. 3. The elastic membranes displayed in Fig. 2 are represented by springs with Young's modulus E, while the viscous compartments are represented by dashpots with coefficients of viscosity η. Each level in Fig. 3 corresponds

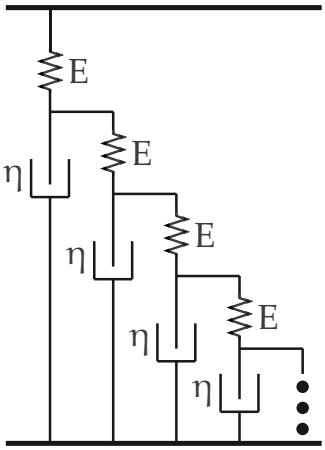


FIG. 3. Fractal ladder model for tissue micro-structure. The continuum model depicted in Fig. 2 is described by an infinite fractal ladder consisting of springs with Young's modulus E and coefficients of viscosity η.

to a pair of elastic/viscous layers shown in Fig. 2 and an individual level of magnification in Fig. 1. This fractal network is analyzed in depth in Sec. II C. Similar fractal networks have previously been used in lumped parameter models of viscoelastic systems[39,44] such as cross-linked polymers and gels.[41]

B. Stress-strain relationships

In this section, the well-known constitutive equation for a viscous Newtonian fluid relating the strain ϵ_{ij} and stress T_{ij} tensors is generalized to self-similar biological media. The strain tensor is defined via $\epsilon_{ij}=\partial w_i/\partial x_j$, where w_i denotes the ith component of displacement. The stress tensor T_{ij} denotes the ith component of stress per unit area along a surface normal to the jth direction, where $1\leq i, j\leq 3$ denote the x, y, and z directions. The viscous stress tensor T_{ij} for a compressible, viscous fluid with pressure p and velocity $\mathbf{u}$ is given by[7]

$$T_{ij}=-p\delta_{ij}-\frac{2}{3}\mu\nabla\cdot\mathbf{u}\delta_{ij}+\mu\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right), \tag{2}$$

where μ is the coefficient of shear viscosity, δ_{ij} is the Kronecker delta operator, and $i,j=1, 2, 3$. For homogeneous gases, μ may be computed via kinetic theory. For more complicated fluids, μ must be measured experimentally. Equation (2) contains three terms: (1) an elastic term involving the thermodynamic pressure p, (2) an isotropic frictional term, and (3) a shearing term. In Eq. (2), the coefficient of bulk viscosity is assumed to be zero.

Although Eq. (2) may describe the stress-strain relationship on a sufficiently small micro-scale with a variable viscosity μ, Eq. (2) fails to predict dissipative behavior on the macro-scale in most biological media. To obtain a constitutive equation on a scale commensurate with an acoustic wavelength (macro-scale), Eq. (2) is averaged over a sufficiently large volume to achieve a constitutive equation with a constant coefficient. This averaging, or up-scaling procedure, should account for the signature micro-heterogeneity and hierarchical micro-structure of biological media. One simple up-scaling procedure employs the lumped parameter model discussed in Sec. II A, wherein the individual components of the medium (cells, membranes, organelles, etc.) are represented via hierarchical arrangements of springs and dashpots.

On the macro-scale, the normal stress is decomposed according to

$$T_{ij}(\mathbf{r},t)=-p(\mathbf{r},t)\delta_{ij}+\sigma_{ij}(\mathbf{r},t), \tag{3}$$

where $\sigma(\mathbf{r},t)$ is the component of stress responsible for dissipation. For a viscous medium, $\sigma(\mathbf{r},t)$ corresponds to the second and third terms in Eq. (2). For a linear material, $\sigma_{ij}(\mathbf{r},t)$ is a linear function of strain $\epsilon_{ij}(\mathbf{r},t)$. The dissipative component of Eq. (2) can be generalized to include memory effects by relating each component of stress and strain to a causal, stationary, hereditary integral (or Boltzmann superposition integral):[33]

$$\sigma_{ij}(\mathbf{r},t) = \int_{-\infty}^{t} g(t-t')\left[-\frac{2}{3}\sum_{t=1}^{3}\epsilon_{il}(\mathbf{r},t')\delta_{ij} + \epsilon_{ij}(\mathbf{r},t') + \epsilon_{ji}(\mathbf{r},t')\right]dt', \tag{4}$$

where $g(t)$ is a *relaxance*, or memory, function[33] which relates the present state of the material to the previous history. Since Eq. (4) is a convolution integral, the stress-strain relationship becomes multiplication in the Laplace domain. In one dimensional (1D), the stress tensor has only the component $\epsilon(t)=\epsilon_{11}(t)$, yielding the convolution $\sigma=4/3g(t)*\epsilon(t)$. Applying a Laplace transform and invoking the convolution theorem yields

$$\hat{\sigma}(\mathbf{r},s) = \tfrac{4}{3}\hat{g}(s)\hat{\epsilon}(\mathbf{r},s). \tag{5}$$

C. Infinite ladder

In this section, a fractal ladder model is constructed in 1D. By assuming that tissue is isotropic on the macro-scale, the 1D model is then extended to three dimensional (3D). The combined viscous and elastic components are modeled as springs and dashpots, respectively. Springs, which model energy storage, represent the nested elastic membranes shown in Fig. 2, while dashpots, which model dissipation, represent the viscous components, such as cytoplasm. The self-similar structure is realized as a fractal ladder in Fig. 3, which provides a lumped parameter description of the geometric model shown in Fig. 2. All of the springs in this model have the same spring constant, or Young's modulus, E and all of the dashpots have the same coefficient of viscosity η. The transfer function $\hat{g}(s)$ for the stress-strain relationship given by Eq. (5) is then evaluated as an infinite, periodic continued fraction:

$$\hat{g}(s) = \eta s + \cfrac{1}{E^{-1} + \cfrac{1}{\eta s + \cfrac{1}{E^{-1} + \cdots}}} = \frac{-\eta s/E + \sqrt{\eta s/E(\eta s/E + 4)}}{2/E}, \tag{6}$$

where the periodic continued fraction is evaluated in closed form.[41,42] For $s\eta/E \ll 1$, the binomial approximation is applied, yielding the low-frequency approximation

$$\hat{g} \approx \sqrt{\eta E s}. \tag{7}$$

Inserting Eq. (7) into Eq. (5), and performing an inverse Laplace transform by applying Eq. (A2) from Appendix A, yields

$$\sigma = \frac{4}{3}\sqrt{\eta E}\frac{\partial^{1/2}\epsilon}{\partial t^{1/2}}, \tag{8}$$

where the fractional derivative operator is defined by Eq. (A1) in Appendix A.

The ladder model can also be considered as a fundamental mechanical component (a "springpot"[45]), allowing more complicated fractal networks, or recursive ladders, to be constructed. For instance, consider a recursive ladder model constructed by replacing the viscous damper in Fig. 3 with a fractal ladder, producing the arrangement shown in Fig. 4. Similarly, a recursive ladder may be constructed by replacing the springs in Fig. 3 with a fractal ladder, producing the arrangement shown in Fig. 5. As shown in Appendix B, the ladder model may be generalized to recursive ladder topologies. Similar fractal tree networks were considered in Ref. 42 to model the power law response of polymers. These recursive ladder models, which are developed in Appendix B, yield the stress-strain relationship

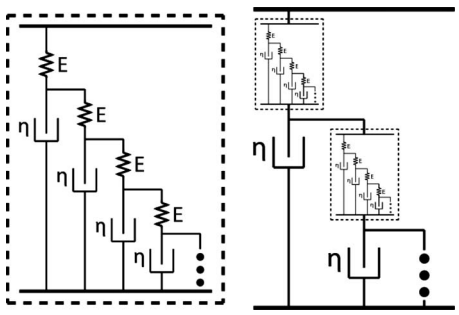


FIG. 4. Example of a recursive fractal ladder model where the dampers in the simple ladder shown in Fig. 3 are replaced with ladders. This particular fractal arrangement is denoted by $(M,N)=(0,1)$ and yields a fractional derivative with $\beta=1/4$.

$$\sigma = \frac{4}{3}\eta^{\beta}E^{1-\beta}\frac{\partial^{\beta}\epsilon}{\partial t^{\beta}}, \tag{9}$$

which is a generalization of Eq. (8) for all $0<\beta\leq 1$. The fractional derivative order β is specified by (see Appendix B)

$$\beta = \frac{1}{2}\left(1 - \frac{1}{2^M} + \frac{1}{2^N}\right), \tag{10}$$

where M is the depth of recursion of the dampers and N is the depth of recursion of the springs. In the special case of the simple ladder $(M,N)=(0,0)$, the recursive ladder reduces to the simple ladder with $\beta=1/2$. For $M\neq N$, different fractional orders β are obtained. For instance, if $(M,N)=(1,0)$, then each damper in Fig. 3 is replaced by a ladder, yielding $\beta=3/4$. If $(M,N)=(0,1)$, then each spring in Fig. 3 is replaced by a ladder, yielding $\beta=1/4$. Note that Eq. (10) does not uniquely define β for given $M,N\geq 0$. For instance, the

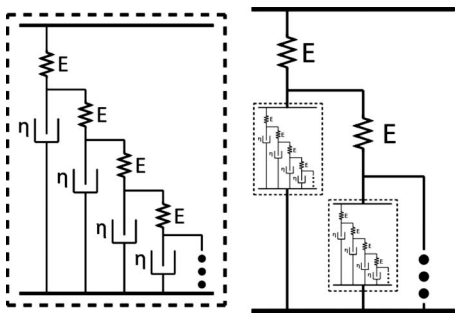


FIG. 5. Example of a recursive fractal ladder model where the springs in the simple ladder shown in Fig. 3 are replaced with ladders. This particular fractal arrangement is denoted by $(M,N)=(1,0)$ and yields a fractional derivative with $\beta=3/4$.

value of $\beta=1/2$ is recovered for any choice of $M=N$.

III. 3D FRACTIONAL CONSTITUTIVE EQUATION AND WAVE EQUATION

A. Fractional constitutive equation

Generalizing Eq. (9) to 3D (under the macro-isotropic assumption) yields

$$\sigma_{ij} = -\left(\frac{2}{3}E_0^{1-\beta}\eta_0^{\beta}\frac{\partial^{\beta}}{\partial t^{\beta}}\sum_{j=1}^{3}\epsilon_{ij}\right)\delta_{ij} + E_0^{1-\beta}\eta_0^{\beta}\frac{\partial^{\beta}}{\partial t^{\beta}}(\epsilon_{ij}+\epsilon_{ji}). \tag{11}$$

Constitutive equations similar to Eq. (14) have been previously proposed for viscoelastic materials within the geology community.[22] Similar to Eq. (2), Eq. (11) is expressed in terms of the particle velocity $\mathbf{u}$, yielding

$$\sigma_{ij} = -\left(\frac{2}{3}E_0^{1-\beta}\eta_0^{\beta}\frac{\partial^{\beta}}{\partial t^{\beta}}\sum_{j=1}^{3}\epsilon_{ij}\right)\delta_{ij} + E_0^{1-\beta}\eta_0^{\beta}\frac{\partial^{\beta-1}}{\partial t^{\beta-1}}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{12}$$

where Einstein summation notation is utilized. Thus, a time-fractional stress-strain relationship follows from the fractal ladder originally proposed for polymer modeling in Ref. 41. Equation (12) consists of two viscoelastic terms involving the fractional derivative of strain with respect to time, where the fractional derivative term is responsible for the coupled processes of attenuation and dispersion. For $\beta<1$, Eq. (12) displays a temporal non-locality commonly utilized in phenomenological viscoelasticity.[22,30] Theoretical justification for constitutive equations similar to Eq. (12) was also established[31] on the basis of dilute solutions of polymers[46] within a homogeneous, Newtonian solvent.

By identifying a generalized coefficient of viscosity

$$\mu = E_0^{1-\beta}\eta_0^{\beta}, \tag{13}$$

the following averaged constitutive equation involving velocity gradients is computed:

$$T_{ij} = -p\,\delta_{ij} - \frac{2}{3}\mu\frac{\partial^{\beta-1}}{\partial t^{\beta-1}}\frac{\partial u_i}{\partial x_i}\delta_{ij} + \mu\frac{\partial^{\beta-1}}{\partial t^{\beta-1}}\left(\frac{\partial u_i}{\partial x_j}+\frac{\partial u_j}{\partial x_i}\right). \tag{14}$$

For $\beta=1$, Eq. (14) reduces to the viscous stress tensor for a compressible, Newtonian fluid given by Eq. (2). Equation (14) contains three terms: (1) an elastic term involving the thermodynamic pressure p, (2) an isotropic frictional term, and (3) a shearing term. The frictional terms tend to diffuse momentum through the flow. In a viscous fluid ($\beta=1$), the frictional term involves only an integer-ordered derivative and is purely local. For a homogeneous fluid with simple molecular structure, this relation properly accounts for momentum diffusion. For biological tissue, however, viscous loss does not properly account for observed dissipation. Tissue is both heterogeneous and has a complex molecular structure that can be modeled as a viscoelastic medium. Physically, momentum may diffuse faster and/or slower in some directions due to the heterogeneity of tissue. To describe these effects, the local constitutive relationship is generalized by a global relation that incorporates memory into the flow. The temporal operator $\partial^{\beta-1}/\partial t^{\beta-1}$ is the Riemann–Liouville fractional derivative for $0<\beta<1$. Thus, the viscous shearing term in a standard Newtonian fluid is replaced with a memory term which relates the stress at time t to the entire history of the velocity gradient. Similar hereditary constitutive equations, which utilize time-fractional and time-convolutional operators, are widely used in theoretical viscoelasticity.[26,47]

B. 3D fractional wave equation

A time-fractional wave equation was originally introduced by Caputo[22] to model dissipative elastic wave motion in geological media. Equation (8) in Ref. 22 contains a FPDE that models 1D, plane wave propagation in a viscoelastic solid. Later, Wismer[23] independently obtained a 3D version of this equation that models ultrasonic wave motion in power law biological media. This section demonstrates how the fractal ladder model and fractional constitutive equation in Sec. II lead to a FPDE that describes dispersive wave propagation in biological media. In particular, the 3D fractional wave equation,[23] which models power law attenuation via a time-fractional derivative, is derived for a linear, macro-homogeneous, and isotropic medium governed by Eq. (14). Since this derivation is similar to the derivation of the thermoviscous wave equation presented in Ref. 7, only the major points are emphasized here.

Linear, longitudinal wave motion is considered in a homogeneous medium with density ρ_0, sound speed c_0, and generalized viscosity μ. Here, shear mode propagation is neglected. The adiabatic hypothesis, whereby entropy is assumed constant, is also adopted; thus the additional dissipative effects of thermal conduction are neglected. Equation (14) is complemented by (1) the linearized Cauchy's equation, (2) the linearized, adiabatic equation of state, and (3) the linearized equation of continuity. The linearized Cauchy equation, which neglects the convective term $\mathbf{u}\cdot\nabla\mathbf{u}$, restates Newton's second law of motion as

$$\rho_0\frac{\partial u_i}{\partial t} = \frac{\partial T_{ij}}{\partial x_j}. \tag{15}$$

The linearized, adiabatic equation of state is given by $p=c_0^2\delta\rho$, where $\delta\rho$ denotes excess density and c_0 is the adiabatic speed of sound. Finally, the linearized equation of continuity is given by

$$\frac{\partial\delta\rho}{\partial t} + \rho_0\frac{\partial u_i}{\partial x_i} = 0, \tag{16}$$

which accounts for local mass conservation. First, the divergence of Eq. (14) is evaluated and inserted into Eq. (15). Neglecting the transverse component of velocity (see Ref. 7 for details) yields

$$\rho_0\frac{\partial u_i}{\partial t} = -\frac{\partial p}{\partial x_i} + \frac{4}{3}\mu\nabla^2 u_i. \tag{17}$$

Applying the divergence operator to both sides of Eq. (17) yields

$$\rho_0 \frac{\partial}{\partial t}\left(\frac{\partial u_i}{\partial x_i}\right) = -\nabla^2 p + \frac{4}{3}\mu \frac{\partial^{\beta-1}}{\partial t^{\beta-1}}\nabla^2\left(\frac{\partial u_i}{\partial x_i}\right). \tag{18}$$

Finally, to derive a wave equation in terms of the pressure, Eq. (16) is combined with the equation of state and inserted into Eq. (18), yielding

$$\frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} = \nabla^2 p + \frac{4\mu}{3c_0^2\rho_0}\frac{\partial^\beta}{\partial t^\beta}\nabla^2 p. \tag{19}$$

Identifying the relaxation time

$$\tau^\beta = \frac{4\mu}{3c_0^2\rho_0} \tag{20}$$

yields the 3D fractional wave equation[23]

$$\nabla^2 p - \frac{1}{c_0^2}\frac{\partial^2 p}{\partial t^2} + \tau^{y-1}\frac{\partial^{y-1}}{\partial t^{y-1}}\nabla^2 p = 0, \tag{21}$$

where $y=\beta+1$ is the power law exponent. Frequency-dependent loss is incorporated via the Riemann–Liouville fractional derivative defined in Appendix A. Equation (21) is identical to Eq. (5) in Ref. 23 and may be expressed as Eq. (8) in Ref. 22 for 1D problems. For $y=2$, Eq. (21) reduces to the Stokes wave equation,[48], which models wave propagation in a homogeneous, viscous medium. As shown in Ref. 23, for $y>1$, Eq. (21) admits an attenuation coefficient with a power law dependence in the low-frequency limit. For $y=1$, however, the loss operator reduces to a spatial Laplacian, and this equation thereby fails to model power law attenuation. For this reason, the exponent is restricted to $1<y\leqslant 2$. Hence, Eq. (21) arises naturally as a wave equation that models small-amplitude, longitudinal disturbances in media governed by the fractional constitutive equation given by Eq. (14). In addition, the parameter τ appearing in Eq. (21) is given physical meaning by Eq. (20), which depends on the micro-structural properties of the medium μ as well as the macroscopic properties c_0 and ρ_0.

C. Power law attenuation

Several important relations, which were discussed in Refs. 22 and 23, are briefly reviewed in this subsection. To derive a power law attenuation coefficient for Eq. (21), the dispersion relationship between angular frequency ω and spatial wavenumber k is calculated. Applying a space-time Fourier transform to Eq. (21) yields

$$-k^2 + \frac{\omega^2}{c_0^2} - k^2\tau^{y-1}(j\omega)^{y-1} = 0. \tag{22}$$

Solving for the wavenumber $k(\omega)$ yields

$$k(\omega) = \frac{\omega}{c_0\sqrt{1+(j\tau\omega)^{y-1}}}. \tag{23}$$

In the low-frequency limit, the binomial approximation is applied, yielding

$$k(\omega) \approx \frac{\omega}{c_0}\left(1 - \frac{\tau^{y-1}}{2}\cos\left(\frac{(y-1)\pi}{2}\right)\omega^{y-1} - j\frac{\tau^{y-1}}{2}\sin\left(\frac{(y-1)\pi}{2}\right)\omega^{y-1}\right) \tag{24}$$

for an outgoing wave. The attenuation coefficient $\alpha(\omega)$ is computed by taking the imaginary part of Eq. (24), yielding the power law coefficient given by Eq. (1) where

$$\alpha_0 = \frac{\tau^{y-1}|\cos(\pi y/2)|}{2c_0}. \tag{25}$$

Likewise, the propagation constant is computed by taking the real part of Eq. (24), yielding Eq. (10) in Ref. 23. Hence, Eq. (24) agrees with the phase velocity predicted by the Kramers–Kronig relations and the local time-causal theory.[49] Note that other FPDE models for power law media, such as the Szabo wave equation[18] and the power law FPDE in Ref. 50, satisfy the same relationship.

IV. RESULTS

This section evaluates the lumped parameter values μ, E_0, η_0, and (M,N) using published values for attenuation coefficients and other acoustic parameters. Given values of the attenuation constant α_0, the power law exponent y, the speed of sound c_0, and the density ρ_0, values for the generalized viscosity μ, an equivalent coefficient of viscosity η_0, and Young's modulus E_0 may be computed using the relations derived in Secs. II and III. In addition, given a specified power law exponent y, a recursive ladder may be constructed using Eq. (10). Measured attenuation, sound speed, and density values from Ref. 4 were utilized in these calculations. To compute the generalized viscosity μ, Eqs. (20) and (25) are combined, yielding

$$\mu = \frac{3\alpha_0 c_0^3\rho_0}{2|\cos(\pi y/2)|}. \tag{26}$$

Once μ is determined, Young's modulus E_0 and coefficient of viscosity η_0 are computed using Eq. (13) subject to the constraint $\eta_0\omega_{\max}/E_0 \ll 1$, where $\omega_{\max}=2\pi f_{\max}$ is the largest angular frequency of interest. These attenuation data are measured over a range of frequencies between 0.5 and 10 MHz, which are typical operating frequencies in diagnostic and therapeutic ultrasound. Since the fractal ladder model allows independent values for both E_0 and η_0, the coefficient of viscosity η_0 was fixed using the total viscosity of water (shear plus bulk) at room temperature[51] $\eta_0 \approx 0.004$ Pa s. This choice for η_0 satisfies the above constraint for all cases considered in this analysis.

Table I displays the results of this procedure for four tissue types: (1) breast fat, (2) liver, (3) spleen, and (4) cyst fluid. From the results of this analysis, there is a large range of predicted generalized viscosity μ values. Breast fat has the smallest generalized viscosity (3.18×10^3 Pa s), while liver has the largest generalized viscosity (2.32×10^6 Pa s). Spleen and cyst fluid have intermediate values of μ given by 8.96×10^4 and 1.30×10^4 Pa s, respectively. This large variation in μ values is attributed to the large variation in (1) the

TABLE I. Equivalent generalized viscosity μ, Young's modulus values E, and ladder parameters (M,N) calculated using density, speed of sound, and attenuation parameters from Ref. 4. The equivalent Young's modulus is calculated assuming an equivalent bulk viscosity of water $\eta_0=0.004$ Pa s.

Tissue	Breast fat	Liver	Spleen	Cyst fluid
Density ρ_0 (kg/m^3)	930	1050	1054	1000
Sound speed c_0 (m/s)	1436	1578	1567	1568
Attenuation constant α_0 (Np/cm MHz^y)	0.086	0.046	0.046	0.0058
Exponent y	1.50	1.14	1.30	1.29
Predicted generalized viscosity μ (Pa s)	3.18×10^3	2.32×10^6	8.96×10^4	1.30×10^4
Predicted Young's modulus E_0 (MPa)	500	24.1	39.7	1.89
Ladder parameters (M,N)	(0,0)	(0,2)	(1,3)	(1,4)

observed attenuation coefficient values over the frequency range 1–10 MHz and (2) the $|\cos(\pi y/2)|$ factor in the denominator of Eq. (26), which amplifies μ for y values near unity. Consequently, there is also a large range of equivalent Young's modulus values, ranging from a maximum value of 500 MPa for breast fat to a minimum value of 1.89 MPa for cyst fluid. Liver and spleen have intermediate equivalent Young's moduli of 24.1 and 39.7 MPa, respectively. Also, the low-frequency limit $\eta_0 s/E_0 \ll 1$ is satisfied for this combination of parameters. For instance, in breast fat, $\eta_0 s/E_0 \ll 1$ for frequencies less than 1 GHz, which is satisfactory for diagnostic and therapeutic applications of ultrasound. On the other hand, for cyst fluid, the low-frequency limit is only valid up to about 10 MHz.

Given a power law exponent y, approximate values of (M,N) for the recursive depth of the ladder are calculated via Eq. (10), yielding

$$y=\frac{3}{2}-\frac{1}{2^{N+1}}+\frac{1}{2^{M+1}}. \tag{27}$$

From the computed values of (M,N) in Table I, breast fat is well modeled by the simple ladder $(M,N)=(0,0)$, while liver, spleen, and cyst fluid require recursive ladder topologies. Liver is approximated by a (0,2) ladder, spleen by a (1,3) ladder, and cyst fluid by a (1,4) ladder. Since the simple ladder has an equal number of springs and dashpots, breast fat has equal elastic and viscous contributions according to the proposed model. However, liver, spleen, and cyst fluid all require greater recursive spring depths N to match the observed power law exponents. In the (0,2) ladder that approximates liver, twice as many springs are present relative to dampers. Hence, according to the recursive fractal ladder, liver has a greater contribution from the elastic, as opposed to the viscous, component. Spleen and cyst fluid, which are modeled by (1,3) and (1,4) ladders, also have more springs than dashpots in the ladder and are hence more elastic than viscous.

The ladder model demonstrates how a small coefficient of viscosity may be combined with springs that have large Young's modulus to produce an attenuation coefficient that is large relative to that of water. Furthermore, in this model, Young's modulus E_0 is more than an order of magnitude smaller in liver than in breast fat. This behavior is expected for two reasons: (1) the attenuation constant α_0 is smaller for liver than for fat and (2) the recursive ladder model for liver $(M=2$ and $N=0$, yielding $y=1.125\approx1.14)$ contains about two times as many springs as fat, which is well modeled by a simple ladder $(M=N=0)$ that predicts a power law exponent $y=1.5$.

V. DISCUSSION

A. Bio-mechanical interpretation

The physical significance of the power law exponent is explored in this section. The order of the fractional derivative β and hence the power law exponent y are determined by the recursion level of the damper-ladders M and spring ladders N using Eq. (10). The fractal ladder model is applicable to four special cases, each of which is discussed below.

Case I: Micro-homogeneous media $(y=2)$. Let $E=0$ and $\eta>0$. The viscous theory is recovered using a single dashpot, yielding a power law coefficient of $y=2$. In this degenerate case, the model is not fractal. Due to the lack of springs, the medium is homogeneous at all scales much smaller than a wavelength, thereby indicating a lack of micro-heterogeneity.

Case II: Simple ladder model $(y=3/2)$. Let $M=N=0$. This case is the simple ladder topology shown in Fig. 3. The behavior of breast fat is captured by the simple ladder model, where the contributions from the elastic and viscous components are roughly equal. The $y=3/2$ case is also recovered if $M=N$ $(\neq0)$ where the relative depth of recursion of dampers and springs is the same.

Case III: Recursive ladder model with $M<N$ $(1<y<3/2)$. In this case, the depth of recursion of springs is greater than the depth of recursion of dampers. Hence, at any given level of the ladder, there are more springs than dampers, indicating that the medium has a greater elastic component than viscous component. Since springs correspond to elastic structures such as cellular and nuclear membranes, while dampers correspond to inter- and intra-cellular fluids such as cytoplasm, the exponent y measures the relative mechanical contributions of elastic versus viscous structures. The power law exponent y ranges from 1 to 1.5 in this case, which is typical for most soft tissue.[4] For instance, anatomical media such as liver have y close to 1 due to the relatively complex tissue structure. This suggests that soft tissue generally has a greater elastic component than viscous component, and these elastic components play a greater role in the dissipation of ultrasonic energy.

Case IV: Recursive ladder model with $M>N$ $(3/2<y<2)$. In this case, the depth of recursion of dampers is

greater than the depth of recursion of springs. Therefore, at any level of the ladder, there are more dampers than springs, indicating that the medium has a greater viscous component than elastic component. The power law y ranges from 1.5 to 2 in this case, which is typical for complex fluids like castor oil and silicone fluid.[52]

The compound ladder model also sheds light on the dependence of the power law exponent on the pathological state of tissue. For example, in Ref. 5, the power law exponent y in normal liver exhibits $y \approx 1.1$, whereas y ranges from 1.25 to 1.4 in fatty liver. The increase in the power law exponent has been explained in terms of an increase in Rayleigh scattering in fatty liver relative to healthy liver.[5] Within the context of the present compound ladder model, the increase in y is explained as an increase in the viscous microstructure relative to healthy liver, which has a greater elastic component.

The recursive ladder model provides an explanation for the combined effects of absorption and incoherent scattering in biological media. First, note that the (local) speed of sound is a function of the spring constant E by combining Eqs. (20) and (25), and the frequency-dependent phase velocity. Thus, the fractal arrangement of springs qualitatively accounts for sound speed inhomogeneity at multiple spatial scales, resulting in incoherent scattering of an incident sound field. However, sound speed inhomogeneity is not solely responsible for the observed power law dependence of the attenuation coefficient. In addition, a viscous mechanism is required to dissipate both the incident and incoherently scattered sound fields. This viscous mechanism, like the sound-speed inhomogeneity, is represented at multiple spatial scales by the fractal model. The interactions between these two mechanisms are mediated by the hierarchical arrangement of springs and dashpots. Finally, the results obtained with the fractal ladder network, which predicts an exponent y ranging between 1 and 2, agrees with the bulk of experimental data collected for soft tissues, which is effectively modeled by a power law attenuation coefficient with $1 < y \leqslant 1.5$.

B. Fractal networks

Although the present model does not correlate the underlying tissue morphology with the power law exponent, some potentially meaningful information can be extracted from the computed values. Since the current model does not consider underlying tissue morphology, the parameters (M,N) cannot be interpreted in terms of tissue complexity. For instance, breast fat has a more complicated structure that cyst fluid, yet breast fat is described by a simple ladder whereas cyst fluid requires a recursive ladder. To extend this model, more sophisticated fractal networks, such as Sierpinski gaskets, may be considered.

Beginning with the pioneering work of Mandelbrot, fractal geometry has been a useful tool to explain the self-similar structure (e.g., alveolar surfaces, cell membranes, etc.) found in biological systems.[53] More recently, fractal geometry has been applied to understanding the pathological architecture of tumors.[54] Since the vasculature of tumors is more tortuous than healthy tissue, the measured fractal dimension of tumor vessels is significantly larger than normal veins and arteries.[54] To interpret the power law exponent within the context of fractal geometry, which was discussed in Ref. 20, a quantitative relationship is provided by applying the analysis presented in Ref. 41. In Ref. 41, a constitutive equation for cross-linked polymers is derived by first assuming a fractal arrangement of springs and dashpots with spectral dimension d_s and then formulating an equivalent random walk problem. The spectral dimension d_s is related to the vibrational properties of the underlying fractal network, such as the density of normal modes in the low-frequency limit.[55,56] Intuitively, d_s measures the connectivity of a fractal network and may be tailored to different tissue types. In future work, the ladder models developed in this paper will be extended to these more general fractal networks.

C. Inhomogeneous media and nonlinear media

In general, biological medium is inhomogeneous on both the microscopic scale (~ 1 μm) and the macroscopic scale (~ 1 mm). The fractional derivative operator in Eq. (21) accounts for the effect of micro-heterogeneity on the macroscopic scale. However, Eq. (21) does not account for the macro-heterogeneity that is responsible for coherent scattering. To incorporate macro-heterogeneity, the material properties of density $\rho_0(\mathbf{r})$, adiabatic compressibility $\kappa_0(\mathbf{r})$, and shear viscosity $\mu(\mathbf{r})$ are assumed to be functions of space. By utilizing the constitutive equation in Eq. (14), an inhomogeneous 3D fractional wave equation [see Eq. (11) in Ref. 23] may be derived via the methods presented in Sec. III. In addition, the 3D fractional wave equation, as derived in Sec. III, assumes small amplitude oscillations and negligible heat conduction by utilizing a linear, adiabatic equation of state. Although the adiabatic hypothesis is justified in most biological media due to negligible thermal conductivity, the linear assumption is not justified in many biomedical applications where large amplitude effects occur.[57] However, most nonlinear models, such as Burgers equation and Westervelt's equation, assume a thermoviscous dissipation mechanism, resulting in an attenuation coefficient with frequency-squared dependence. In order to combine the effects of power law attenuation with nonlinearity, several authors have formulated nonlinear FPDEs.[18,20,24] Finite amplitude effects may be incorporated into the 3D fractional wave model by augmenting the equation of state with a quadratic term. Utilizing the stress tensor given by Eq. (14), a nonlinear generalization of the 3D fractional wave equation may be derived for both homogeneous and inhomogeneous media. The competing effects of nonlinearity and power law dissipation may then be studied within the presented framework.

VI. CONCLUSION

This paper proposes a fractal ladder network of springs and dashpots to model wave propagation in power law attenuation media. Both a simple and a recursive fractal ladder model are considered in order to capture the viscoelastic, self-similar, and hierarchical properties of biological tissue. These fractal ladders capture the hierarchical arrangement of elastic and viscous components present in biological media.

The simple fractal ladder network produces a stress-strain relationship with a fractional derivative of order 1/2, while the recursive fractal ladder produces fractional derivatives of all orders between 0 and 1. Hence, the resulting constitutive equation interpolates between a Hookean solid and Newtonian fluid via the Riemann–Liouville fractional derivative operator.

When the constitutive equation in Eq. (14) is combined with the linear equation of state and the linear equations of mass and momentum conservation, Eq. (21), which models longitudinal wave propagation in power law media via a time-fractional derivative, is derived. Hence, a fractional PDE is derived from a fractal description of the medium. The attenuation coefficient computed from this constitutive equation follows a power law in the low-frequency limit. The ladder model is compared with measured attenuation data, thereby determining an equivalent Young's modulus and the topology of the ladder model.

ACKNOWLEDGMENTS

The authors thank Mark M. Meerschaert, Department of Statistics and Probability, Michigan State University, and Stephen W. Wheatcraft, Department of Geological Sciences and Engineering, University of Nevada, Reno, for useful discussion and advice. The drawings in Fig. 1 were prepared by Amy Albin, Department of Zoology, Michigan State University. Figures 3–5 were prepared by Christopher Johnson, Department of Electrical and Computer Engineering, Michigan State University. This work was funded in part by NIH Grant No. 1R21 CA121235 J.F.K. also acknowledges support from the NRC Postdoctoral Associateship program.

APPENDIX A: RIEMANN LIOUVILLE FRACTIONAL DERIVATIVES

The Riemann–Liouville fractional derivative is formally defined via a hyper-singular integral[58]

$$\frac{d^y f}{dt^y} = \frac{1}{\Gamma(-y)} \int_{-\infty}^{t} \frac{f(t')}{(t-t')^{1+y}} dt', \tag{A1}$$

where $\Gamma(z)$ is the gamma function. By letting $y<0$ in Eq. (A1), a fractional integration is realized. The following Laplace transform relationship for fractional derivatives is necessary:

$$\mathcal{L}\left(\frac{d^y g}{dt^y}\right) = s^y \mathcal{L}(g). \tag{A2}$$

Letting $s=j\omega$ yields the Fourier transform relationship

$$\mathcal{F}\left(\frac{d^y g}{dt^y}\right) = (j\omega)^y \mathcal{F}(g). \tag{A3}$$

APPENDIX B: RECURSIVE FRACTAL LADDER MODELS

Recursive ladders are constructed in this section. Consider a ladder model constructed by replacing each of the viscous dampers in Fig. 3 with a fractal ladder, producing the arrangement shown in Fig. 4. That is, a simple ladder is embedded within a larger ladder along with springs that have elastic coefficients E, denoted by a (0,1) network, as shown in Fig. 4. Computation of $\hat{g}(s)$ for this model using the low-frequency approximation given by Eq. (7) yields $\hat{g}(s) \approx E^{3/4}\eta^{1/4}s^{1/4}$. This construction may be extended by embedding a ladder within a ladder, yielding a (0, 2) network. This recursive ladder network is further generalized to N-1 level ladders alternating with springs to create an N-level ladder-spring network, yielding $\hat{g}(s) \approx E^{1-1/2^{N+1}}\eta^{1/2^{N+1}}s^{1/2^{N+1}}$. By performing an inverse Laplace transform, fractional derivative stress-strain relationships of order $1/2, 1/4, 1/8, \ldots$ are generated. As the depth of the ladder increases $(N\to\infty)$, $\hat{g}(s)\to E$, yielding a purely elastic response.

A similar recursive mechanical network is constructed with dashpots and fractal ladders, as shown in Fig. 5. Evaluating the transfer function for this recursive ladder yields $\hat{g}(s) \approx E^{1/4}\eta^{3/4}s^{3/4}$. This model may also be generalized to dampers alternating with N-1 level ladders, producing $\hat{g}(s) \approx E^{1/2^{N+1}}\eta^{1-1/2^{N+1}}s^{1-1/2^{N+1}}$. By performing an inverse Laplace transform, fractional derivative stress-strain relationships of order $1/2, 3/4, 7/8, \ldots$ are generated. As the depth of the ladder increases $(N\to\infty)$, $\hat{g}(s)\to \eta s$, yielding a purely viscous response.

In order to generate fractional derivatives of all orders within the unit interval, fractal ladders containing alternating damper-M ladders and N ladder-spring networks are denoted as an (M,N) ladder. The (M,N) ladder is constructed by replacing the dampers in Fig. 3 with M-level damper-ladders and the springs in Fig. 3 with N-level spring ladders. Using this notation, the recursive ladder shown in Fig. 4 is denoted as (0,1), while the recursive ladder shown in Fig. 5 is denoted as (1,0). Computing the frequency-domain modulus for an (M,N) network yields

$$\hat{g}(s) \approx E^{1/2(1+(1/2^M)-(1/2^N))}\eta^{1/2(1-(1/2^M)+(1/2^N))} \times s^{1/2(1-(1/2^M)+(1/2^N))}. \tag{B1}$$

Letting $\beta=1/2(1-1/2^M+1/2^N)$ and performing an inverse Laplace transform yields Eq. (9).

[1] S. A. Goss, L. A. Frizzell, and F. Dunn, "Ultrasonic absorption and attenuation in mammalian tissues," Ultrasound Med. Biol. **5**, 181–186 (1979).
[2] 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).
[3] P. He, "Experimental verification of models for determining dispersion from attenuation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 706–714 (1999).
[4] F. A. Duck, *Physical Properties of Tissue*, 1st ed. (Academic, London, 1990), pp. 99–124.
[5] P. A. Narayana and J. Ophir, "On the frequency dependence of attenuation in normal and fatty liver," IEEE Trans. Sonics Ultrason. **SU-30**, 379–383 (1983).
[6] T. Lin, J. Ophir, and G. Potter, "Frequency-dependent ultrasonic differentiation of normal and diffusely diseased liver," J. Acoust. Soc. Am. **82**, 1131–1138 (1987).
[7] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Pres, Princeton, NJ, 1968), pp. 270–300.
[8] M. A. Biot, "Theory of propagation of elastic waves in fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[9] 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).

[10]A. C. Kak and K. A. Dines, "Signal-processing of broad-band pulsed ultrasound—Measurement of attenuation of soft biological tissues," IEEE Trans. Biomed. Eng. **25**, 321–344 (1978).
[11]K. V. Gurumurthy and R. M. Arthur, "A dispersive model for the propagation of ultrasound in soft tissue," Ultrason. Imaging **4**, 355–377 (1982).
[12]T. L. Szabo, "The material impulse response for broadband pulses in lossy media," in Proceedings of the IEEE Ultrasonics Symposium, Honolulu, HI (2003), pp. 748–751.
[13]P. He, "Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 114–125 (1998).
[14]M. G. Wismer and R. Ludwig, "An explicit numerical time domain formulation to simulate pulsed pressure waves in viscous fluids exhibiting arbitrary frequency power law attenuation," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1040–1049 (1995).
[15]R. S. C. Cobbold, N. V. Sushilov, and A. C. Weathermon, "Transient propagation in media with classical or power-law loss," J. Acoust. Soc. Am. **116**, 3294–3303 (2004).
[16]S. Leeman, "Ultrasound pulse propagation in dispersive media," Ultrasound Med. Biol. **25**, 481–488 (1980).
[17]J. M. Blackledge and S. Leeman, "Green's functions for acoustic fields in dispersive media," J. Phys. D: Appl. Phys. **16**, L247–L250 (1983).
[18]T. L. Szabo, "Time-domain wave-equations for lossy media obeying a frequency power-law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[19]A. Hanyga and M. Seredyńska, "Power-law attenuation in acoustic and isotropic anelastic media," Geophys. J. Int. **155**, 830–838 (2003).
[20]W. Chen and S. Holm, "Fractional Laplacian time-space models for linear and nonlinear lossy media exhibiting arbitrary frequency power-law dependency," J. Acoust. Soc. Am. **115**, 1424–1430 (2004).
[21]R. Schumer, D. A. Benson, M. M. Meerschaert, and S. W. Wheatcraft, "Eulerian derivation of the fractional advection-dispersion equation," J. Contam. Hydrol. **48**, 69–88 (2001).
[22]M. Caputo, "Linear models of dissipation whose *Q* is almost frequency independent II," Geophys. J. R. Astron. Soc. **13**, 529–539 (1967).
[23]M. G. Wismer, "Finite element analysis of broadband acoustic pulses through inhomogeneous media with power law attenuation," J. Acoust. Soc. Am. **120**, 3493–3502 (2006).
[24]M. Ochmann and S. Makarov, "Representation of the absorption of nonlinear waves by fractional derivatives," J. Acoust. Soc. Am. **94**, 3392–3399 (1993).
[25]Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: A time-domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).
[26]T. L. Szabo and J. Wu, "A model for longitudinal and shear wave propagation in viscoelastic media," J. Acoust. Soc. Am. **107**, 2437–2446 (2000).
[27]M. Giona and H. E. Roman, "Fractional diffusion equation on fractals: One-dimensional case and asymptotic behavior," J. Phys. A **25**, 2093–2105 (1992).
[28]R. Metzler, W. G. Glöckle, and T. F. Nonnenmacher, "Fractional model equation for anomalous diffusion," Physica A **211**, 13–24 (1994).
[29]W. G. Glöckle and T. F. Nonnenmacher, "Fox function representation of non-Debye relaxation processes," J. Stat. Phys. **71**, 741–756 (1993).
[30]H. Schiessel, R. Metzler, A. Blumen, and T. F. Nonnenmacher, "Generalized viscoelastic models: Their fractional equations with solutions," J. Phys. A **28**, 6567–6584 (1995).
[31]R. L. Bagley and P. J. Torvik, "A theoretical basis for the application of fractional calculus to viscoelasticity," J. Rheol. **27**, 201–210 (1983).
[32]J. Wu and C. Layman, "Wave equations, dispersion relations, and van Hove singularities for applications of doublet mechanics to ultrasound propagation in bio- and nanomaterials," J. Acoust. Soc. Am. **115**, 893–900 (2004).
[33]N. W. Tschoegl, *The Phenomenological Theory of Linear Viscoelastic Behavior: An Introduction* (Springer-Verlag, Berlin, 1989).
[34]C. T. Lim, E. H. Zhou, and S. T. Quek, "Mechanical models for living cells A review," J. Biomech. **39**, 195–216 (2006).
[35]A. Yeung and E. Evans, "Cortical shell-liquid core model for passive flow of liquid-like spherical cells into micropipets," Biophys. J. **56**, 139–149 (1989).
[36]E. M. Darling, M. Topel, S. Zauscher, T. P. Vail, and F. Guilak, "Viscoelastic properties of human mesenchymally-derived stem cells and primary osteoblasts, chondrocytes, and adipocytes," J. Biomech. **41**, 454–464 (2008).
[37]J. D. Humphrey, "Continuum biomechanics of soft biological tissues," Proc. R. Soc. London, Ser. A **459**, 3–46 (2003).
[38]C. Verdier, "Rheological properties of living materials. From cells to tissues," J. Theoretical Medicine **5**, 67–91 (2003).
[39]B. Gross and R. M. Fuoss, "Ladder structures for representation of viscoelastic systems," J. Polym. Sci. **19**, 39–50 (1956).
[40]H. Schiessel and A. Blumen, "Hierarchical analogues to fractional relaxation equations," J. Phys. A **26**, 5057–5069 (1993).
[41]H. Schiessel and A. Blumen, "Mesoscopic pictures of the sol-gel transition: Ladder models and fractal networks," Macromolecules **28**, 4013–4019 (1995).
[42]N. Heymans and J.-C. Bauwens, "Fractal rheological models and fractional differential equations for viscoelastic behavior," Rheol. Acta **33**, 210–219 (1994).
[43]B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 4th ed. (Garland Science, New York, 2002).
[44]B. Gross, "Ladder structures for representation of viscoelastic systems. II," J. Polym. Sci. **20**, 123–131 (1956).
[45]R. C. Koeller, "Applications of fractional calculus to the theory of viscoelasticity," J. Appl. Mech. **51**, 299–307 (1984).
[46]P. E. Rouse, "A theory of linear viscoelastic properties of dilute solutions of coiling polymers," J. Chem. Phys. **21**, 1272–1280 (1953).
[47]A. Kreis and A. C. Pipkin, "Viscoelastic pulse propagation and stable probability distributions," Q. Appl. Math. **44**, 353–360 (1986).
[48]M. J. Buckingham, "Causality, Stokes' wave equation, and acoustic pulse propagation in a viscous fluid," Phys. Rev. E **72**, 026610 (2005).
[49]T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power-law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[50]J. F. Kelly, M. M. Meerschaert, and R. J. McGough, "Analytical time-domain Green's functions for power-law media," J. Acoust. Soc. Am. **124**, 2861–2872 (2008).
[51]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 4th ed. (Wiley, New York, 2000), pp. 210–245.
[52]K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers-Krönig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[53]E. R. Weibel, "Fractal geometry: A design principle for living organisms," Am. J. Physiol. Lung Cell. Mol. Physiol. **261**, L361–L369 (1991).
[54]J. W. Baish and R. K. Jain, "Fractals and cancer," Cancer Res. **60**, 3683–3688 (2000).
[55]D. Dhar, "Lattices of effectively nonintegral dimensionality," J. Math. Phys. **18**, 577–585 (1977).
[56]S. Alexander and R. Orbach, "Density of states on fractals: Fractons," J. Phys. (Paris) **43**, L625–L631 (1982).
[57]G. Wojcik, J. Mould, F. Lizzi, N. Abboud, M. Ostromogilsky, and D. Vaughn, "Nonlinear modeling of therapeutic ultrasound," in Proceedings of the IEEE Ultrasonics Symposium, Cannes, France (1995), pp. 1617–1622.
[58]A. A. Kilbas, H. M. Srivastava, and J. J. Truhillo, *Theory and Applications of Fractional Differential Equations* (Elsevier, Amsterdam, 2006).

Why do Chinese alligators (*Alligator sinensis*) form bellowing choruses: A playback approach

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

Ding Wang[a)]
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Song Zhang, Chaolin Wang, and Renping Wang
Anhui Research Center for the Chinese Alligator Reproduction, Xuancheng 242034, People's Republic of China

Xiaobing Wu
College of Life Sciences, Anhui Normal University, Wuhu 241000, People's Republic of China

(Received 17 September 2008; revised 8 February 2009; accepted 13 July 2009)

Crocodilians are quite vocal relative to other reptile groups, and the alligators are among the most vocal of the crocodilians. The Chinese alligator, *Alligator sinensis*, is usually solitary but engages in bellowing choruses in certain waters during the mating season. This paper reports the organization of Chinese alligator's bellowing choruses based upon field observations and playback experiments. Alligators of both genders engaged in the choruses, remaining immobile throughout and inclining toward bellowing synchronously (i.e., starting and finishing at about the same time). The choruses lasted about 10 min with abrupt onset and offset. Moreover, playback experiments revealed that both male and female alligators responded equally to bellowing stimuli from the same and opposite sexes and that none of the tested alligators approached the loudspeaker in spite of playback of male or female stimuli. These suggest that Chinese alligators may not bellow to compete for or attract mates during the choruses. Instead, when their ecological behaviors, namely, dispersed inhabitation, multi-copulation, restricted mating season, etc., are considered, we hypothesize that they may synchronize bellows to enhance group detectability for assembling individuals into certain waters for subsequent copulations.
 [DOI: 10.1121/1.3203667]

PACS number(s): 43.80.Ka [MCH] Pages: 2082–2087

I. INTRODUCTION

An efficient system of signals and responses is important for successful reproduction (Patterson, 1985). The role of acoustic signals in reproduction has been well investigated in insects (Minckley *et al.*, 1995; Snedden, 1996; Snedden and Greenfield, 1998), amphibians (Wells, 1988; Schwartz, 1993, 2001; Cocroft and Ryan, 1995; Kime *et al.*, 2004), birds (Morse, 1989; Lind *et al.*, 1996; Liu *et al.*, 1998), and mammals (Southall *et al.*, 2003; Notman and Rendall, 2005; Deecke *et al.*, 2005). Diverse vocal behaviors and acoustic tactics are employed in different species' reproduction patterns. Many studies of anurans, for example, have shown that overlapping calls are less attractive to females than non-overlapping calls (Schwartz and Wells, 1983, 1984; Wells and Schwartz, 1984), especially if the overlap obscures fine temporal properties of the call (Schwartz, 1987). Males may respond to the calls of their neighbors in ways that help them maintain or increase their relative attractiveness. For instance, they may elevate the rate, complexity, duration, and intensity of their call, and may also shift the timing of their advertisement calls or call elements to reduce the potential for acoustic interference or to increase the likelihood that their signals will lead rather than follow those of their neighbors (reviewed in Schwartz and Buchanan, 2002).

Reptiles, in general, are relatively nonvocal, and consequently their acoustic signals and vocal behaviors have received relatively little attention (Marcellini, 1977; Vliet, 1989; Young, 1991). Some reptile groups, especially the crocodilian species, produce striking vocalizations during the mating season (Zhu, 1957; Campbell, 1973; Gans and Maderson, 1973; Herzog and Burghardt, 1977; Chen *et al.*, 1985, 2003; Vliet, 1989; Wang *et al.*, 2006, 2007), but there are very few reports that address the precise roles of these acoustic signals in successful reproduction.

The Chinese alligator, *Alligator sinensis*, is one of the most endangered of the 23 crocodilian species in the world (Thorbjarnarson, 1992). Wild individuals currently number less than 150 and are restricted to a small area in southern Anhui Province and perhaps in adjacent Zhejiang and Jiangsu Provinces of China (Ding *et al.*, 2001; Chen *et al.*,

[a)]Author to whom correspondence should be addressed. Electronic mail: wangd@ihb.ac.cn

TABLE I. Characteristic parameters of 26 Chinese alligators (13 males and 13 females) used in playback experiments.

		Mean	Std. deviation	Minimum	Maximum
Male	BL (cm)	164	6.85	151	175
	Snout-vent length (cm)	78.9	3.59	74.0	86.0
	Weight (kg)	15.5	2.12	11.7	19.0
Female	BL (cm)	158	8.9	146	177
	Snout-vent length (cm)	76.8	4.20	70.0	86.0
	Weight (kg)	13.6	2.59	10.5	19.3

2003). Chinese alligators are usually solitary and territorial but move into specific ponds, lakes, swamps, and marshes for copulation during mating season from late May through middle June (Chen *et al.*, 1985, 2003; Wang *et al.*, 2006, 2007). Previous investigations have shown that vocal communication plays a vital role in social interactions and reproduction of Chinese alligators (Zhu, 1957; Chen *et al.*, 1985, 2003; Wu and Wang, 2004; Wang *et al.*, 2006). Though both male and female alligators bellow throughout the active season, these vocalizations are greatly heightened during the mating season (Wu and Wang 2004; Wang *et al.*, 2006). During the mating period, the bellows of one alligator stimulate neighboring others to bellow, eventually forming a chorus (Wang *et al.*, 2007). The choruses are believed to be highly related to mating, but the mechanisms are still unknown.

Because both male and female Chinese alligators participate in these bellowing choruses and it is difficult to differentiate males from females by appearance alone, or by different vocal behaviors as American alligator (*Alligator mississippiensis*) has, clear observations seem rarely possible. The present report describes the first experiments to address this issue by playing back male and female bellowing stimuli to both male and female alligators individually to determine the mechanism involved in the bellowing chorus.

II. METHODS

A. Study site and subjects

Experiments were carried out during the mating season of Chinese alligator in 2006 and 2008 within Anhui Research Center for Chinese Alligator Reproduction (ARCCAR), located in Xuancheng City in Southern Anhui Province, China. The study site is a reproductive enclosure (30°54′N, 118°46′E) that primarily consists of four permanent ponds. Most individuals of the reproductive cohort are wild-caught, although a few are first- and second-generation offsprings from artificial reproduction [see Zhang *et al.* (2005) and Wang *et al.* (2006) for more details]. The focal field observation pond was approximately 2000 m^2 with a 400 m^2 island densely covered by shrub situated in the center.

During the study, a total of 26 Chinese alligators (13 males and 13 females) were harmlessly captured and transported to a $10\times6\times1$ m^3 outdoor arena individually. All captured alligators were sexually mature (snout-head rubbing, mounting behaviors, and bellowing activities had been observed before capture). The gender of each alligator was confirmed by touching the cloaca. Immediately after capture, the body length (BL) and snout-vent length of each alligator were measured, and the weight was calculated based on the following formula: $\text{weight}=6.7-0.1713\text{BL}+0.001\,189\text{BL}^2+0.000\,001\,1\text{BL}^3$, provided by Chen *et al.* (2003) (Table I).

B. Field observations

To avoid disturbing the animals, field observations were performed behind a brick boundary wall (1.5 m high) using the naked eye and/or Fujinon 7×50 binoculars at distances ranging from 2 to 30 m. We collected information on behavioral context, number of bellows in each bellowing sequence [i.e., a succession of bellows produced in a regularly repeated unit (Garrick *et al.*, 1978)], total number of sequences, and responses from neighboring animals to focal bellowers using the "target objective observation" method (Zhang *et al.*, 2005) during choruses. We also noted individuals' activities after choruses.

C. Playback experiments

The signals used for playback experiments were recorded from 19 male and 23 female alligators in captivity within ARCCAR using an omnidirectional Fidek KSM-81 microphone (frequency response: 30–20 000 Hz, +5.5/−11.4 dB, Japan) connected to a Sony TCD-D100 digital audio tape recorder (sampling frequency: 48 kHz, Japan) and Sony digital audio tape. All vocalizations were recorded at a distance of approximately 3 m. The recordings were replayed from the recorder and were then transformed into, and saved as, sound files (in wav format) on a computer by PC-based signal processing software, COOL EDIT PRO 2.0, with a 22 kHz sampling frequency and 16-bit. The software was also used to edit and playback sound files during playback experiments. We chose signals with high signal-to-noise ratio and representative of the population mean as playback stimuli.

After each alligator was transported to the arena, it was given more than 48 h to adapt to its new surroundings before testing began. Some fish were provided according to the feeding arrangement of ARCCAR during the playback period. The tested alligator was released at the same place where it was captured when tests were completed, and another alligator was brought into the arena. Because Chinese alligator has two pairs of musk glands that are believed to release substances that influence social interaction (Chen *et al.*, 1985, 2003), the arena was always washed before a new alligator was introduced to it. As bellowing activities mostly occurred in shallow water (Wang *et al.*, 2006), water depth in the arena during playback experiments was controlled and ranged from 11.6 to 19.0 cm according to the BL

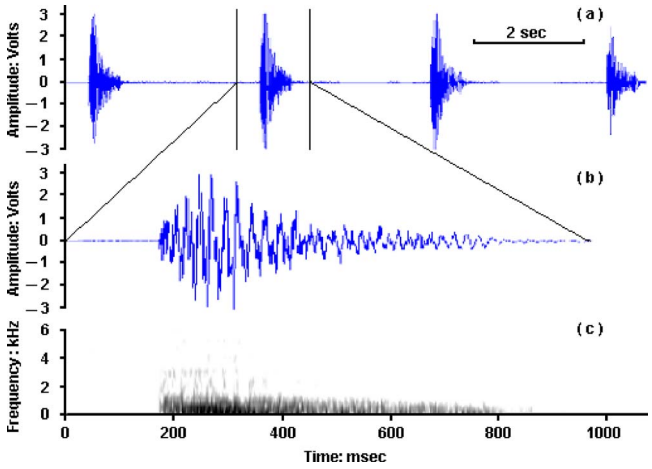


FIG. 1. (Color online) Wave form and spectrogram of bellows of a male Chinese alligator. (a) Wave form of a sequence of four bellows used in playback experiments. Panels (b) and (c) show the wave form and spectrogram, respectively, of the second bellow of the sequence.

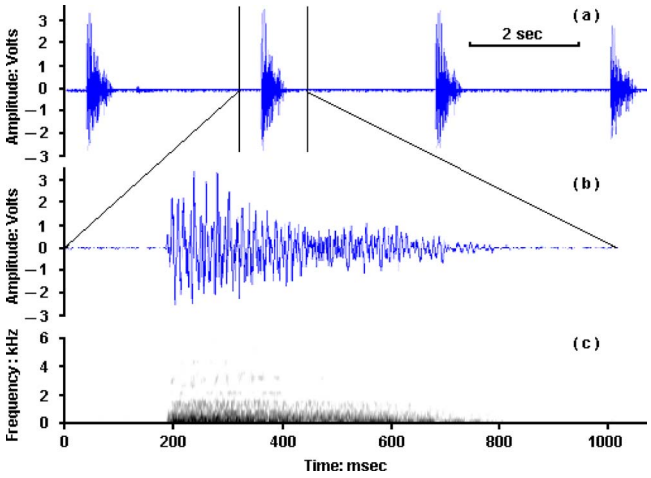


FIG. 2. (Color online) Wave form and spectrogram of bellows of a female Chinese alligator. (a) Wave form of a sequence of four bellows used in playback experiments. Panels (b) and (c) show the wave form and spectrogram, respectively, of the second bellow of the sequence.

of each alligator, which may facilitate the "head oblique and tail arched" bellowing posture (see Wang *et al.*, 2007). Both air and water temperatures were measured during the playback period (air temperature: 22.0–32.0 °C; water temperature: 22.4–32.1 °C).

Playback experiments were conducted between 0600 and 0700 h, a period corresponding to the time of natural bellowing activity of Chinese alligators (Wang *et al.*, 2006) during clam and dry weather conditions to avoid sound propagation problems due to wind. Stimuli were broadcast from an IBM computer using the COOL EDIT PRO 2.0 software over a Shock wave SC-2105 amplified speaker (effective range 25–20 000 Hz), which was mounted on the arena wall, 1 m above the floor. Peak sound pressure levels (SPLs, in dB) of playback stimuli were verified at 90 dB (re 20 μPa), measured 5 m from the loudspeaker using a hand-held Realistic sound level meter (type DT-805, re 20 μPa, C-weighting, fast response). This level corresponded to that of natural bellows at that distance (Wang *et al.*, 2007).

After the equipment were installed and verified, more than 30 min was spent on acclimation before playback tests commence at about 0600 every morning. Four male (or female) bellows were broadcast with a silent interval of 2.5 s [i.e., a bellowing sequence, Fig. 1(a) and 2(a)], a rhythm that corresponds to that of natural bellows' emission. Then, after at least a 5-min period of silence, another bellowing sequence was broadcast. Each alligator was presented with three male sequences and three female sequences. The order of sequence presentation was randomized to avoid habituation. Totally, 78 male and 78 female bellowing sequences were edited and played back, each broadcast only once.

D. Criteria of responses and data analysis

On the basis of natural observations, responses of alligators to bellows of neighboring individuals were ranked as follows: 0 (none), no reaction to the stimulus (tagged as "no reaction"); 1 (weak), presented a head oblique and tail arched posture but did not bellow (tagged as "HOTA"); and 2 (strong), responded to the stimulus by bellowing with a HOTA posture (tagged as "bellowing"). If the emitted stimuli induced no change in the behavior of the focal animal within 5 min of being presented, we scored it as a no reaction. We monitored each alligator's response and timed response latency (i.e., time between stimulus onset and the alligator's response) in every playback trial behind the arena wall.

Statistical analysis was done using the mean response per subject as the sample points; thus, the sample size is the number of subjects. The two-tailed Wilcoxon paired-sample signed rank test was used to compare the response intensities of male and female alligators to the same- and opposite-sex stimuli and the differences in the latency of responses to male and female stimuli. Statistical significance was set at $P<0.05$. All statistical calculations were carried out using SPSSs 13.0.

III. RESULTS

A. General description of bellowing choruses

Choruses were usually initiated by vocalizations from one alligator (other times by meteorological thunders) and then intensified as neighboring individuals joined in. Alligators tended to bellow synchronously (i.e., starting and finishing at about the same time), and choruses were characterized by abrupt onsets and offsets. Choruses lasted for a period of 9.99 ± 2.51 min (means $\pm$ SD; $N=103$). During the mating season, the inter-chorus silence was 5.33 ± 3.94 h (means $\pm$ SD; $N=123$). Alligators remained immobile during choruses, but vigorous swimming, snout-head rubbing, and mounting behaviors were frequently observed afterwards.

B. Both male and female alligators responded equally to the same- and opposite-sex stimuli

Playback tests showed that male alligators responded equally to the same- and opposite-sex stimuli ($Z=-0.597$, $N=13$, and $P=0.551$) [Fig. 3(a)]. Male stimuli evoked 79.5% bellowing, 12.8% HOTA, and 7.7% no reaction. Female stimuli evoked 76.9% bellowing, 12.8% HOTA, and 10.3% no reaction. Response latencies to male and female stimuli

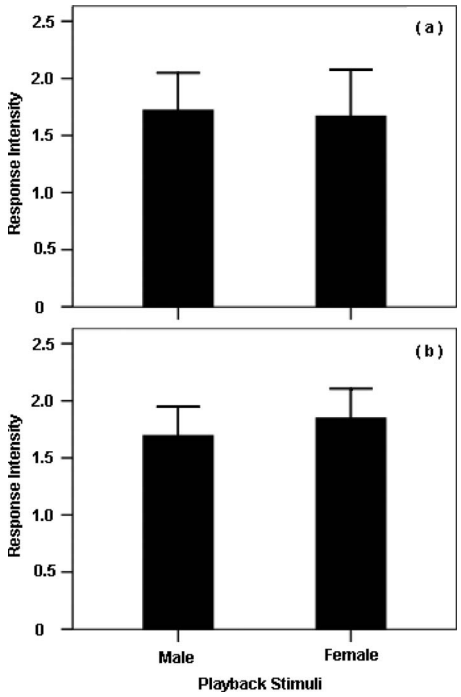


FIG. 3. Response intensities of tested alligators to male and female bellowing stimuli. (a) Male and (b) female alligators responded equally well to the same- and opposite-sex bellowing stimuli. Bars show means+SD and $N=13$.

were not significantly different ($Z=-0.490$, $N=13$, and $P=0.624$) [Fig. 4(a)]. Female alligators also responded equally well to male and female stimuli ($Z=-1.51$, $N=13$, and $P=0.131$) [Fig. 3(b)], with 76.9% bellowing, 15.4% HOTA, and 7.7% no reaction to male stimuli and 87.2% bellowing, 10.3% HOTA, and 2.5% no reaction to female stimuli. Response latencies to male and female stimuli were also not significantly different ($Z=-1.58$, $N=13$, and $P=0.114$) [Fig. 4(b)]. None of the tested alligators approached the loudspeaker during any trial in spite of playback of male or female stimuli.

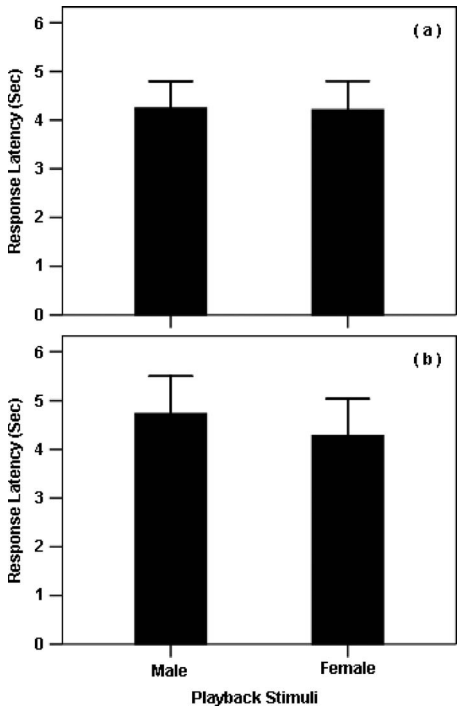


FIG. 4. Response latencies of tested alligators to male and female bellowing stimuli. There is no significant difference in the response latencies of (a) male and (b) female alligators to the same- and opposite-sex bellowing stimuli. Bars show means+SD and $N=13$.

IV. DISCUSSION

In numerous densely congregative insects and anuran species, competitive interactions between males are the best explanation for the emergence of synchronized vocalizations (Greenfield and Roizen, 1993; Greenfield, 1994; Greenfield *et al.*, 1997). During choruses, males advertise vocally to attract females for mating (Minckley *et al.*, 1995; Schwartz and Buchanan, 2002). Females, on the other hand, are in general voiceless and move among the males to choose appropriate mates on the basis of their advertisement calls (Morris, 1991; Howard and Young, 1998; Wollerman and Wiley, 2002; Smith and Roberts, 2003).

Unlike many insect and anuran species, however, the following reasons indicate that Chinese alligators may not bellow to compete for or attract mates during choruses: (i) both male and female Chinese alligators engage in bellowing choruses and bellow synchronously; (ii) attending alligators remain stationary throughout; and (iii) the choruses last only about 10 min with abrupt onset and offset, with an average of 14.3 ± 6.40 bellows per alligator during each chorus (Wang *et al.*, 2007). Male and female alligators bellow synchronously and remain stationary throughout the chorus, suggesting that male bellows may not be attractive to neighboring females; otherwise, females would approach bellowing males (and vice versa). In addition, unlike those of insects and anurans, which last several hours or even throughout an entire night (Brenowitz and Rose, 1999; Grafe, 1999), a 10-min alligator chorus might not be sufficient for every individual to find an appropriate mate. No alligator bellowed alone after the end of the choruses. Our deduction is further supported by studies by Chen *et al.* (1985, 2003) and Wang *et al.* (2006), in which they found that bellowing activities mostly occur during 0600–0700 and 1100–1200 h, which is more than 12 h ahead of the peak time of copulation (around midnight).

Bellowing activity involves a significant energy cost (Chen *et al.*, 1985, 2003). The bellows have an average SPL of 90.8 dB (re 20 μPa) at 5 m distance. Chinese alligators may not have enough energy to bellow as long as insects and anurans do, so the chorus lasts just 10 min and the total number of bellows per alligator is 14.3 ± 6.40 during a chorus (Wang *et al.*, 2007). This might explain why the Chinese alligator does not bellow to compete for or attract mates.

The deduction is also strengthened by results from the playback experiments that both male and female alligators respond equally to same- and opposite-sex bellowing stimuli without ambiguity and that none of the tested alligators approach the loudspeaker in spite of playback of male or female stimuli, which indicate that bellowing choruses may have other biological functions rather than as signal to compete for or attract mates. Besides, for attracting mates, the

following hypotheses have also been proposed to explain why certain animals' synchronous chorusing might be adaptive, none but the last one of which seems reasonable for Chinese alligators: Species synchronize calling (i) to establish dominance hierarchy, (ii) to confuse predators, and (iii) to enhance group detectability (i.e., by maximizing the peak amplitude of group signaling) (reviewed in Grafe, 1999).

Studies of the American alligator show that bellowing activity provides information about the body size and social status of the active bellower (Garrick and Lang, 1977; Garrick *et al.*, 1978). Hierarchy is obvious in some crocodilian species, such as Nile crocodile (*Crocodylus niloticus*), Indian gavial (*Gaviails gangeticus*), and Estuarine crocodile (*Crocodylus porosus*) (reviewed in Chen *et al.*, 2003). Thus, one might suggest the possibility that Chinese alligators bellow to increase their dominance status, with large, aggressive individuals controlling access to mates and resources. In reality, however, Chinese alligators live in separate territories throughout most time of the year, and dominance hierarchies do not seem to exist (Chen *et al.*, 1985, 2003). Hence, this hypothesized function for Chinese alligator bellowing seems implausible.

Study on neotropical tree frog (*Smilisca sila*) shows that overlapping calls attract fewer predatory bats than alternating calls (Tuttle and Ryan, 1982). Yet, Chinese alligator is at the top of the local food chain—no predator is known to attack adult Chinese alligator (Chen *et al.*, 1985, 2003). So, it is hard to say that Chinese alligators synchronize bellowing to reduce predation.

Chinese alligator lives in temperate climates, which force mating to occur within a restricted time frame (Chen *et al.*, 1985, 2003). Male alligator mates with more than one female and female alligator can produce clutches fathered by more than one male (Chen *et al.*, 1985, 2003; Davis *et al.*, 2001). This means that the females can carry offspring of various genetic compositions, which help the alligators tolerate and thrive in new environments and habitats (Davis *et al.*, 2001). Consequently, with widely distributed individuals congregating into a small number of waters, it becomes quite cost-efficient for alligators to copulate with more mates in a limited time frame. Meanwhile, an efficient communication signal is required for them to locate certain aggregative waters in the dense vegetative habitat.

Compared to visual and olfactory signals, auditory signal is an effective way for Chinese alligators to locate certain aggregative waters. Dense vegetation restricts visual communication, and efficacy of olfactory signaling is low (Chen *et al.*, 2003; Wang *et al.*, 2007). The bellow, characterized by low dominant frequency and high SPL, which facilitate propagation in the high densely vegetated environment (Wang *et al.*, 2007), can serve as an efficient beacon that allows widely dispersed alligators of both genders to locate a certain breeding aggregation. The synchronous bellows of many alligators help to enhance group detectability over a further spatial range and also provide information about the number of attending alligators, which would contribute to a larger number of copulations. These may explain why both male and female alligators respond equally to the same- and opposite-sex bellowing stimuli to form bellowing choruses and why there is a 12-h interval between peak times of bellowing and copulations. The field observation of Wang *et al.* (2006) finds that bellowing chorus does have congregative effect during the mating season.

As the current wild alligators are mainly scattered in the 43 300 h m^2 national reserve, it results in a lack of communication between individuals and, in turn, a lack of reproductive success (Ding *et al.*, 2001). Although the tested alligators are living in semi-natural condition (though once wild-caught), their higher density may lead to a population with reproductive behaviors that more closely reflect the wild alligators prior to reaching the endangered situation of the present. This study would be useful in the reintroduction of artificially reproduced alligators back to the wild. But further works, such as those on Chinese alligator's hearing capability, active space of the bellows, and alligator's responses at different distances from the source to single and chorused bellows, are required to directly test the hypothesis about enhancing group detectability.

ACKNOWLEDGMENTS

The authors are extremely grateful to Hongxing Zhu, Tongsheng Xia, Jialong Zhu, Xuesong Zhang, Jiangjiang Zhou, Yongkang Zhou, and Yijiang Ou of ARCCAR, Songhai Li and Kexiong Wang of Research Group on Conservation Biology of Aquatic Animals of the Institute of Hydrobiology, Chinese Academy of Sciences, and Baoshan Xia of College of Life Sciences, Anhui Normal University for providing helpful suggestions in conducting playback experiments. The authors would also like to express appreciation to the two reviewers for their valuable comments on this manuscript. Financial support was provided by National Basic Research Program of China (Grant No. 2007CB411600) and National Natural Science Foundation of China (Grant No. 30730018).

Brenowitz, E. A., and Rose, G. J. (**1999**). "Female choice and plasticity of male calling behaviour in the Pacific treefrog," Anim. Behav. **57**, 1337–1342.

Campbell, H. H. (**1973**). "Observations on the acoustic behavior of crocodilians," Zoologica (N.Y.) **58**, 1–11.

Chen, B. H., Hua, T. M., Wu, X. B., and Wang, C. L. (**2003**). *Research on the Chinese Alligator* (Shanghai Scientific and Technological Education, Shanghai), pp. 18–252.

Chen, B. H., Hua, Z. H., and Li, B. H. (**1985**). *Alligator sinensis* (Anhui Technology, Hefei), pp. 115–216.

Cocroft, R. B., and Ryan, M. J. (**1995**). "Patterns of advertisement call evolution in toads and frogs," Anim. Behav. **49**, 283–303.

Davis, L. M., Glenn, T. C., Elsey, R. M., Dessauer, H. C., and Sawyer, R. H. (**2001**). "Multiple paternity and mating patterns in the American alligator *Alligator mississippiensis*," Mol. Ecol. **10**, 1011–1024.

Ding, Y. Z., Wang, X. M., He, L. J., Shao, M., Xie, W. S., Thorbjarnarson, J., and McMurry, T. S. (**2001**). "Study on the current population and habitat of the wild Chinese alligator (*Alligator sinensis*)," Biodiversity Sci. **9**, 102–108.

Gans, C., and Maderson, P. (**1973**). "Sound production mechanisms in recent reptiles: Review and comments," Am. Zool. **13**, 1195–1203.

Garrick, L. D., and Lang, J. W. (**1977**). "Social signals and behavior of adult alligators and crocodiles," Am. Zool. **17**, 225–239.

Garrick, L. D., Lang, J. W., and Herzog, H. A. (**1978**). "Social signals of adult American alligators," B. Am. Mus. Nat. Hist. **160**, 153–192.

Grafe, T. U. (**1999**). "A function of synchronous chorusing and a novel female preference shift in an anuran," Proc. R. Soc. London, Ser. B **266**, 2331–2336.

Greenfield, M. D. (**1994**). "Cooperation and conflict in the evolution of signal interactions," Annu. Rev. Ecol. Syst. **25**, 97–126.
Greenfield, M. D., and Roizen, I. (**1993**). "Katydid synchronous chorusing is an evolutionarily stable outcome of female choice," Nature (London) **364**, 618–620.
Greenfield, M. D., Tourtellot, M. K., and Snedden, W. A. (**1997**). "Precedence effects and the evolution of chorusing," Proc. R. Soc. London, Ser. B **264**, 1355–1361.
Herzog, H. A., and Burghardt, G. M. (**1977**). "Vocalization in juvenile crocodilians," Z. Tierpsychol. **44**, 294–304.
Howard, R. D., and Young, J. R. (**1998**). "Individual variation in male vocal traits and female mating preferences in *Bufo americanus*," Anim. Behav. **55**, 1165–1179.
Kime, N. M., Burmeister, S. S., and Ryan, M. J. (**2004**). "Female preferences for socially variable call characters in the cricket frog, *Acriscrepitans*," Anim. Behav. **68**, 1391–1399.
Lind, H., Dabelsteen, T., and McGregor, P. K. (**1996**). "Female great tits can identify mates by song," Anim. Behav. **52**, 667–671.
Liu, R. S., Yu, Q., Lei, F. M., Ding, W. N., and Zhao, X. R. (**1998**). *Research of Bird's Singing* (Science, Beijing), pp. 53–57.
Marcellini, D. (**1977**). "Acoustic and visual display behavior of Gekkonid lizards," Am. Zool. **17**, 251–260.
Minckley, R. L., Greenfield, M. D., and Tourtellot, M. K. (**1995**). "Chorus structure in tarbush grasshoppers: Inhibition, selective phonoresponse, and signal competition," Anim. Behav. **50**, 579–594.
Morris, M. R. (**1991**). "Female choice of large males in the treefrog *Hyla ebraccata*," J. Zool. (Lond.) **223**, 371–378.
Morse, D. H. (**1989**). "Song patterns of warblers at dawn and dusk," Wilson Bull. **101**, 26–35.
Notman, H., and Rendall, D. (**2005**). "Contextual variation in chimpanzee pant hoots and its implications for referential communication," Anim. Behav. **70**, 177–190.
Patterson, H. E. H. (**1985**). "The recognition concept of species," in *Species and Speciation*, edited by E. S. Vrba (Transvaal Museum, Pretoria), Monograph No. 4, pp. 21–29.
Smith, M. J., and Roberts, J. D. (**2003**). "Call structure may affect male mating success in the quacking frog *Crinia georgiana* (Anura: Myobatrachidae)," Behav. Ecol. Sociobiol. **53**, 221–226.
Schwartz, J. J. (**1987**). "The function of call alternation in anuran amphibians: A test of three hypotheses," Evolution (Lawrence, Kans.) **41**, 461–471.
Schwartz, J. J. (**1993**). "Male calling behavior, female discrimination and acoustic interference in the neotropical treefrog *Hyla microcephala* under realistic acoustic conditions," Behav. Ecol. Sociobiol. **32**, 401–414.
Schwartz, J. J. (**2001**). "Call monitoring and interactive playback systems in the study of acoustic interactions among male anurans," in *Anuran Communication*, edited by M. J. Ryan (Smithsonian Institution, Washington, DC), pp. 183–204.
Schwartz, J. J., and Buchanan, B. W. (**2002**). "Acoustic interactions among male gray treefrogs, *Hyla versicolor*, in a chorus setting," Behav. Ecol. Sociobiol. **53**, 9–19.
Schwartz, J. J., and Wells, K. D. (**1983**). "An experimental study of acoustic interference between two species of neotropical treefrogs," Anim. Behav. **31**, 181–190.
Schwartz, J. J., and Wells, K. D. (**1984**). "Interspecific acoustic interactions of the neotropical treefrog *Hyla ebraccata*," Behav. Ecol. Sociobiol. **14**, 211–224.
Snedden, W. A. (**1996**). "Lifetime mating success in male sagebrush crickets: Sexual selection constrained by a virgin male mating advantage," Anim. Behav. **51**, 1119–1125.
Snedden, W. A., and Greenfield, M. D. (**1998**). "Females prefer leading males: Relative call timing and sexual selection in katydid chorus," Anim. Behav. **56**, 1091–1098.
Southall, B. L., Schusterman, R. J., and Kastak, D. (**2003**). "Acoustic communication ranges for northern elephant seals (*Mirounga angustirostris*)," Aquat. Mamm. **29**, 202–213.
Thorbjarnarson, J. (**1992**). *Crocodiles: An Action Plan for Their Conservation* (IUCN, Gland, Switzerland), pp. 38–39.
Tuttle, M. D., and Ryan, M. J. (**1982**). "The role of synchronized calling, ambient light, and ambient noise, in anti-bat-predator behavior of a treefrog," Behav. Ecol. Sociobiol. **11**, 125–131.
Vliet, K. A. (**1989**). "Social displays of the American alligator (*Alligator mississippiensis*)," Am. Zool. **29**, 1019–1031.
Deecke, V. B., Ford, J. K. B., and Slater, P. J. B. (**2005**). "The vocal behaviour of mammal-eating killer whales: Communicating with costly calls," Anim. Behav. **69**, 395–405.
Wang, X. Y., Wang, D., Wu, X. B., Wang, R. P., and Wang, C. L. (**2006**). "Congregative effect of the Chinese alligator's bellowing chorus in mating season and its function in reproduction," Acta Zool. Sinica. **52**, 663–668.
Wang, X. Y., Wang, D., Wu, X. B., Wang, R. P., and Wang, C. L. (**2007**). "Acoustic signals of Chinese alligators (*Alligator sinensis*): Social communication," J. Acoust. Soc. Am. **121**, 2984–2989.
Wells, K. D. (**1988**). "The effect of social interactions on anuran vocal behaviour," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczinski, T. E. Hetherington, and W. Walkowiak (Wiley, New York), pp. 433–454.
Wells, K. D., and Schwartz, J. J. (**1984**). "Vocal communication in a neotropical treefrog, *Hyla ebraccata*: Aggressive calls," Behaviour **91**, 128–145.
Wollerman, L., and Wiley, R. H. (**2002**). "Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog," Anim. Behav. **63**, 15–22.
Wu, J. S., and Wang, X. M. (**2004**). "Regulation of bellowing of Chinese alligators (*Alligator sinensis*) in the wild," Zool. Research. **25**, 281–286.
Young, B. A. (**1991**). "Morphological basis of 'growling' in the king cobra, *Ophiophagus hannah*," J. Exp. Zool. **260**, 275–287.
Zhang, F., Wu, X. B., Zhu, J. L., and Zhang, S. (**2005**). "Primary research on the activity rhythm and the behavior coding of capative-bred Chinese alligator in summer and autumn," Acta Hydrobiol. Sin. **29**, 484–494.
Zhu, C. G. (**1957**). "Preliminary study on the life history of Chinese alligator," Acta Zool. Sinica. **9**, 132–138.

Note types and coding in Parid vocalizations: The chick-a-dee call of the chestnut-backed chickadee (*Poecile rufuscens*)

Marisa Hoeschele
Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

David E. Gammon
Department of Biology, Elon University, Campus Box 2625, Elon, North Carolina 27244

Michele K. Moscicki
Department of Psychology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

Christopher B. Sturdy[a)]
Department of Psychology and Centre for Neuroscience, University of Alberta, P217 Biological Sciences Building, Edmonton, Alberta T6G 2E9, Canada

(Received 3 February 2009; revised 22 April 2009; accepted 13 July 2009)

A first step to understanding how a species communicates acoustically is to identify, categorize, and quantify the acoustic parameters of the elements that make up their vocalizations. The "chick-a-dee" call notes of the chestnut-backed chickadee (*Poecile rufescens*) were sorted into four call note categories, A, C, D, and Dh notes, based on their acoustic structure as observed in sound spectrograms, and evaluated based on the syntactical ordering of the note types within calls. The notes were then analyzed using quantitative measures and it was determined which features have the potential to convey information to discriminate note type, individual, and the geographic origin of the producer. The findings were comparable to previous research of congeners in that chestnut-backed chickadee calls were produced with a relatively fixed syntax and contained similarly structured note types across all geographic regions. Overall this information will form a base for future research on chestnut-backed chickadee vocalizations and will strengthen the foundation for future comparative evolutionary studies.
 [DOI: 10.1121/1.3203736]

PACS number(s): 43.80.Ka [MCH] Pages: 2088–2099

I. INTRODUCTION

In order to understand acoustic communication in songbirds, as in all animal species, we must first understand the acoustic structure and note composition that comprise their vocalizations. With this information we can begin to evaluate which aspects of the vocalizations are potentially important for individual recognition, species recognition, and to ascribing meaning. Further, by using similar methods across a number of closely related species, we can gain insights on the evolution of complex acoustic signals.

Chickadees (genus *Poecile*) are a frequently studied songbird genus for several reasons. First, chickadees are common in most areas of North America, making them readily available for scientific investigation. Second, chickadees are an excellent group for comparative study because there are many species of chickadees that all have unique, yet comparable, vocalizations such as the chick-a-dee call (e.g., Hailman and Ficken, 1996). Third, chickadees have a social structure that varies seasonally, whereby they form mated pairs in the spring and summer and in the fall and winter form cooperative flocks of typically six to eight individuals (Smith, 1991). These reasons, combined with the fact that chickadees have a complex and well-studied vocal repertoire, make them an excellent choice for studies of communication (Smith, 1991).

The North American clade of chickadees is often broken up into two sibling groups: black-headed and brown-headed chickadees based on both their appearance and their phylogenetic relationships (Gill *et al.*, 2005). To date, the vocalizations of all the black-headed species of the North American chickadee clade [black-capped, mountain, and Carolina (*atricapillus*, *gambeli*, and *Carolinensis*) chickadees] have been studied extensively, including their whistled song (black-capped: Weisman *et al.*, 1990; Kroodsma *et al.*, 1999; mountain: Wiebe and Lein, 1999; and Carolina: Lohr *et al.*, 1991, 1994) and chick-a-dee call (black-capped: Charrier *et al.*, 2004; mountain: Bloomfield *et al.*, 2004; and Carolina: Bloomfield *et al.*, 2005). The whistled song is not part of the repertoire of brown-headed chickadees (chestnut-backed chickadee, *Poecile refescens*, boreal chickadee, *Poecile hudsonicus*, gray-headed chickadee, *Poecile cinctus*, and Mexican chickadee *Poecile sclateri*), but they all produce chick-a-dee calls (Hailman, 1989). In fact, the chick-a-dee call of these species is used in social contexts normally reserved for the whistled song in the black-headed group. Specifically, the chick-a-dee call of the chestnut-backed chickadee is the most common vocalization of the dawn chorus, and is thought to be used for territory defense (Dahlsten *et al.*, 2002). In con-

[a)]Author to whom correspondence should be addressed. Electronic mail: csturdy@ualberta.ca

trast to the black-headed group, chick-a-dee calls from members of the brown-headed group have not been studied nearly as extensively. Here we present the first in depth study examining the acoustic structure of the chestnut-backed chickadee chick-a-dee call.

In addition to examining the acoustic structure of the chestnut-backed chickadee call, we also looked at whether there was any geographic variation in the structure and syntax of the call; what some researchers refer to as different "dialects." Vocal geographic variation has often been demonstrated for songbird populations separated by physical barriers (e.g., Slabbekoorn *et al.*, 2003), and for populations lacking these barriers (e.g., Wright *et al.*, 2008).

Because of the many regional and seasonal differences that occur in the vocalizations of many species, we examined and compared calls from a group of chestnut-backed chickadees recorded in May 2006 on Vancouver Island, as well as recordings of chestnut-backed chickadees recorded across many regions at different times of year. These regions contained areas of both sympatry and allopatry with other chickadee species. Our methodology for determining note types is similar to that used when studying the chick-a-dee calls of other chickadee species (Charrier *et al.*, 2004; Bloomfield *et al.*, 2004, 2005). We then measure and summarize several acoustic features from each note type to determine which aspects of the vocalizations the chestnut-backed chickadee could potentially use to discriminate among individuals and to discriminate among note types. Using this information, we made comparisons between notes, individuals, and geographic regions. We then postulate mechanisms by which the birds could extract information from conspecific calls. Finally, we examine syntactical trends across a vast library of calls recorded from several geographic areas.

II. STUDY 1: CALL NOTE CLASSIFICATION

The purpose of this study was to examine sound spectrograms of chick-a-dee call notes in order to determine the different note types present in the call. This information was then used as a basis for the subsequent studies.

A. Recordings

Recordings were obtained from two sources: a library of archived calls, and our own field recordings. Our field recordings from Vancouver Island in May 2006 (recorded by DEG) were used as a source from which we had detailed information about the nature of the recordings and individuals present in the recordings, which was important for some of our analyses. However, because this source contained recordings from only one location at one time of year, we also obtained recordings from Macaulay Library of Natural Sounds at the Cornell Laboratory of Ornithology, which consisted of recordings from many different people, with different recording equipment, in different months of the year, in different locations, and no detailed information regarding the individual identities of birds contained in the recordings.

The first sample (hereafter referred to as the Vancouver 2006 sample) of chestnut-backed chickadees was recorded in the field at eight different locations on Vancouver Island, Canada during May 2006 between the hours of 0500 and 1500 Pacific daylight time. All birds were unbanded, so sex was usually not known. A total of 43 pairs of chickadees were recorded. Because birds were unbanded, only recordings from one visit were used if locations were visited more than once. In these recordings it was clear which individual in the pair was vocalizing; this was important information for any analyses that required knowing which individual was calling. Recordings were made using a MiniDisc recorder (model MZ-N1, Sony Corp., Tokyo, Japan) connected to a Sennheiser omnidirectional microphone (model ME62, Sennheiser Corp., Wedemark, Germany), with a 20–20 000 Hz frequency response, mounted in a 60 cm parabola (Telinga Pro-universal model, Tobo, Sweden). The calls were then edited into individual wave files using SYRINX software, sampling at 22.05 kHz.

The goal was to obtain large samples of calls from each of several individuals with a representative sample of their call note type repertoire. In order to achieve this, there were two potential problems to overcome. First, calls were generally produced by birds high in the canopy at very low amplitudes, thus making for challenging sample acquisition in the field. Second, C and D notes in this species were relatively rare compared with other chickadee species studied to date (e.g., Charrier *et al.*, 2004). To ameliorate these issues, we excluded calls from focal individuals if they had poor signal to noise ratio, high levels of background noise, a small sample of calls (<10), and/or had few (<5) calls that contained note types appearing similar to the C and D notes of other chickadee species. We had no numerical cutoff for poor signal to noise ratio, as this was determined by visually examining the spectrograms. However, all spectrograms had a cutoff of -35 dB below peak frequency (PF). Therefore, signals with noise that was louder than -35 dB from the peak of the signal and obscured our ability to properly distinguish the note from background noise were omitted. With all of these considerations in mind, nine individuals were selected from the Vancouver 2006 sample. From each of these individuals we then randomly selected ten calls such that the first five calls from each individual had to contain either a C or D type note, and additional five calls were randomly sampled from the remaining set of recorded calls such that at least five of the calls had notes resembling A or B type notes from other chickadee species. Thus, we sampled in total 10 calls each from 9 individuals for a total of 90 calls consisting of 490 notes.

The second set of calls was obtained from the Macaulay Library of Natural Sounds at Cornell Laboratory of Ornithology. A total of 564 calls and 2242 notes were obtained from this source, with recordings taken from a variety of regions (California, Washington, Oregon, and British Columbia) with various recording equipment (see Table I for details). Using the same procedure as was used for the Vancouver 2006 sample, we selected 12 individual recordings and randomly selected 10 calls from each of these such that at least 5 calls contained C or D type notes, and at least 5 calls contained A type notes for a total of 490 notes. These calls were evaluated after we had already established note types using the

TABLE I. Details about location, recordist, month and year of recording, and equipment used for the sample received from the Cornell Laboratory of Ornithology's Macaulay Library of Natural Sounds.

Region	Local	Recordist	Month	Year
Vancouver Island	Miracle Beach Provincial Park	Gunn, William W. H.	5	1962
Vancouver Island	Miracle Beach Provincial Park	Gunn, William W. H.	5	1962
Northwestern U.S.	Fort Worden State Park	Hewitt, Oliver H.	5	1980
Northwestern U.S.	Powers; Eden Valley	Keller, Geoffrey A.	5	1989
Northwestern U.S.	Powers; Eden Valley	Keller, Geoffrey A.	5	1989
Northwestern U.S.	Charleston, South Slough Sanctuary	Keller, Geoffrey A.	3	2000
Northwestern U.S.	Unknown	Stillwell, J.E.	5	1955
Northwestern U.S.	Lake Crescent	Robbins, Mark B.	6	1989
Northwestern U.S.	Lake Crescent	Robbins, Mark B.	6	1989
Northwestern U.S.	Lake Crescent	Robbins, Mark B.	6	1989
Northwestern U.S.	Cape Perpetua; 10 km in on Tachats Loop Road	Little, Randolph S.	5	1995
Northwestern U.S.	Corvallis; Valley of the Giants	Little, Randolph S.	5	1995
Northwestern U.S.	Fort Worden State Park	Hewitt, Oliver H.	5	1987
California	Carmel	Allen, Arthur A. and Allen, D. G.	4	1959
California	Carmel: PT Lobos	Sutherland, C. A.	7	1961
California	Marin County	Sander, Thomas G.		
California	Crescent City; Lake Earl Wildlife Viewing Area	Keller, Geoffrey A.	5	2001
California	Devil's Gulch	Sander, Thomas G.	3	1988
California	Monta-a de Oro State Park, Coon Creek	Marantz, Curtis A.	6	2005
California	Monta-a de Oro State Park, Coon Creek	Marantz, Curtis A.	6	2005

Vancouver 2006 sample, to see if our results generalize to other regions and seasons. Recordings were from a broad range of years; however, for all analyses that required us to split the library into subgroups, year ranges were balanced as well as possible across subgroups (see Table II for a description of the total samples and the samples used in this study).

B. Classification procedure

To classify the notes into note type categories, a method similar to that used in previously published papers was employed (e.g., Bloomfield *et al.*, 2004). All 490 of the notes selected from the Vancouver 2006 sample were saved in separate sound files with a duration of 300 (non-D notes) or 500 ms (D notes) from which spectrograms were then generated using SIGNAL sound analysis software (Engineering Design, Berkeley, CA) at a window size of 512 points with a cut off amplitude of −35 dB relative to the peak amplitude of the note. These spectrograms were then printed on glossy photo paper and arranged with 15 spectrograms per page, each accompanied with a random 4-digit number. The spectrograms were then cut out such that each note had its own small card.

In a preliminary analysis, we sorted the notes into six categories based on visual similarity. Exemplars were then identified for each of the note types by searching for notes with similar acoustic characteristics from the notes in the calls that were not selected for individual analysis. The exemplars were then printed on note cards in the same fashion as the 490 notes used for the analysis and were labeled as exemplars for each category. These exemplars, as well as a written description of each note type (the final version of which is provided in the results section), were presented to

TABLE II. Total number of subsamples (recordings) and the years and months of recording for each region are presented as well as selected subsamples for specific analysis.

Region	No. of subsamples	Years recorded	Months recorded
Vancouver Island and Macaulay Library total sample information			
Vancouver Island	37	1962–2006	May
Northwestern United States	12	1955–2001	March–June
California	6	1959–2005	March–July
Note card sort and quantitative analysis subsample information			
Vancouver Island	11	1962–2006	May
Northwestern United States	5	1980–2001	May–June
California	5	1959–2005	March–July
Regional comparison subsample information			
Vancouver Island	6	1962–2006	May
Northwestern United States	5	1980–2001	May–June
California	5	1959–2005	March–July

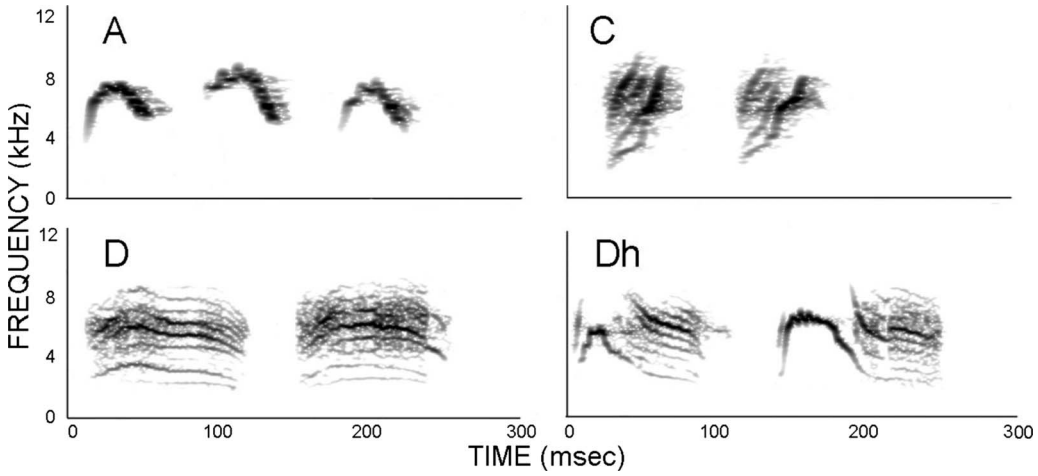


FIG. 1. Spectrogram exemplars of chestnut-backed chickadee note types [fast Fourier transform (FFT) window=512 points].

two additional sorters (Sturdy and Moscicki) who then independently sorted the notes. Afterwards, the percent agreement was calculated between the sorters and the average agreement between two individuals was determined. The sorters then met to discuss any disagreements and refine the note type definitions.

After agreeing on and finalizing the note types of the chestnut-backed chickadee based on the Vancouver 2006 sample, this procedure was repeated with the sample obtained from the Macaulay Library to assess whether the same note types from one location and season are also found in other locations and seasons.

C. Results

In the first (i.e., preliminary) note sort, six note types were identified among the 490 notes in the sample of chestnut-backed chick-a-dee calls. However, due to less than desired percent agreement among sorters (<80%) on three of the note types, the six note types were collapsed into four. Specifically, three of the preliminary note types were collapsed into a single "A" note-type. The percent agreement among sorters after collapsing into four note types was 100%.

The same four note types were also found in the Macaulay Library sample as no notes were identified that did not clearly conform to the categories found in the Vancouver 2006 sample. However, there was slight disagreement among sorters (97% agreement) that resulted in refining the description of the Dh note, such that A notes with a noisy terminal portion would not be misclassified as Dh notes. Below are the final note descriptions that achieved consensus among the sorters. The nomenclature used for the note types observed here is similar to that used in previous studies with chick-a-dee calls. For a visual explanation, Fig. 1 shows exemplars of each note type.

1. A notes

The A note, much like A and B notes in other chickadee species, is highly tonal in nature. It increases in frequency, peaks, and then decreases in frequency, thus forming an inverted U shape, or, in some cases, an inverted V shape. The A note almost always occurs first in the call, if it occurs at all.

2. C notes

The C note of the chestnut-backed chickadee appears visually similar to the C notes of black-capped, mountain, and Carolina chickadees (Charrier *et al.*, 2004; Bloomfield *et al.*, 2004, 2005) in that it consists of a "stack" of frequency bands that ascend in frequency to a peak and are then occasionally observed to rapidly decrease over a very brief duration. The main (loudest) amplitude band is frequency modulated from the beginning of the note to the peak and then occasionally decreases rapidly over a short duration. In addition, there are usually two parallel frequency bands that ascend prior to the main band (see Fig. 1). The C note typically appears in isolation from other call note types in the chestnut-backed chickadee.

3. D notes

The D note is also similar to the D notes of other chickadees in that they are composed of a stack of frequency bands that are relatively unmodulated in frequency over their duration. In the chestnut-backed chickadee, D notes have several harmonic-like bands that span a wide frequency range (approximately 2–9 kHz). The individual bands remain fairly constant in frequency, sometimes ascending slightly at the beginning and descending again at the end. D notes do not occur in every call, but when they do, they almost always appear at the end of the call.

4. D hybrid notes (Dh)

The D hybrid note resembles both the D note and A note, and could be considered a "hybrid" of the two note types, hence the name. It begins with a more tonal structure (like the A note) and ends with a harmonic-like structure (like the D note). It starts out much like the A note with an inverted U or V shape, but with a slightly longer descending than ascending frequency band. The end of the note uses the descending frequency band of the A note-like portion as the

lowest harmonic band for a D note type ending of the note. Most Dh notes are relatively long in duration (~150 ms) and have a portion without frequency modulation that looks almost exactly like a short D note, whereas some Dh notes only have frequency modulation on the descending portion of the tonal part of the note. There are always at least two additional frequency bands that initiate at or just after the tonal portion of the note begins to descend in frequency. When they occur, Dh notes appear after A notes and before D notes, if there are any D notes in the call. There is never more than one Dh note in a call.

III. STUDY 2: QUANTITATIVE NOTE ANALYSIS

After first determining and then classifying notes into types in the previous study, we made quantitative measurements of several measures of frequency and duration on each note. From the results of this analysis, it was possible to make quantitative acoustical comparisons between note types and individual birds.

A. Note measurement procedure

All measurements were completed using SIGNAL 5.10.25 software (Engineering Design, 2001–2007) according to the methods used with other chickadee species (Charrier *et al.*, 2004; Bloomfield *et al.*, 2004, 2005). Because all the Macaulay Library recordings were sampled at 44.1 kHz, which is also the standard sample rate used in other chickadee species, we resampled the Vancouver Island recordings from 22.05 to 44.1 kHz before conducting any measurements.

All frequency measurements were taken in a 1024-point window size spectrogram, which permitted a frequency precision of 43 Hz. We used this window to measure the start frequency (SF), PF, and end frequency (EF) of the frequency band with the highest amplitude for A, C, and the opening portion of D hybrid notes.

Duration measurements were taken in a 256-point window size spectrogram, which has a duration precision of 5.8 ms. Ascending durations (ADs) and descending durations (DDs) to and from the PF were measured in all notes with tonal portions (A, C, and Dh notes) as were the total durations (TDs) for each note type.

Based on the frequency and duration measurements for A, C, and the opening portion of D hybrid notes, we were also able to calculate ascending (FMasc) and descending (FMdes) frequency modulations. These were calculated as follows: $\mathrm{FMasc}=(\mathrm{PF}-\mathrm{SF})/\mathrm{AD}$ and $\mathrm{FMdes}=(\mathrm{PF}-\mathrm{EF})/\mathrm{DD}$.

Finally, a spectrum was generated for each individual note with a smoothing width of 88.2 Hz, in which the loudest frequency (frequency of maximum amplitude, FMax) for all notes was measured. Additionally, the fundamental frequency (f_0) was measured for C, D, and D hybrid notes. See Fig. 2 for a visual explanation of each measurement.

B. Statistical analysis

All initial analyses of the two samples were conducted separately. This was because for the Vancouver 2006 sample, but not the Macauley sample, the same recordist made all of the field recordings using the same equipment. In the Macauley sample, some of the acoustic differences may have been due to different recordists and/or equipment. Thus, the Vancouver 2006 sample was ideal to set the standard for the quantification of chestnut-backed chickadee vocalizations. The Macaulay sample was examined to confirm that the results would generalize to chickadees in other regions and months of the year.

To compare the four note types separately for each sample, potential for note type coding (PNTC) (Charrier *et al.*, 2004) was examined to see whether there was more variation of acoustic features between note types than within a note type. The PNTC is the ratio of the coefficient of variation for a particular acoustic feature, calculated between all note types (CVb) and the mean of the coefficients of variation (for the same acoustic feature) calculated within note types (CVw). The coefficient of variation between note types is calculated as $\mathrm{CVb}=(\mathrm{SD}/\bar{x})\times 100$ where SD is the standard deviation and $\bar{x}$ is the mean for an acoustic feature calculated on the total sample of call notes. The coefficient of variation within (CVw) is an average of the coefficients of variation for all the note types separately for a particular feature, calculated with the same formula. For example, the CVw for SF was calculated from the average coefficient of variation of SF for A, C, D, and Dh notes calculated separately. This allows us to look at the average variation within a note type to compare it to the total variation. If the variation for a given acoustic feature is greater between note types than within note types, then the ratio will be >1, and it is theoretically possible that these differences can be used by the birds to discriminate note types. Conversely, if similar amounts of variation exist between and within types, then the ratio will be ~1, indicating that the particular acoustic feature is unlikely to contribute in any meaningful way to note-type perception.

Similar to the PNTC, potential for individual coding (PIC) (Charrier *et al.*, 2004) was examined. This allowed us to determine whether there are any acoustic features that are specific to an individual, and thus could be used by conspecifics for individual identification. We could distinguish individuals of the Vancouver 2006 sample, but not for the Macaulay Library sample, so only the Vancouver 2006 sample was used for this analysis. The PIC is calculated in the same manner as the PNTC (i.e., PIC=CVb/CVw). However, PIC values were calculated separately for each note type and feature, unlike the PNTC. CVb, therefore, instead of being calculated for each feature across all note types and individuals, was calculated for each feature separately for each note type but across all individuals. CVw is an average of the coefficients of variations within each individual, rather than note type like in PNTC. Because the sample of a single note type within an individual was small, an adjusted version of the CVw was used: $\mathrm{CVw}=(\mathrm{SD}/\bar{x})(1+1/(4n))\times 100$ where n is the number of calls in each bird's sample.

For each sample, a one-way analysis of variance (ANOVA) was also conducted using SPSS 11.5 (SPSS Inc. 1989–2002) to determine what measurements of each note type best predicted individual. A Bonferroni correction of alpha=0.005 (alpha of 0.05/10 measurements) was used to

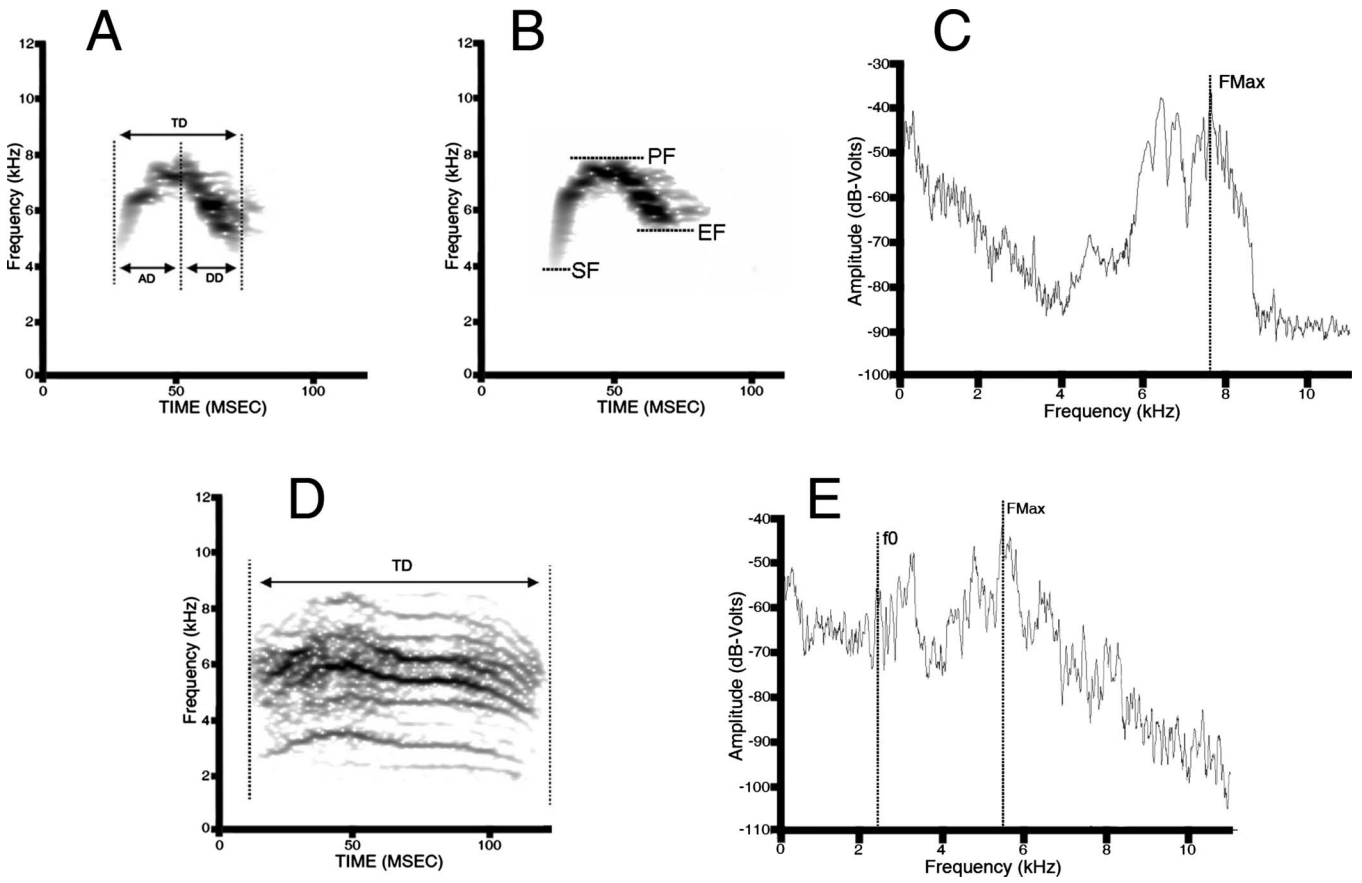


FIG. 2. Exemplars of quantitative measurements. (A) Spectrogram showing duration measurements on non-D notes. TD=total duration, AD=ascending duration, and DD=descending duration. (B) Spectrogram showing frequency measurements on non-D notes. SF=start frequency, EF=end frequency, and PF=peak frequency. (C) Power spectrum calculated on the entire note duration showing frequency at maximum amplitude (FMax) for non-D notes. (D) Spectrogram measuring duration of D notes. (E) Power spectrum calculated on the entire note duration to measure FMax and f_0 in D notes; f_0 =fundamental frequency within note.

account for the multiple measurements that could be used as predictors of individual.

Finally, we conducted linear discriminant analyses (LDAs) using Wilks' lambda stepwise method, and multivariate analyses of variance (MANOVAs) on the Macaulay Library sample for each note type separately to determine whether within-region similarities and/or between-region differences could be observed. Three *ad-hoc* regions were formed and evaluated. The first region was the Vancouver Island chestnut-backed chickadees, which is an allopatric group of chickadees. No other chickadee species are present on the island (Dahlsten *et al.*, 2002). The second region was the Northwestern United States that consisted of Washington, Oregon, and the very Northwestern tip of California. This region contains both black-capped chickadees and chestnut-backed chickadees, making it a region of sympatry for chickadees (Smith, 1993; Dahlsten *et al.*, 2002). The final sample was chestnut-backed chickadees in other regions of California, which live in a region separate from black-capped chickadees. No calls were sampled from chestnut-backed chickadees north and east of Vancouver Island, because none were available from the Macauley Library. Both United States regions also contain some mountain chickadees in the more mountainous regions of the states (McCallum *et al.*, 1999). However, based on the locations that recordings were obtained (when this information was available), there was likely minimal overlap between mountain chickadees and chestnut-backed chickadees in this study since most of the recordings in each of the United States appeared to be taken by the coast rather than in the mountains. Additionally, because the Vancouver 2006 sample was larger than the entire sample received from the Macaulay Library, and since the Macaulay Library contained far less recordings from Vancouver Island than the other two regions, a subsample of calls from the Vancouver 2006 sample was chosen to boost the Vancouver Island region in the regional comparisons. Because we wanted to use samples with the largest potential amount of acoustic variation, we looked for the individuals that had the largest sample of calls from which we pseudorandomly chose the ten calls for analysis. Four birds had $\geq$40 (40–73) calls and the other five had $<$20 (13–17) calls. We therefore included only these four birds. See Table II for a full description of the recordings used.

C. Sample specific results

PNTC values indicated that there were numerous acoustic features that the birds could potentially use to discriminate among note types. Both the Vancouver 2006 sample and the Macaulay Library sample had PNTC values $>$1 for TD, AD, DD, SF, EF, FMasc, and FMdes, suggesting that these measures differ for all note types in both samples. The Van-

TABLE III. PNTC values for both the Vancouver 2006 sample and the Macaulay Library sample. A value >1 indicates a potential for note type coding.

	Vancouver 2006 sample			Macaulay Library sample		
Measurement	CVb	Mean CVw	PNTC	CVb	Mean CVw	PNTC
TD	64.5	55.8	1.2	50.2	17.1	2.9
AD	29.9	24.0	1.2	37.7	22.5	1.7
DD	64.6	53.9	1.2	70.9	45.1	1.6
SF	25.6	20.8	1.2	31.3	24.1	1.3
EF	21.5	20.1	1.1	25.4	17.9	1.4
PF	9.4	8.2	1.1	13.1	15.3	0.9
FMax	9.5	7.7	1.2	11.1	11.2	1.0
f_0	17.5	21.0	0.8	24.4	23.6	1.0
FMasc	33.6	27.7	1.2	52.3	39.7	1.3
Fmdes	36.5	29.6	1 2	705.1	214.0	3.3

couver 2006 sample had more PNTC values >1, as PF and FMax also could be used to discriminate note types in this sample. However, the Macaulay Library sample showed that, across regions, seasons, and years, TD and FMdes are the clearest indicators of note type with much higher PNTC values of 2.9 and 3.3, respectively. See Table III for all PNTC values.

PIC values for the Vancouver 2006 sample were >1 for all measurements for all notes except AD for A notes, SF and FMdes for C notes, and EF, FMax, and f_0 for Dh notes. The highest PIC value for A notes was 1.3 for PF and FMdes. For C notes, TD, EF, and PF had the highest PIC value of 1.5. For D notes, both FMax and f_0 had PIC values of 2.2, and Dh notes had the highest PIC value of 2,1 for PF. The ANOVAs showed similar results to those obtained using the PIC measure: All A measurements $[F(8,362) \geq 2.931,\ p \leq 0.003]$ and D note $[F(7,56) \geq 6.679,\ p < 0.001]$, TD, AD, f_0, and FMasc for C notes $[F(4,31) \geq 4.862,\ p \leq 0.004]$, and PF and FMdes for Dh notes $[F(4,14) \geq 7.115,\ p \leq 0.002]$ were significantly different across individuals. These results suggest that all note types have the potential to be used by birds to discriminate among individuals (See Table IV for all PIC information on the Vancouver 2006 sample). However, it is possible, as with PNTC, that some of these results could be due to the limited variation within the particular sample.

D. Comparative results

The LDAs conducted for each note type on the three regions suggested that A notes are similar across the three regions. In spite of this, the model was able to sort them at a better than chance level. The other three note types (i.e., C, D, and Dh) form clearly distinguishable categories based on region. In agreement with these results, all MANOVAs conducted on all note types showed significant differences across regions for all note types.

A notes were correctly classified by location of origin by the LDA model an average of 56.5% of the time. However, California A notes were classified correctly 68.4% of the time and Vancouver Island and Northwestern United States A notes were only classified as being Californian 14.4% and 13.2% of the time, respectively. When looking at the MANOVA to try and further evaluate the differences among groups, we found that the model returned a significant result $[F(18,942) = 11.858,\ p < 0.001]$. There were main effects for all A note measurements across regions except FMasc and FMdes $[F(2,478) \geq 5.108,\ p \leq 0.006]$. To look at these further, we conducted Tukey *post-hoc* analyses and found that, for all significant measurements except FMax, SF, and PF, California was significantly different from the other two regions ($p \leq 0.001$). That said, the two more northerly regions were not significantly different from each other. For SF only Vancouver Island and California were significantly different from each other ($p = 0.003$), and for PF and FMax the opposite was true: All regions except Vancouver Island and California were different from each other ($p \leq 0.035$).

C notes were classified correctly 77.8% of the time by the LDA model. In this case, both Vancouver Island and the Northwestern United States group were classified correctly frequently (81.0% and 83.3% of the time, respectively) whereas California classified correctly 66.7% of the time. The MANOVA was significant $[F(20,140) = 6.527,\ p < 0.001]$, with main effects for all measurements except DD, FMax, and FMdes $[F(2,78) \geq 9.147,\ p \leq 0.001]$. When looking at the results of Tukey *post-hoc* comparisons for all measurements with significant main effects, Vancouver Island was significantly different from the other two regions ($p \leq 0.008$) with the exception of SF, while California was significantly different from the other two regions only for SF ($p < 0.001$).

D notes were classified correctly 74.4% of the time by the LDA model. Interestingly, Northwestern United States and Vancouver Island D notes were least likely to be misclassified as Californian (0.0% and 3.8%, respectively). In general, the LDA had the highest accuracy at classifying the Northwestern United States region, with an accuracy of 90.5%. Vancouver Island and California had 65.4% and 71.0% accuracies, respectively. This suggests that the Northwestern United States D notes were the least variable, but they were intermediate between Vancouver Island and California D notes, which are more distinct from one another. The MANOVA for D notes was again significant $[F(6,146) = 15.258,\ p < 0.001]$ and showed main effects for all measurements $[F(2,75) \geq 8.831,\ p < 0.001]$. Tukey *post-*

TABLE IV. Mean, standard deviation, CVb, mean CVw, and PIC values for each measurement on each note type for the Vancouver 2006 sample. *F* statistics and *df* are presented as well as comparison. A PIC value >1 suggests potential for individual coding.

Note type		TD	AD	DD	SF	EF	PF	Fmax	f_0	FMasc	FMdes
A	Mean ± SD	51.7 ± 10.2 ms	23.7 ± 5.9 ms	28.5 ± 8.6 ms	4996.4 ± 997.7 Hz	4653.2 ± 667.5 Hz	7853.3 ± 684.0 Hz	6599.9 ± 546.0 Hz		127.8 ± 44.6 Hz/ms	121.1 ± 38.8 Hz/ms
	CVb	19.8	25.0	30.1	20.0	14.3	8.7	8.3		34.9	32.1
	Mean CVw	17.8	25.1	26.3	18.3	13.6	6.6	7.3		29.3	24.5
	PIC	1.1	1.0	1.1	1.1	1.1	1.3	1.1		1.2	1.3
	F	12.184[a]	2.931[a]	12.660[a]	13.055[a]	4.163[a]	30.723[a]	12.189[a]		13.153[a]	36.565[a]
	df	370.0	370.0	370.0	370.0	370.0	370.0	370.0		370.0	370.0
C	Mean ± SD	51.8 ± 7.0 ms	40.8 ± 3.6 ms	11.0 ± 5.1 ms	2425.2 ± 280.6 Hz	6652.7 ± 918.2 Hz	8041.8 ± 570.6 Hz	6153.7 ± 575.3 Hz	2854.5 ± 467.7 Hz	138.0 ± 15.4 Hz/ms	127.1 ± 54.3 Hz/ms
	CVb	13.5	8.8	45.9	11.6	13.8	7.1	9.3	16.38364971	11.1	42.7
	Mean CVw	9.2	6.4	33.0	11.8	9.4	4.8	7.9	12.0	9.7	90.4
	PIC	1.5	1.4	1.4	1.0	1.5	1.5	1.2	1.4	1.2	0.5
	F	5.050[a]	9.818[a]	1.832	0.436	2.730	4.372	3.255	7.681[a]	4.862[a]	1.626
	df	35	35	35	35	35	35	35	35	35	35
D	Mean ± SD	175.5 ± 21.9 ms						5731.4 ± 375.2 Hz	2648.8 ± 507.0 Hz		
	CVb	12.5						6.5	19.1		
	Mean CVw	8.4						3.0	8.7		
	PIC	1.5						2.2	2.2		
	F	9.373[a]						24.960[a]	8.650[a]		
	df	63						63	63		
Dh	Mean ± SD	149.0 ± 20.0 ms	25.5 ± 4.4 ms	107.6 ± 21.6 ms	3682.0 ± 767.7 Hz	2303.6 ± 299.6 Hz	6432.5 ± 579.5 Hz	5999.6 ± 294.9 Hz	2529.8 ± 210.0 Hz	110.6 ± 39.3 Hz/ms	39.6 ± 10.1 Hz/ms
	CVb	13.4	17.4	20.1	20.9	13.0	9.0	4.9	8.3	35.5	25.4
	Mean CVw	9.3	13.7	17.2	19.5	12.9	4.4	5.4	8.0	24.2	14.3
	PIC	1.4	1.3	1.2	1.1	1.0	2.1	0.9	1.0	1.5	1.8
	F	4.141	0.918	0.781	1.824	0.503	22.440[a]	0.654	3.488	2.299	7.115[a]
	df	18	18	18	18	18	18	18	18	18	18

[a]Significant with $p < 0.005$.

hoc comparisons revealed that for TD, Vancouver Island was significantly different from the other two regions ($p < 0.001$). For FMax and f_0, California was significantly different from the other two regions ($p \leq 0.002$).

The LDA model classified Dh notes by region correctly 63.8% of the time. The model was high for California (83.3%) and did less well with Vancouver Island (64.3%) and even less with the Northwestern United States region (52.4%). There were no misclassifications of Vancouver Island as California, or California as the Northwestern United States. Only the Northwestern United states was misclassified as both other regions, suggesting that these notes were intermediate between the other two. Because, in this case, the Northwestern United States was more likely to be misclassified than both the other two regions together, it appears that this region was the most variable. The MANOVA looking at the differences in measurements across regions was, again, significant [$F(20,72)=3.958$, $p<0.001$]. However, only AD, SF, PF, and FMax showed significant main effects [$F(2,44) \geq 3.880$, $p \leq 0.028$]. When examining the *post-hoc* tests with these variables, Vancouver Island and Northwestern United States differed on AD ($p=0.027$), California differed from the other two for SF and PF ($p \leq 0.039$), and Vancouver Island differed from the other two for FMax ($p \leq 0.035$).

In summary, for the majority of the measurements for the notes, Vancouver Island and California were the most different with the Northwestern United States falling somewhere in between, suggesting that the acoustic differences may be correlated with geographic distance between the populations.

IV. STUDY 3: SYNTACTICAL ANALYSIS

After distinguishing note types and analyzing their features, the next step was to examine their ordering, or syntax, within the call. The subsamples in our studies were pseudorandomly selected; that is, we purposefully selected calls that contained a variety of note types. Because of this, and because we wanted as large a sample as possible to determine call syntax trends, we analyzed every usable call (i.e., high quality, low noise with discernable note types) in both samples. This not only gave us descriptive information of the call, but also allowed us to evaluate how our large Vancouver 2006 sample compared to a more general sample, and whether there are any differences in syntax types and rules present. If there are differences between the samples, this would suggest that regional or seasonal differences could also play a role in syntax.

A. Call note ordering procedure

Whole calls were printed with 4 spectrograms per page on 8.5×11 pages of paper in a binder of calls. Sorters recorded the note type ordering observed within each call. All three sorters were blind to the classifications from the other sorters, and were the same sorters from study 1. The sorts were then compared, and any disagreements discussed together in a group until the group came to a general consensus to determine the final agreed syntax.

From the data obtained from the full call sets, all syntactical arrangements were analyzed and the total number of times a call was produced was examined. Additionally, we looked at probabilities within the samples of note type order and placement in order to evaluate whether there are standard syntactical rules followed in both samples. First we looked at the probability of notes occurring given the previous note type. For example, what is the probability that the next note is a C note given that the previous note is an A note? These probabilities were calculated by summing up the total number of all occurrences (e.g., total occurrences of A followed by C notes) and finding the proportion of each occurrence given the preceding note (e.g., total number of A followed by C occurrences/total transitions from A notes).

We then looked at the probabilities of note type placement within individual calls. For example, what is the probability that the first note in a call will be an A note? Or, vice versa: What is the probability that an A note will be the first note in a call? These were calculated in the same way as the order probabilities, e.g., the total number of A notes that are the first note in the call/the total number of notes that are first in the call. An example for the opposite calculation would be the total number of A notes that were the first note in the call/the total number of A notes. Comparable studies have been conducted in both black-capped chickadees (Hailman *et al.*, 1987) and Mexican chickadees (Ficken *et al.*, 1994) that found both species have fixed syntactical rules used to generate calls based on Markov chain analyses.

Finally, Pearson correlations were conducted to compare the proportions of each syntax type between the Vancouver 2006 sample and the Macaulay Library sample to see whether there are any syntactical differences between a specific and more general sample. Similar correlations were conducted on the probabilities of each note type followed by each other note type, as well as the probabilities of the placement for each note type within the call.

B. Results

The percent agreement in classifying note types was extremely high (99.6%). Sixty-eight specific syntax patterns were found in the two samples together, suggesting that there is a wide range of patterns the birds can use in communicating. However, 55.6% of all chick-a-dee calls contain just A notes, and another 22.7% contained just A notes that terminate in either a single D note or a single Dh note. Overall, 88.7% of all calls in our sample begin with A notes and most of the remaining calls contain either just C notes (5.4%) or just D notes (4.8%). Only 1.1% of the calls (17/1532) have any other syntax not described above. The chestnut-backed chickadee does, therefore, appear to have a more primitive syntactical structure to its call than any of the black-headed chickadee species (Bloomfield *et al.*, 2004, 2005; Charrier *et al.*, 2004), although there is still room for variation within these parameters.

The frequency of each syntactic sequence was significantly correlated between the two samples, but the correlation was not high ($r=0.534$, $p<0.001$). When looking again at the data, it is apparent that the Vancouver 2006

sample is dominated by certain vocalizations, whereas the Macaulay sample seems to have more variation, which makes sense considering that the Macaulay sample is more varied in terms of location and season. In fact, 34.3% of the calls in the Vancouver 2006 sample have identical syntax (six A notes) whereas the most common call in the Macaulay sample (four A notes) only makes up 11.7% of the total calls. See Table V for detailed information on the call types and their proportions in the different samples.

However, when looking at what proportion of the time a note, e.g., an A note, is followed by every other note, the two samples are extremely similar ($r=0.949$, $p<0.001$). This suggests that the difference between syntax types is not due to different syntactical rules about which note should follow which other note. Overall, A notes are most commonly followed by more A notes, C notes are most commonly followed by more C notes, while D and Dh notes are most commonly the last notes within the call. See Table VI for more information on the proportions of the note types following each note type.

V. DISCUSSION

This paper provides the first detailed acoustic description of the note types of the chick-a-dee call of the chestnut-backed chickadee. This species is the first member of the brown-headed chickadee clade to have its chick-a-dee calls studied in such depth, which opens up new opportunities for comparative study. We looked at both a specific and broad sample of calls in terms of location and time of year, making our study the most comprehensive yet.

We found that the chestnut-backed chick-a-dee call was composed of 4 distinct note types (A, C, D, and Dh) that were arranged in 68 syntactical call types in our sample of 1532 calls with 7628 notes. Based on the 68 call types, we described syntactical rules based on the probabilities of note types being followed by all the other note types, and on the probabilities of the different placements within the call. For all note types there were several acoustical features that could be used by listening birds to differentiate between note types, as well as the individual producing the note. We also found regional differences in the structure of all note types suggesting that, although all regions use the same note types, the way they are produced varies from region to region.

A. Individual discrimination

Based on the PIC results of the Vancouver 2006 sample, there appear to be many potential mechanisms for birds to be able to differentiate between individuals. This information provides us with numerous possible hypotheses on how chickadees could discriminate between individuals. As one example, a listener might distinguish between individuals by attending to the fundamental frequency of D notes, and this possibility could be tested using operant conditioning paradigms (Sturdy *et al.*, 2000). However, because we recorded unbanded birds, it was only possible to record each bird used for analysis once, and it is possible that the current motivational and contextual states of the birds may have contributed to the differences we found among individuals. We believe

TABLE V. Syntax types present in each sample (Vancouver 2006 and Macaulay Library) and totaled over the entire sample.

Syntax pattern	Grand total	Vancouver Island (%)	Macaulay (%)
AA	26	0.7	3.4
AAA	73	1.1	11.2
AAAA	113	4.9	11.7
AAAAA	189	15.5	6.8
AAAAAA	399	34.3	11.5
AAAAAAA	8	0.5	0.5
AAAAAAAA	17	1.5	0.4
AAAAAAAAA	6	0.6	0.0
AAAAAAAAAA	18	1.8	0.0
AAAAAAAAAAA	2	0.1	0.2
AAAAAAAAAAADh	1	0.0	0.2
AAAAAAAAADh	2	0.2	0.0
AAAAAAAACCC	1	0.0	0.2
AAAAAAAAD	2	0.2	0.0
AAAAAAAD	1	0.1	0.0
AAAAAAADh	1	0 1	0.0
AAAAAACC	1	0.1	0.0
AAAAAAD	16	1.6	0.0
AAAAAADD	2	0.2	0.0
AAAAACCCD	1	0.0	0.2
AAAAAD	22	2.0	0.4
AAAAADh	8	0.8	0.0
AAAAADhD	2	0.0	0.4
AAAACC	1	0.1	0.0
AAAACCC	3	0.3	0.0
AAAACCCC	12	1.2	0.0
AAAACCCCC	1	0.1	0.0
AAAACCCD	1	0.1	0.0
AAAAD	113	9.2	4.1
AAAADD	9	0.5	0.7
AAAADh	92	9.0	0.7
AAAADhD	41	4.1	0.2
AAACC	4	0.0	0.7
AAACCC	4	0.0	0.7
AAAD	16	0.8	1.4
AAADD	23	1.3	1.8
AAADDD	1	0.1	0.0
AAADh	28	0.0	5.0
AAADhD	26	0.1	4.5
AAC	1	0.0	0.2
AACC	7	0.1	1.1
AACCD	1	0.0	0.2
AAD	15	0.4	2.0
AADD	3	0.0	0.5
AADh	24	0.4	3.6
AADhD	12	0.6	1.1
AD	2	0.0	0.4
ADD	4	0.0	0.7
ADh	4	0.0	0.7
C	2	0.2	0.0
CAAAD	8	0.8	0.0
CC	20	0.9	2.0
CCAA	1	0.0	0.2
CCAAAADh	1	0.1	0.0
CCC	26	0.9	3.1
CCCC	16	0.3	2.3
CCCCC	7	0.3	0.7
CCCCCC	3	0.3	0.0

TABLE V. (*Continued.*)

Syntax pattern	Grand total	Vancouver Island (%)	Macaulay (%)
CCCCCCC	7	0.6	0.2
CCCCCCCC	1	0.1	0.0
DADh	1	0.0	0.2
DD	64	0.0	11.5
DDAAAA	1	0.0	0.2
DDD	6	0.0	1.1
DDDD	3	0.0	0.5
DDDDD	1	0.0	0.2
DDh	2	0.0	0.4
DhD	3	0 3	0.0

that this is most likely not the case, as other chickadee species that were recorded on multiple occasions in the laboratory also showed similar PIC values (Bloomfield *et al.*, 2004; Charrier *et al.*, 2004).

B. Note discrimination and syntax

The PNTC results suggest that all note types found in our study can be distinguished using several features. The Vancouver 2006 sample shows that many features measured on the notes can be used to tell note types apart. Fewer features can be used across populations as shown through the PNTC values for the Macaulay sample, possibly because of variations between regions. However, there are two features, namely, TD and FMdes, which stand out in the Macaulay Library sample with extremely high PNTC values, which suggests that these two features are especially important in differentiating note types across all regions.

We found numerous syntactical arrangements of the chestnut-backed chick-a-dee call, which might also be used in different contexts, as in other chickadee species. For example, black-capped chickadees produce more D notes when a model of a more threatening species of predator is presented (Templeton *et al.*, 2005), and Carolina chickadee respond to calls rich in C notes at feeders (Freeberg and Lucas, 2002). By determining what features of the notes the bird could potentially use to discriminate between note types, potential mechanisms for processing meaning of call syntax have also been described. This has been shown with both black-capped and mountain chickadees, which are able to discriminate each others vocalizations into categories (Bloomfield and Sturdy, 2008) and the mechanisms for which are currently being explored in our laboratory.

In addition to this, our findings from study 3 on the syntax of the calls suggest that once a note type has been generated in a call, the probability for the next note type is the same across samples. Thus, the differences in frequencies of specific call syntax between the Vancouver 2006 sample and the Macaulay Library sample seem not to be due to different syntactical rules; instead they may be due either to sampling error, or to different contexts, which could be seasonal contexts such as food availability. This is interesting because even though the chestnut-backed chickadee appears to use its call for the same functions that song usually serves (Dahlsten *et al.*, 2002), and components of songs that are present often vary across regions (e.g., Slabbekoorn *et al.*, 2003), the birds appear capable of producing the same call set in all regions. This finding is not trivial and was not necessarily to be expected from the outset; gargle calls in chickadees have been shown to vary significantly across regions, with more syllable sharing within than between regions (Baker *et al.*, 2000).

C. Regional differences

The LDAs conducted for each note type on the three regions suggested that all note types except A notes were easily differentiated across regions, and A notes were also distinguished based on region at a greater than chance level. In addition, the MANOVAs showed that all notes were significantly different across regions. Thus it is possible that these differences were indicative of different dialects in different regions that chestnut-backed chickadees inhabit, in line with other research on geographic variation in avian vocalizations. C, D, and Dh notes were the most different between the Vancouver Island and California samples, suggesting a geographic correlation, such that, as geographic distance increases, so do differences in call note production. However, the reasons for these differences are not yet clear. Thorough comparative investigations of sympatrically living species could help us determine whether these differences are due to acoustic character displacement (Brown and Wilson, 1956). Alternatively, by evaluating whether the different regions consist of different major habitat types, we could evaluate whether differences in vocalizations are due to acoustic adaptation (Morton, 1975). Finally, regional differences might reflect genetics differences (Burg, 2007). Given our sample, it was not possible to address other potential vocal differences due to season or time, or look at the regions in more depth, but this is something that could be looked at with a more controlled sample in the future.

TABLE VI. Probability of the next note type given the preceding note. These probabilities are based on the entire sample.

	Probability of next note (%)				
Preceeding note	A	C	D	Dh	Call end
A	78.90	0.59	3.70	3.50	13.31
C	2.39	68.90	0.72	0.00	27.99
D	0.37	0.00	24.86	0.37	74.39
Dh	0.00	0.00	33.19	0.00	66.81

VI. Conclusions

Overall, our results can be used as a starting point for further investigations of the chick-a-dee call. In addition, the chestnut-backed chickadee is an especially important addition to current knowledge on chickadee vocalizations because it will allow us to conduct comparative evolutionary studies between black-headed and brown-headed chickadees, which were not possible before.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant, an Alberta Ingenuity New Faculty grant, Canada Foundation for Innovation (CFI) New Opportunities and Infrastructure Operating Fund grants, and start-up and CFI partner funding from the University of Alberta to C.B.S. M.H. was supported by the Queen Elizabeth II master's scholarship (QEII) scholarship (University of Alberta). D.E.G. was supported by a Faculty Development grant from St. Edwards University in Austin, Texas. M.K.M. was supported by NSERC post-graduate scholarship (PGS) and Alberta Ingenuity Fund scholarship.

Baker, M. C., Howard, T. M., and Sweet, P. W. (**2000**). "Microgeographic variation and sharing of the gargle vocalization and its component syllables in black-capped chickadee (Aves, Paridae, *Poecile atricapillus*) populations," Ethology **106**, 819–838.

Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (**2004**). "Note types and coding in Parid vocalizations II: The chick-a-dee call of the mountain chickadee (*Poecile gambeli*)," Can. J. Zool. **82**, 780–793.

Bloomfield, L. L., Phillmore, L. S., Weisman, R. G., and Sturdy, C. B. (**2005**). "Note types and coding in Parid vocalizations III: The chick-a-dee call of the Carolina chickadee (*Poecile carolinensis*)," Can. J. Zool. **83**, 820–833.

Bloomfield, L. L., and Sturdy, C. B. (**2008**). "All 'chick-a-dee' calls are not created equally. Part I: Open-ended categorization of chick-a-dee calls by sympatric and allopatric chickadees," Behav. Processes **77**, 73–86.

Brown, W. L., and Wilson, E. O. (**1956**). "Character displacement," Syst. Zool. **5**, 49–65.

Burg, T. M. (**2007**). "Phylogeography of chestnut-backed chickadees in western North America," in *Ecology and Behaviour of Chickadees and Titmice: An Integrated Approach*, edited by K. A. Otter (Oxford University Press, London), pp. 77–94.

Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (**2004**). "Note types and coding in Parid vocalizations I: The chick-a-dee call of the black-capped chickadee (*Poecile atricapillus*)," Can. J. Zool. **82**, 769–779.

Dahlsten, D. L., Brennan, L. A., Mccallum, D. A., and Gaunt, S. L. (**2002**). "Chestnut-backed Chickadee (*Poecile rufescens*)," in *The Birds of North America Online*, edited by A. Poole (Cornell Laboratory of Ornithology, Ithaca, NY). Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/689 (Last viewed 2/3/2009).

Ficken, M. S., Hailman, E. D., and Hailman, J. P. (**1994**). "The chick-a-dee call system of the Mexican chickadee," Condor **96**, 70–82.

Freeberg, T. M., and Lucas, J. R. (**2002**). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*," Anim. Behav. **63**, 837–845.

Gill, F. B., Slikas, B., and Sheldon, F. H. (**2005**). "Phylogeny of titmice (Paridae): II. Species relationships based on sequences of the mitochondria cytochrome-B gene," Auk **122**, 121–143.

Hailman, J. P. (**1989**). "The organization of major vocalizations in the Paridae," Wilson Bull. **101**, 305–343.

Hailman, J. P., and Ficken, M. S. (**1996**). "Comparative analysis of vocal repertoires, with reference to chickadees," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 136–159.

Hailman, J. P., Ficken, M. S., and Ficken, R. W. (**1987**). "Constraints on the structure of combinatorial 'chick-a-dee' calls," Ethology **75**, 62–80.

Kroodsma, D. E., Byers, B. E., Halkin, S. L., Hill, C., Minis, D., Bolsinger, J. R., Dawson, J. A., Donelan, E., Farrington, J., Gill, F. B., Houlihan, P., Innes, D., Keller, G., Macaulay, L., Marantz, C. A., Ortiz, J., Stoddard, P. K., and Wilda, K. (**1999**). "Geographic variation in black-capped chickadee songs and singing behavior," Auk **116**, 387–402.

Lohr, B., Nowicki, S., and Weisman, R. (**1991**). "Pitch production in Carolina chickadee songs," Condor **93**, 197–199.

Lohr, B., Weisman, R., and Nowicki, S. (**1994**). "The role of pitch cues in song recognition by Carolina chickadees (*Parus carolinensis*)," Behaviour **130**, 1–15.

Mccallum, D. A., Grundel, R., and Dahlsten, D. L. (**1999**). "Mountain chickadee (*Poecile gambeli*)," in *The Birds of North America Online*, edited by A. Poole (Cornell Laboratory of Ornithology, Ithaca, NY). Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/453 (Last viewed 2/3/2009).

Morton, E. S. (**1975**). "Ecological sources of selection on avian sounds," Am. Nat. **109**, 17–34.

Slabbekoorn, H., Jesse, A., and Bell, D. A. (**2003**). "Microgeographic song variation in island populations of the white-crowned sparrow (*Zonotrichia leucophrys nutallii*): Innovation through recombination," Behaviour **140**, 947–963.

Smith, S. M. (**1991**). *The Black-Capped Chickadee: Behavioral Ecology and Natural History* (Cornell University Press, Ithaca, NY).

Smith, S. M. (**1993**). "Black-capped chickadee (*Poecile atricapillus*)," in *The Birds of North America Online*, edited by A. Poole (Cornell Laboratory of Ornithology, Ithaca, NY). Retrieved from the Birds of North America Online: http://bna.birds.cornell.edu/bna/species/039 (Last viewed 2/3/2009).

Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**2000**). "Call-note discriminations in black-capped chickadees (*Poecile atricapillus*)," J. Comp. Psychol. **114**, 357–364.

Templeton, C. N., Greene, E., and Davis, K. (**2005**). "Allometry of alarm calls: Black-capped chickadees encode information about predator size," Science **308**, 1934–1937.

Weisman, R., Ratcliffe, L., Johnsrude, I., and Hurly, T. A. (**1990**). "Absolute and relative pitch production in the song of the black-capped chickadee," Condor **92**, 118–124.

Wiebe, M. O., and Lein, M. R. (**1999**). "Use of song types by mountain chickadees (*Poecile gambeli*)," Wilson Bull. **111**, 368–375.

Wright, T. F., Dahlin, C. R., and Salinas-Melgoza, A. (**2008**). "Stability and change in vocal dialects of the Yellow-Naped Amazon," Anim. Behav. **76**, 1017–1027.

Sound pressure and particle acceleration audiograms in three marine fish species from the Adriatic Sea

Lidia Eva Wysocki[a)]
Department of Behavioural Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

Antonio Codarin
Department of Biology, University of Trieste, viale Giorgieri 7, 34127 Trieste, Italy

Friedrich Ladich
Department of Behavioural Biology, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria

Marta Picciulin
Department of Biology, University of Trieste, viale Giorgieri 7, 34127 Trieste, Italy and Etho-ecology Marine Laboratory, Natural Marine Reserve of Miramare-WWF, viale Miramare 349, 34014 Trieste, Italy

(Received 8 August 2008; revised 9 March 2009; accepted 10 July 2009)

Fishes show great variability in hearing sensitivity, bandwidth, and the appropriate stimulus component for the inner ear (particle motion or pressure). Here, hearing sensitivities in three vocal marine species belonging to different families were described in terms of sound pressure and particle acceleration. In particular, hearing sensitivity to tone bursts of varying frequencies were measured in the red-mouthed goby *Gobius cruentatus*, the Mediterranean damselfish *Chromis chromis*, and the brown meagre *Sciaena umbra* using the non-invasive auditory evoked potential-recording technique. Hearing thresholds were measured in terms of sound pressure level and particle acceleration level in the three Cartesian directions using a newly developed miniature pressure-acceleration sensor. The brown meagre showed the broadest hearing range (up to 3000 Hz) and the best hearing sensitivity, both in terms of sound pressure and particle acceleration. The red-mouthed goby and the damselfish were less sensitive, with upper frequency limits of 700 and 600 Hz, respectively. The low auditory thresholds and the large hearing bandwidth of *S. umbra* indicate that sound pressure may play a role in *S. umbra's* hearing, even though pronounced connections between the swim bladder and the inner ears are lacking.
 [DOI: 10.1121/1.3203562]

PACS number(s): 43.80.Lb, 43.64.Ri, 43.30.Xm, 43.66.Gf [ADP] Pages: 2100–2107

I. INTRODUCTION

Fishes are known to have a great variability in hearing abilities in terms of hearing bandwidth, hearing sensitivity, and the appropriate stimulus component for the inner ear. The otolithic end organs in the inner ear are thought to act as biological accelerometers (Popper and Fay, 1999). Accordingly, fishes are thought to be primarily sensitive to the particle motion component of sound. Particle motion can be either described as acoustic displacement, particle velocity, or particle acceleration, each of which can be calculated based on the other two units. It has been suggested by several researchers that particle acceleration may be the most appropriate component for describing particle motion of sound in the context of fish hearing (Kalmijn, 1988; Fay and Edds-Walton, 1997; Popper and Fay, 1999; Bass and McKibben, 2003) Fishes without swim bladders are only sensitive to particle motion (Enger and Andersen, 1967; Banner, 1967; Chapman and Sand, 1974).

Several groups from unrelated taxa (often termed "hearing specialists") have independently evolved the ability to perceive also the pressure component of sound. This considerably lowers their hearing thresholds and extends the hearing bandwidth to higher frequencies up to several kilohertz (Hawkins and Myrberg, 1983; Ladich and Popper, 2004). Such enhanced hearing abilities are primarily based on accessory hearing structures consisting of air-filled cavities. These transmit oscillations of their walls in the pressure field to the inner ear. Examples for such accessory hearing structures are the Weberian apparatus of otophysans (Ladich and Wysocki, 2003), the suprabranchial chambers (labyrinths) of anabantoids, and the auditory bullae of mormyrids and clupeids (Stipetić, 1939; Schneider, 1941; Blaxter *et al.*, 1981; Fletcher and Crawford, 2001). The degree of the connection between the air-filled cavity and the inner ear, as well as the size of the cavity, influence the hearing bandwidth and sensitivity. Catfishes with large swim bladders and a higher number of connecting ossicles have better high-frequency hearing than groups with small swim bladders and fewer Weberian ossicles (Lechner and Ladich, 2008). Within holocentrids, a shorter distance between anterior swim bladder extensions and inner ear results in lower hearing thresholds and broader hearing bandwidth (Coombs and Popper, 1979; Hawkins, 1993). In sciaenids, this dependence of hearing on the structure of the auditory periphery seems to be less pro-

[a)]Author to whom correspondence should be addressed. Electronic mail: lidia.wysocki@univie.ac.at

nounced, although a smaller distance between the inner ear and anterior swim bladder protrusions may be associated with a higher upper frequency detected (Ramcharitar *et al.*, 2006a).

The cod *Gadus morhua* and damselfish of the genus *Stegastes* (syn. *Eupomacentrus*) have been shown to detect sound pressure at the higher frequencies within their hearing range. At low frequencies (below about 100 Hz), however, these species are particle motion sensitive (Chapman and Hawkins, 1973; Sand and Enger, 1973; Myrberg and Spires, 1980). Cahn *et al.* (1969) found that two species of the grunt *Haemulon* could also shift from particle motion sensitivity to pressure sensitivity as frequency increased. Whereas at 400 Hz the grunts were sensitive to pressure only, they responded to either pressure or particle motion at frequencies of 100 and 200 Hz. The swim bladder of these fishes appears to serve as an accessory hearing structure: oscillations are obviously transmitted through the surrounding tissue to the inner ear, even though there is no apparent specialized anatomical link to the inner ear.

In many fish species, it is unknown which sound component (particle motion or sound pressure) is more relevant for detecting sound at the hearing threshold. Masking studies in cod have led Buwalda (1981) to assume that signal detection is ruled by whatever of the two components has the highest signal-to-noise ratio in a given instance and location.

These observations highly limit the possibility to estimate the auditory sensitivity, hearing bandwidth, or sound component (particle motion or sound pressure) detected by a previously unstudied species even within a given taxonomic group. This underlines the need for assessing hearing in many more species than has been currently done, whereby both sound components must be examined.

Due to the lack of commercially available sensors, most fish audiograms have been described in terms of sound pressure level (SPL), although this may not always be appropriate. The recent development of miniature sensors allows the simultaneous measurement of sound pressure and particle motion (McConnell, 2003; McConnell and Jensen, 2006). This overcomes earlier limitations and enables to characterize the sound stimulus at threshold for both components of sound in the small tanks typically used during fish audiometry tests whose acoustics are very complex and do not allow to calculate particle motion out of SPL measurements based on typical assumptions for an acoustic free-field environment.

The aim of the present study was therefore to describe the hearing range and sensitivity of three vocal Mediterranean species from three perciform families in terms of SPL and three-dimensional particle acceleration level at hearing threshold. All three species occur within the Miramare Natural Marine Reserve near Trieste (Italy) and are currently target species for assessing the impact of human activities, namely boat noise, on their biology and physiology (Codarin *et al.*, 2009; Picciulin *et al.*, 2008).

II. MATERIALS AND METHODS

A. Animals

Test subjects were six brown meagre *Sciaena umbra* [142–173 mm standard length (SL); 53.3–128 g body mass (BM)], six damselfish *Chromis chromis* (72–89 mm SL; 13.9–20.9 g BM), and six red-mouthed gobies *Gobius cruentatus* (97–121 mm SL; 17.6–37.7 g BM). All test subjects were captured with trap nets at rocky reefs facing the Gulf of Trieste (North Adriatic Sea, Italy) prior to the study, and then transported to Vienna. Each species was kept in 250 l tanks for at least 1 week before starting the experiments. The bottoms of the aquaria were covered with sand and equipped with several plastic shelters. The aquaria were fitted with external filters and protein skimmers for salt water, water temperature was kept at 20 °C, and a 12 h:12 h L:D cycle was maintained. Fish were fed with frozen mussels (*Mytilus galloprovincialis*), crustaceans (*Penaeus* spp. and *Daphnia* spp.) and commercial food for sea water fish (TetraMin®, TetraWerke, Germany). All experiments were performed with the permission of the Austrian Federal Ministry for Education, Science and Culture (GZ 66.006/2-BrGT/2006).

B. Auditory sensitivity measurements

Hearing thresholds were obtained using the AEP-recording technique. The AEP protocol was based on that introduced by Kenyon *et al.* (1998) and adapted by Wysocki and Ladich (2005); therefore only a brief summary of the technique is given here. During the experiments, fish were mildly immobilized with Flaxedil (gallamine triethiodide; Sigma Aldrich Handels GmbH, Vienna, Austria) injected intramuscularly. The dosage used was 9.8 ± 1.8 μg/g fish for *S. umbra*, 4.8 ± 0.6 μg/g for *C. chromis*, and 12.4 ± 0.3 μg/g for *G. cruentatus*. This dosage allowed the fish to retain slight opercular movements during the experiments but without significant myogenic noise to interfere with the recordings. Test subjects were placed secured in an oval plastic tub (diameter 45×30 cm; water depth 12 cm; 1.5 cm layer of sand) which was lined on the inside with acoustically absorbent material (air-filled packing wrap) to reduce resonances and reflections (see Fig. 1 in Wysocki and Ladich, 2002). Fishes were positioned below the water surface (except for the contacting points of the electrodes, which were maximally 1–2 mm above the surface) in the center of the tub. Fish respiration was secured through a simple temperature-controlled (20.7 ± 0.1 °C), gravity-fed water system using a pipette inserted into the subject's mouth. Tissue paper (Kimwipes®) was placed on the fish's head to keep it moist and ensure proper contact of electrodes during experiments.

The AEPs were recorded by using silver wire electrodes (0.38 mm diameter) pressed firmly against the skin. The recording electrode was placed in the midline of the skull over the region of the medulla and the reference electrode cranially between the nares. Shielded electrode leads were attached to the differential input of a Grass P-55 AC preamplifier (Grass Instruments, West Warwick, RI, gain 100×, high-pass at 30 Hz, low-pass at 1 kHz). The plastic tub was

positioned on an air table (TMC Micro-g 63Y540, Technical Manufacturing Corporation, Peabody, MA), which rested on a vibration-isolated concrete plate. The entire setup was enclosed in a walk-in soundproof room, which was constructed as a Faraday cage (interior dimensions: $3.2 \times 3.2 \times 2.4$ m). A ground electrode was placed in the water.

Both presentation of sound stimuli and AEP waveform recordings were achieved using a modular rack-mount system [Tucker-Davis Technologies (TDT) System 3, Gainesville, FL] controlled by a personal computer containing a TDT digital signal processing board and running TDT BIOSIG RP software. Sound stimuli waveforms were created using TDT SIGGEN RP software and played back by a transducer system consisting of two speakers (Fostex PM-0.5 Sub and PM-0.5 MKII, Fostex Corporation, Tokyo, Japan) which was positioned 50 cm above the water surface. Acoustic stimuli consisted of tone bursts (ranging from 100 to 3000 Hz) presented at a repetition rate of 21/s and subsequently at opposite polarities (90° and 270°). The number of cycles in a tone burst was adjusted according to frequency in order to obtain the best compromise between stimulus rapidity (greater rapidity of onset means greater efficacy at generating AEPs) and peak frequency bandwidth (longer duration implies sharper spectral peak) (Silman and Silverman, 1991). The duration of sound stimuli increased from 2 cycles at 100 Hz up to 6 cycles at 3000 Hz. Rise and fall times increased from 1 cycles at 100 Hz to 3 cycles at 3000 Hz. All bursts were gated using a Blackman window. For each test condition, 1000 stimuli (or less, if the response was unambiguous) of each polarity were presented and the corresponding AEPs averaged by BIOSIG RP software to eliminate stimulus artifacts. At each tested frequency, this procedure was performed twice and the AEP traces were overlaid to examine if they were repeatable. The lowest sound pressure level (SPL) where a repeatable AEP trace could be obtained, as determined by overlying replicate traces, was considered the threshold. SPLs were attenuated in 4-dB steps until recognizable and repeatable waveforms could no longer be produced.

C. SPL and particle acceleration measurements

A hydrophone (Brüel & Kjaer 8101, Naerum, Denmark; frequency range: 1 Hz–80 kHz ± 2 dB; voltage sensitivity: −184 re 1 V μPa^{-1}) was placed on the right side of the animals (~1 cm away) in order to control for absolute stimulus SPLs under water in close proximity to the subjects during each experimental session. In order to compare SPL and particle acceleration level for all frequencies tested, a calibrated underwater miniature acoustic pressure-acceleration (p-a) sensor (S/N 2007-001, Applied Physical Sciences Corp., Groton, CT) was placed at the fish's position in the test tub. This p-a sensor (with a frequency bandwidth from 20 Hz to 2 kHz) accelerometer allows the simultaneous recording of sound pressure and particle acceleration. It consists of two built-in units: a piezoelectric, omnidirectional hydrophone (sensitivity: −173.7 dB re 1 V/μPa or −193.7 dB re 1 V/μPa without preamplifier gain) and a bi-directional accelerometer (sensitivity: −137.6 dB re 1 V/μm/s^2 or 12.99 mV/g without preamplifier gain). The sensor was calibrated by the manufacturer in an acoustic waveguide and accounts for buoyancy effects owing to the fact that the sensor is slightly negatively buoyant (McConnell, 2003; McConnell and Jensen, 2006).

TABLE I. SPL and particle acceleration levels in the three orthogonal Cartesian directions and for the magnitude of the three axes combined at each test frequency. SPL—sound pressure level (dB re 1 μPa), L_a—acceleration level (dB re 1 μm/s^2) in the vertical (vert), rostrocaudal (rc), and lateral (lat) axes; Mag *a*—magnitude level of particle acceleration of the three direction combined (*sensu* Casper and Mann, 2006). The magnitude was calculated based on the particle acceleration of each axis in μm/s^2 as $20 \log(\sqrt{(x^2+y^2+z^2)})$.

Frequency (Hz)	SPL	L_a vert	L_a rc	L_a lat	Mag *a*	Ratio SPL-L_a (dB)
100	100	59	52	46	60	40
200	100	61	49	47	62	38
300	100	66	51	51	66	34
500	100	63	53	49	64	36
600	100	63	52	51	64	36
700	100	62	48	50	62	38
1000	100	63	57	58	65	35

Measurements of all stimulus frequencies at various levels, including the hearing threshold levels of the fish, were measured with the acceleration sensor subsequently oriented in all three orthogonal directions. Consistent with previous studies (Casper and Mann, 2006; Horodysky *et al.*, 2008), the *x*-axis was considered to be anterior-posterior along each subject's body, the *y*-axis was considered to be lateral (right-left) relative to the subject, and the *z*-axis to be vertical (i.e., up-down) relative to the subject.

This approach yielded simultaneous measurements of sound pressure and particle acceleration in all three directions over the entire stimulus range, except for the 3000 Hz stimulus used for *S. umbra* because this frequency is above the sensitivity of the p-a sensor. SPLs were calculated in dB rms re 1 μPa and particle acceleration levels (L_a) in dB rms re 1 μm/s^2. These are the international units for sound pressure and particle acceleration according to ISO standards (ISO 1683, 1983).

III. RESULTS

A. Sound stimuli/sound field

Due to the vertical speaker axis, the vertical component (*z*-axis) of particle acceleration had substantially greater amplitudes than the two horizontal axes (*x*- and *y*-axes) at each frequency and attenuation (Table I). Any 4-dB attenuation of the speaker output resulted in a 4-dB decrease of both sound pressure and particle acceleration (in all three orthogonal directions). This was valid for each frequency tested at the position of the fish in the test tub. The proportion between sound pressure and particle acceleration varied among frequencies by up to 6 dB, with similar proportions for 70–200 Hz and the same proportion for 500 and 600 Hz (Table I). For comparative purposes, additional measurements were performed after lowering the sensor by 5 cm. Here again, any 4-dB attenuation of the speaker output re-

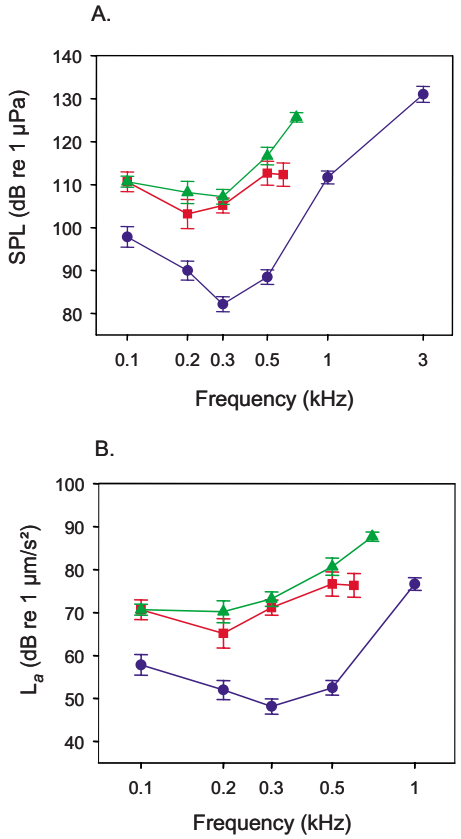


FIG. 1. (Color online) Mean (± S.E.M.) SPLs (A) and particle acceleration levels (B) at auditory thresholds of the three species investigated. triangle—*Gobius cruentatus*; square—*Chromis chromis*; circle—*Sciaena umbra*.

sulted in a 4-dB decrease of both sound pressure and particle acceleration (in all three orthogonal directions) for all frequencies tested. In addition, the vertical component (z-axis) of particle acceleration again had substantially greater amplitudes than the two horizontal axes (x- and y-axes) at each frequency and attenuation.

B. Hearing thresholds

In general, the audiograms expressed in terms of SPL and L_a were very similarly shaped within each species, with sensitivity maxima and minima at the same frequencies, except in the red-mouthed goby, where the best sensitivity of the particle acceleration audiogram was slightly lower (Fig. 1, Table II).

The drum *S. umbra* showed broadest hearing bandwidth and lowest thresholds of all three species tested. Consistent AEPs could be obtained for frequencies up to 3 kHz [Fig. 1(A)] in the drum, whereas no consistent AEPs could be obtained for frequencies higher than 600 Hz in the damselfish or higher than 700 Hz in the goby at the highest level possible (136 dB re 1 μPa).

The brown meagre showed highest auditory sensitivity at 300 Hz regardless of whether the threshold was expressed in terms of SPL or particle acceleration level; this was followed by a steep sensitivity drop-off toward lower and especially higher frequencies. In the damselfish, hearing sensitivity was highest at 200 Hz [Figs. 1(A) and 1(B)]. Hearing sensitivity of the goby was slightly lower than that of the damselfish.

IV. DISCUSSION

A. Sound pressure and particle acceleration field

The vertical speaker axis resulted in a large vertical component (z-axis) of particle acceleration with substantially greater amplitudes than in the two horizontal directions (x- and y-axes) at each frequency and attenuation independently of the water depth. Similar findings have been made in other studies (Casper and Mann 2006—x-axis (anterior-posterior) with underwater speaker anterior to the fish; Horodysky *et al.*, 2008—vertical axis with same experimental setup as in the current study). In all three cases, the particle motion component of the two axes perpendicular to the speaker axis were negligible, and it has thus been proposed to consider only the particle motion values in the speaker axis for plotting audiograms (Casper and Mann, 2006; Horodysky *et al.*, 2008).

Acoustics in small tanks are very complex, and contrary to a freely propagating plane wave in an unbounded acoustic free-field, the relative proportion of particle acceleration and sound pressure at a given frequency is unpredictable. This proportion will vary depending on frequency, tank acoustics, sound source used, and distance to the sound source (Parvulescu, 1967; Rogers and Cox, 1988). Underwater speakers

TABLE II. Mean (± S.E.M) hearing thresholds of the three species investigated expressed in terms of SPLs and magnitude level of particle acceleration of the three directions combined (*sensu* Casper and Mann, 2006).

	Sciaena Umbra		*Chromis chromis*		*Gobius cruentatus*	
Frequency (Hz)	SPL threshold (dB re 1 μPa)	Mag a threshold (dB re 1 μm/s^2)	SPL threshold (dB re 1 μPa)	Mag a threshold (dB re 1 μm/s^2)	SPL threshold (dB re 1 μPa)	Mag a threshold (dB re 1 μm/s^2)
100	97.83 ± 2.39	57.83 ± 2.39	110.67 ± 2.30	70.67 ± 2.30	110.67 ± 1.28	70.67 ± 1.28
200	90.00 ± 2.21	52.00 ± 2.21	103.17 ± 3.41	65.17 ± 3.41	108.17 ± 2.56	70.17 ± 2.56
300	82.17 ± 1.76	48.17 ± 1.76	105.17 ± 1.76	71.71 ± 1.76	107.17 ± 1.70	73.17 ± 1.70
500	88.50 ± 1.69	52.50 ± 1.69	112.67 ± 2.80	76.67 ± 2.80	116.67 ± 2.01	80.67 ± 2.01
600			112.33 ± 2.74	76.33 ± 2.74		
700					125.67 ± 1.09	87.67 ± 1.09
1000	111.67 ± 1.48	76.67 ± 1.48				
3000	131.00 ± 1.86					

inside small tanks are considered to generate high amounts of particle motion relative to pressure as compared to what would be the case in an unbounded medium (Parvulescu, 1967). On the other hand, air speakers reduce the amount of particle motion relative to the pressure component as compared to what would be the case with underwater speakers (Parvulescu, 1964). Therefore, air speakers have been widely used in earlier behavioural fish audiometry studies (Enger, 1966; Jacobs and Tavolga, 1967; Fay, 1969; Popper, 1971). More recently, they are also used in combination with the AEP recording technique [e.g., Kenyon *et al.*, 1998; Ladich and Yan, 1998; Ladich, 1999; Wysocki and Ladich, 2001; 2005, Cordova and Braun, 2007; Horodysky *et al.*, 2008]. In agreement with previous findings (Horodysky *et al.*, 2008), air speakers produce both components of sound during these experiments.

The proportion between both sound characteristics (motion and pressure) differs considerably in various studies. This is because each experimental chamber would have different impedance depending on the construction of the tank and where the measurements were made.

This variance in the relative proportion between particle acceleration and sound pressure shows that it is very important to measure both sound characteristics at the hearing thresholds of the fish. This is especially true when investigating a presumptive "hearing generalist" species, believed to be sensitive only to the particle motion component of sound. On the other hand, the issue of the appropriate sound component seems to be less crucial for describing the hearing range and audiogram shape. Thus, despite frequency-dependent differences in the relative proportion of particle acceleration to sound pressure, the few studies measuring both components have found no significant difference in terms of best hearing range or general audiogram shape. Horodysky *et al.* (2008) reported that the general shapes of the particle acceleration audiograms of six sciaenid species were similar to the sound pressure audiograms, except that they were flatter at low frequencies. A similar conclusion can be drawn by comparing sound pressure and particle acceleration thresholds of two elasmobranch species (Casper and Mann, 2006). Our current results also agree well with these previous findings. Lugli and Fine (2007) characterized stream ambient noise and sound spectra of two types of goby sounds in terms of sound pressure and particle velocity in the field. They also found no noticeable difference in the shapes between the particle velocity spectrum and the sound pressure spectrum in the ambient noise and in one type of goby sound. For the second goby sound type, however, they found larger differences between the pressure and particle velocity spectra. They nonetheless concluded that one of the two sound components is sufficient for characterizing the energy distribution in stream ambient noise and goby sounds.

B. Inter- and intra-family variabilities in hearing abilities

Audiograms have only been assessed in a few representative species of the three fish families investigated in the current study and different techniques (behavioral conditioning versus electrophysiological recordings have been applied). Most of audiograms were expressed in terms of sound pressure.

1. Sciaenidae or drums

Horodysky *et al.* (2008) measured hearing thresholds in six sciaenid species and expressed them in terms of SPL as well as particle acceleration (calculated from the directly measured particle velocity). Similar to Ramcharitar *et al.* (2006b), they did not find significant differences in hearing sensitivity between species with and without anterior swim bladder diverticulae. In both studies, however, thresholds tended to be slightly lower in the former. Surprisingly, a species with swim bladder atrophy in adults was among those with lowest hearing thresholds (i.e., highest sensitivity) at frequencies above 600 Hz. Ramcharitar *et al.* (2006a, 2006b) only measured AEPs up to a frequency of 700 Hz in the spot *Leiostomus xanthurus*, a sciaenid species with no anterior swim bladder extensions, but recorded AEP up to 2000 Hz in the weakfish *Cynoscion regalis*, a species with anterior swim bladder horns extending close to the inner ear. In contrast, Horodysky *et al.* (2008) recorded AEPs in all six species up to 1200 Hz (including the spot), the highest frequency tested, regardless of their swim bladder shape. They might have found responses to higher frequencies in some of those species.

Only one report has been published on hearing in *S. umbra*: Dijkgraaf (1952), using an automated sound production procedure, reported responses of the fish (syn. *Corvina nigra*) up to 1000 Hz or even up to 4000 Hz at very high intensities. Since his thresholds were related to human hearing, no direct comparison to our data set can be made. However, the fish proved to be extremely difficult to train to positive food reward stimuli compared to other species.

Accordingly, their hearing sensitivity and bandwidth have probably been underestimated. When comparing our current data on *S. umbra* to other sciaenid species, the brown meagre shows lowest sound pressure hearing thresholds and the broadest hearing range (up to 3000 Hz). This is surprising because it has a simple, carrot-shaped swim bladder without appendages which is still well developed in adults (Chao, 1986; pers. obs.). One assumption is that sciaenid species without such appendages are more likely to respond solely to the particle motion fields, while species with enhanced swim bladder connections to the inner ear may also be able to detect sound pressure (Horodysky *et al.*, 2008). However, the responses of *S. umbra* to sound stimuli of 3000 Hz cannot be explained by particle motion detection. In the few studies attempting to separate the two components of sound, particle motion sensitivity in teleost fish has never been demonstrated at frequencies above 400 Hz (reviewed in Fay, 1988; Fay and Megela-Simmons, 1999). Moreover, when comparing the audiogram of *S. umbra* to that of another species with no swim bladder extension, *L. xanthurus* (Horodysky *et al.*, 2008), the inter-species differences are much larger in particle acceleration versus pressure audiograms [Figs. 2(A) and 2(B)].

One possibility is that, despite the lack of anterior extensions, the swim bladder can still transmit oscillations in

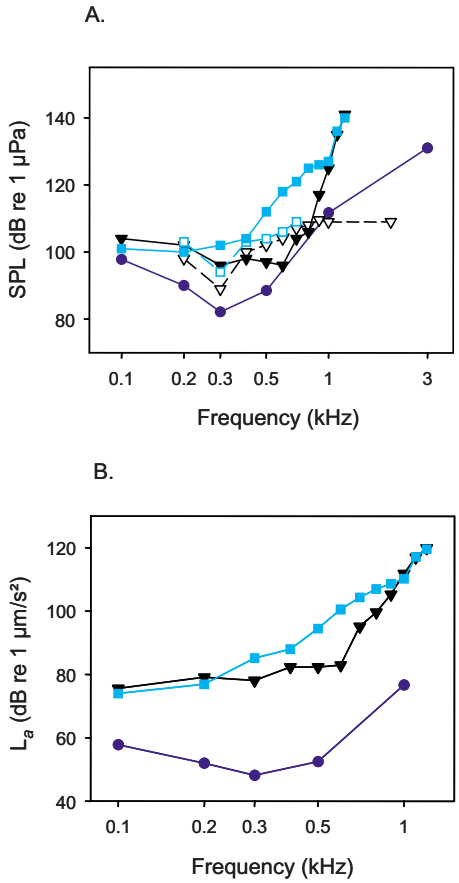


FIG. 2. (Color online) Comparison of (A) SPLs and (B) particle acceleration levels at auditory thresholds of sciaenid species described elsewhere with *S. umbra* (current study). Filled circles—*S. umbra* (current study); filled square—*Leiostomus xanthurus* (Horodysky *et al.*, 2008); open square—*Leiostomus xanthurus* (Ramcharitar *et al.*, 2006b); filled triangle—*Cynoscion regalis* (Horodysky *et al.*, 2008); open triangle—*Cynoscion regalis* (Ramcharitar *et al.*, 2006b).

the sound pressure field to the inner ear. Altogether, the form-function relationship in different sciaenid species is much less clear than in holocentrids, where the degree of swim bladder extension toward the ear nicely matches hearing sensitivity and bandwidth (Coombs and Popper, 1979). This currently precludes extrapolating from the few species investigated to other sciaenid species.

2. Pomacentridae or damselfishes

The highest within-family variation in hearing seems to occur in damselfish. For example, six species within the genus *Stegastes* (syn. *Eupomacentrus*) yielded quite similar behavioral audiograms in terms of absolute pressure sensitivity, hearing range, best frequency, and audiogram shape (Myrberg and Spires, 1980; Kenyon, 1996). In contrast, representatives of the genus *Abudefduf* were up to 50 dB less sensitive to sound when thresholds were measured as SPL (Egner and Mann, 2005; Maruska *et al.*, 2007) by auditory evoked potentials. While all *Stegastes* species had a pronounced sensitivity maximum at 500 Hz, matching well the dominant frequency of their courtship sounds (Myrberg and Spires, 1980), the *Abudefduf* audiograms were much flatter with a region of highest sensitivity between 100 and 400 Hz (Fig. 3). The AEP-pressure audiogram of *Chromis chromis* is intermediate between the two other genera, with a best sensitivity at 200 Hz again, matching the dominant frequency of their sounds (Picciulin *et al.*, 2002) and the narrowest hearing range. Note that responses to higher frequencies might have been recorded if it would have been technically possible to produce levels above 140 dB. While some of the inter-species differences may be explained by methodological differences in audiometry, the large inter-generic differences cannot. So far, no swim bladder diverticulae or other peripheral adaptations have been described in pomacentrids. Myrberg and Spires (1980) found that hearing in *S. dorsopunicans* was governed by particle motion around 100 Hz, but that its hearing was dominated by pressure detection at frequencies of 300 Hz. That experiment involved changing the distance between fish and speaker and thus the proportion between sound pressure and particle motion.

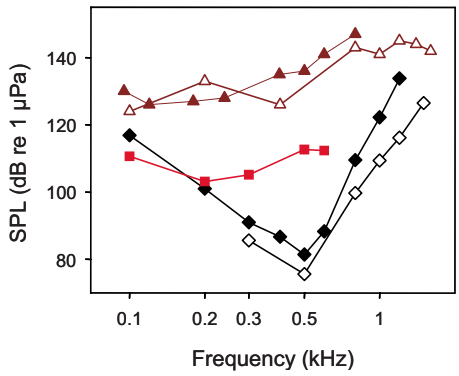


FIG. 3. (Color online) Comparison of SPL audiograms of various pomacentrid species with *C. chromis* (current study). Filled squares—*C. chromis* (current study); filled diamond—*Stegastes adustus* (syn. *Eupomacentrus*, Myrberg and Spires, 1980); open diamond—*Stegastes partitus* (syn. *Pomacentrus*, Kenyon, 1996); filled triangle—*Abudefduf abdominalis* (Maruska *et al.*, 2007); open triangle—*Abudefduf saxatilis* (Egner and Mann, 2005).

Similar to *S. umbra, Stegastes* has no swim bladder extension despite its ability to detect sound pressure. Based on their audiograms, particle motion may be more relevant for *Abudefduf* and *Chromis*, at least in the low frequency range. Given the enormous inter-generic variation in pomacentrid audiograms, it would be desirable to undertake comparative physiological, behavioral, and morphological studies on various pomacentrid species and genera to elucidate the factors behind such differences.

3. Gobiidae or gobies

The red-mouthed goby showed the lowest hearing sensitivity of all three species investigated, regardless of whether expressed as sound pressure or particle acceleration (Fig. 4). Gobies are generally considered to be hearing generalists (Lugli *et al.*, 2003), and some species—though not the *G. cruentatus* (Gil *et al.*, 2002)—even lack a swim bladder. They therefore probably only detect the particle motion component of sound. Currently, only sound pressure audiograms of gobies are available (Dijkgraaf, 1952—behavioral thresholds; Lugli *et al.*, 2003—AEP thresholds). Their shape resembles that of *G. cruentatus*. A particle motion audiogram is available for one representative of the related family Eleotridae, the sleeper goby *Dormitator latifrons* (Lu and Xu,

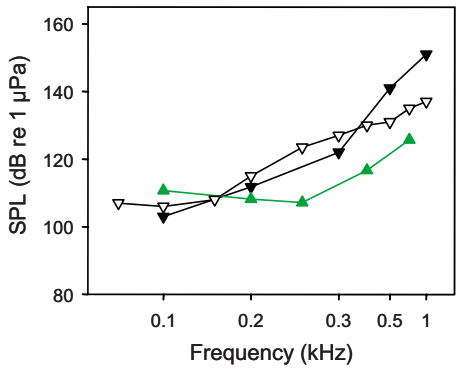


FIG. 4. (Color online) Comparison of sound pressure audiograms of various gobiids with *G. cruentatus* (current study). Filled triangles up—*G. cruentatus* (current study); filled triangles down—*G. niger* (Dijkgraaf, 1952); open triangles down—*P. bonelli* (syn. *P. martensii*, Lugli *et al.*, 2003).

2002—AEP thresholds). Given the high inter-specific variability in other families such as pomacentrids, the audiogram of *D. latifrons* looks surprisingly similar to the particle acceleration audiogram of *G. cruentatus* at comparable frequencies when calculated as particle acceleration level in dB re 1 $\mu m/s^2$ (Fig. 5).

V. CONCLUSIONS

In summary, the three investigated Adriatic species showed quite different hearing sensitivity and hearing bandwidth. Additionally, intra-family comparisons showed large variations at least among pomacentrids, variations that do not always coincide with morphological adaptations (accessory hearing structures). This shows that among species extrapolations must be interpreted with extreme caution and that it is better to directly investigate the target species of any environmental acoustic impact study. The standard setup used in most fish hearing studies is unsuitable to determine whether a fish reacts to the particle motion or the pressure component of sound below approximately 400 Hz. Special equipment that separates both sound components would be necessary for this purpose. In addition, sound in the natural environment always consists of both components, whose relative proportions may vary depending on distance to the sound source and the acoustic field in the environment. Therefore, the best option if information on absolute sensitivity to sound of a given fish species is required is to measure both components at the hearing threshold.

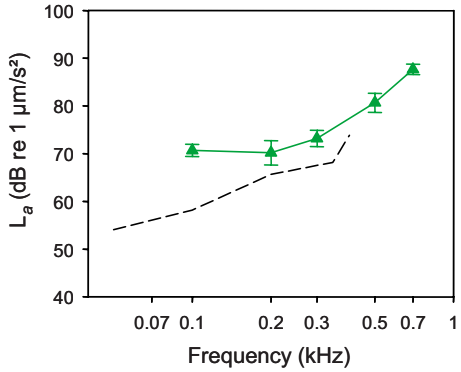


FIG. 5. (Color online) Comparison of particle acceleration levels at auditory thresholds of *G. cruentatus* (Gobiidae; filled triangles up, current study) and the sleeper goby *Dormitator latifrons* (Eleotridae; dashed line). Threshold values for *D. latifrons* calculated from Lu and Xu (2002).

ACKNOWLEDGMENTS

We would like to thank Jim McConnell for developing the miniature p-a sensor for our purpose, for many helpful discussions, and for his patience in answering all our questions; Valter Žiža for help in catching fish; Enrico Vinzi and Alessandro Di Bartolomeo for help with fish transport to Vienna; Michael Pollirer and the Department of Marine Biology for providing salt water for fish keeping and Maurizio Spoto for the permission to record noise in the Natural Marine Reserve of Miramare. We thank Michael Stachowitsch for professional scientific English proofreading. This research was supported by the Austrian Science Fund (FWF Grant No. 17263 to F.L.) and the Italian Ministry for Environment, Territory and Sea.

Banner, A. (**1967**). "Evidence of sensitivity to acoustic displacements in the lemon shark, *Negaprion brevirostris* (Poey)," in *Lateral-Line Detectors*, edited by P. H. Cahn (Indiana U.P., Bloomington, IN), pp. 265–273.

Bass, A. H., and McKibben, J. R. (**2003**). "Neural mechanisms and behaviors for acoustic communication in teleost fish," Prog. Neurobiol. **69**, 1–26.

Blaxter, J. H. S., Denton, E. J., and Gray, J. A. B. (**1981**). "Acousticolateralis system in clupeid fishes," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 39–56.

Buwalda, R. J. (**1981**). "Segregation of directional and nondirectional acoustic information in the cod," in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 139–171.

Cahn, P. H., Siler, W., and Wodinsky, J. (**1969**). "Acoustico-lateralis system of fishes. Tests of pressure and particle-velocity sensitivity in grunts, *Haemulon sciurus* and *Haemulon parrai*," J. Acoust. Soc. Am. **46**, 1572–1578.

Casper, B. M., and Mann, D. A. (**2006**). "Evoked potential audiograms of the nurse shark (*Ginglymostoma cirratum*) and the yellow stingray (*Urobatis jamaicensis*)," Environmental Biology of Fishes **76**, 101–108.

Chao, L. N. (**1986**). "Sciaenidae," in *Fishes of the North-Eastern Atlantic and the Mediterranean*, edited by P. J. P. Whitehead, M.-L. Bauchot, J.-C. Hureau, J. Nielsen, and E. Tortonese (Unesco, Paris), Vol. **1**, pp. 865–874.

Chapman, C. J., and Hawkins, A. D. (**1973**). "Field study of hearing in cod, *Gadus morhua* L," J. Comp. Physiol. **85**, 147–167.

Chapman, C. J., and Sand, O. (**1974**). "Field studies of hearing in 2 species of flatfish *Pleuronectes platessa* (L) and *Limanda limanda* (L) (Family Pleuronectidae)," Comp. Biochem. Physiol. **47**, 371–385.

Codarin, A., Wysocki, L. E., Ladich, F., and Picciulin, M. (**2009**). "Effects of ambient and boat noise on hearing and communication in three fish species living in a marine protected area (Miramare, Italy)," Mar. Pollution Bull.

Coombs, S., and Popper, A. N. (**1979**). "Hearing differences among Hawaiian squirrelfish (Family Holocentridae) related to differences in the peripheral auditory system," J. Comp. Physiol. **132**, 203–207.

Cordova, M. S., and Braun, C. B. (**2007**). "The use of anesthesia during evoked potential audiometry in goldfish (*Carassius auratus*)," Brain Res. **1153**, 78–83.

Dijkgraaf, S. (**1952**). "Über die Schallwahrnehmung bei Meeresfischen (On sound reception in marine fishes)," Zeitschrift für Vergleichende Physiologie **34**, 104–122.

Egner, S. A., and Mann, D. A. (**2005**). "Auditory sensitivity of sergeant major damselfish *Abudefduf saxatilis* from post-settlement juvenile to adult," Mar. Ecol. Prog. Ser. **285**, 213–222.

Enger, P. S. (**1966**). "Acoustic threshold in goldfish and its relation to sound source distance," Comp. Biochem. Physiol. **18**, 859–868.

Enger, P. S., and Andersen, R. (**1967**). "An electrophysiological field study of hearing in fish," Comp. Biochem. Physiol. **22**, 527–538.

Fay, R. R. (**1969**). "Behavioural audiogram for the goldfish," J. Aud Res. **9**,

112–121.
Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Fay Hill Assoc., Winnetka, IL).
Fay, R. R., and Edds-Walton, P. L. (**1997**). "Directional response properties of saccular afferents of the toadfish, *Opsanus tau*," Hear. Res. **111**, 1–21.
Fay, R. R., and Megela-Simmons, A. (**1999**). "The sense of hearing of fishes and amphibians," in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 269–318.
Fletcher, L. B., and Crawford, J. D. (**2001**). "Acoustic detection by sound-producing fishes (Mormyridae): The role of gas-filled tympanic bladders," J. Exp. Biol. **204**, 175–183.
Gil, F., Borges, R., Faria, C., and Goncalves, E. J. (**2002**). "Early development of the red mouthed goby, *Gobius cruentatus* (Pisces: Gobiidae)," J. Mar. Biol. Assoc. U.K. **82**, 161–163.
Hawkins, A. D. (**1993**). "Underwater sound and fish behaviour," in *Behaviour of Teleost Fishes*, edited by T. J. Pitcher (Chapman and Hall, London), pp. 129–169.
Hawkins, A. D., and Myrberg, A. A. (**1983**). "Hearing and sound communication underwater," in *Bioacoustics: A Comparative Approach*, edited by B. Lewis (Academic, London), pp. 347–405.
Horodysky, A. Z., Brill, R. W., Fine, M. L., Musick, J. A., and Latour, R. J. (**2008**). "Acoustic pressure and particle motion thresholds in six sciaenid fishes," J. Exp. Biol. **211**, 1504–1511.
Jacobs, D. W., and Tavolga, W. N. (**1967**). "Acoustic intensity limens in the goldfish," Anim. Behav. **15**, 324–335.
ISO 1683 (**1983**). "Acoustics—Preferred reference quantities for acoustic levels," International Organization for Standardization.
Kalmijn, A. J. (**1988**). "Hydrodynamic and acoustic field detection," in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, A. N. Popper, and W. N. Tavolga, (Springer-Verlag, New York), pp. 83–130.
Kenyon, T. N. (**1996**). "Ontogenetic changes in the auditory sensitivity of damselfishes (Pomacentridae)," J. Comp. Physiol. [A] **179**, 553–561.
Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). "A comparative study of hearing ability in fishes: the auditory brainstem response approach," J. Comp. Physiol. [A] **182**, 307–318.
Ladich, F. (**1999**). "Did auditory sensitivity and vocalization evolve independently in otophysan fishes?," Brain Behav. Evol. **53**, 288–304.
Ladich, F., and Wysocki, L. E. (**2003**). "How does tripus extirpation affect auditory sensitivity in goldfish?," Hear. Res. **182**, 119–129.
Ladich, F., and Popper, A. N. (**2004**). "Parallel evolution in fish hearing organs," in *Evolution of the Vertebrate Auditory System*, edited by G. Manley, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 95–127.
Ladich, F., and Yan, H. Y. (**1998**). "Correlation between auditory sensitivity and vocalization in anabantoid fishes," J. Comp. Physiol. [A] **182**, 737–746.
Lechner, W., and Ladich, F. (**2008**). "Size matters: Diversity in swimbladders and Weberian ossicles affects hearing in catfishes," J. Exp. Biol. **211**, 1681–1689.
Lu, Z., and Xu, Z. (**2002**). "Effects of saccular otolith removal on hearing sensitivity of the sleeper goby (*Dormitator latifrons*)," J. Exp. Biol. **188**, 595–602.
Lugli, M., and Fine, M. L. (**2007**). "Stream ambient noise, spectrum and propagation of sounds in, the goby *Padogobius martensii*: Sound pressure and particle velocity," J. Acoust. Soc. Am. **122**, 2881–2892.
Lugli, M., Yan, H. Y., and Fine, M. L. (**2003**). "Acoustic communication in two freshwater gobies: the relationship between ambient noise, hearing thresholds and sound spectrum," J. Comp. Physiol. [A] **189**, 309–320.
Maruska, K. P., Boyle, K. S., Dewan, L. R., and Tricas, T. C. (**2007**). "Sound production and spectral hearing sensitivity in the Hawaiian sergeant damselfish, *Abudefduf abdominalis*," J. Exp. Biol. **210**, 3990–4004.
McConnell, J. A. (**2003**). "Analysis of a compliantly suspended acoustic velocity sensor," J. Acoust. Soc. Am. **113**, 1395–1405.
McConnell, J. A., and Jensen, S. C. (**2006**). "Development of a miniature pressure-acceleration probe for bioacoustic applications," J. Acoust. Soc. Am. **119**, 3446.
Myrberg, A. A., and Spires, J. Y. (**1980**). "Hearing in damselfishes—An analysis of signal-detection among closely related species," J. Comp. Physiol. **140**, 135–144.
Parvulescu, A. (**1964**). "Problems of propagation and processing," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 87–100.
Parvulescu, A. (**1967**). "The acoustics of small tanks," in *Marine Bio-Acoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 7–13.
Picciulin, M., Costantini, M., Hawkins, A. D., and Ferrero, E. A. (**2002**). "Sound emissions of the Mediterranean damselfish *Chromis chromis* (Pomacentridae)," Bioacoustics **12**, 236–238.
Picciulin, M., Codarin, A., and Spoto, M. (**2008**). "Characterization of small boat noises compared with the chorus of *Sciaena umbra* (Sciaenidae)," Bioacoustics **17**, pp. 210–212.
Popper, A. N. (**1971**). "The effects of size on auditory capacities of the goldfish," J. Aud. Res. **11**, 239–247.
Popper, A. N., and Fay, R. R. (**1999**). "The auditory periphery in fishes," in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 43–100.
Ramcharitar, J., Gannon, D. P., and Popper, A. N. (**2006a**). "Bioacoustics of fishes of the family Sciaenidae (Croakers and Drums)," Trans. Am. Fish. Soc. **135**, 1409–1431.
Ramcharitar, J. U., Higgs, D. M., and Popper, A. N. (**2006b**). "Audition in sciaenid fishes with different swim bladder-inner ear configurations," J. Acoust. Soc. Am. **119**, 439–443.
Rogers, P. H., and Cox, H. (**1988**). "Underwater sound as a biological stimulus," in *Sensory Biology of Aquatic Animals*, edited by J. Atema, R. R. Fay, and A. N. Popper (Springer-Verlag, New York), pp. 131–149.
Sand, O., and Enger, P. S. (**1973**). "Evidence for an auditory function of swimbladder in cod," J. Exp. Biol. **59**, 405–414.
Schneider, H. (**1941**). "Die Bedeutung der Atemhöhle der Labyrinthfische für ihr Hörvermögen (The significance of the respiratory cavity for hearing in labyrinth fishes)," Zeitschrift für Vergleichende Physiologie **29**, 172–194.
Silman, S., and Silverman, C. A. (**1991**). *Auditory Diagnosis: Principles and Applications* (Academic Press, San Diego).
Stipetić, E. (**1939**). "Über das Gehörorgan der Mormyriden (On the hearing organ of mormyrids)," Zeitschrift für Vergleichende Physiologie **26**, 740–752.
Wysocki, L. E., and Ladich, F. (**2001**). "The ontogenetic development of auditory sensitivity, vocalization and acoustic communication in the labyrinth fish *Trichopsis vittata*," J. Comp. Physiol. [A] **187**, 177–187.
Wysocki, L. E., and Ladich, F. (**2002**). "Can fishes resolve temporal characteristics of sounds? New insights using auditory brainstem responses," Hear. Res. **169**, 36–46.
Wysocki, L. E., and Ladich, F. (**2005**). "Hearing in fishes under noise conditions," J. Assoc. Res. Otolaryngol. **6**, 28–36.

Simulation of shear wave propagation in a soft medium using a pseudospectral time domain method

Cécile Bastard,[a] Jean-Pierre Remeniéras, and Samuel Callé
Inserm U930, CNRS ERL 3106, Université François Rabelais de Tours, 10 boulevard Tonnellé, 37032 Tours, France

Laurent Sandrin
Department of Research and Development, Echosens, 153 avenue d'Italie, 75013 Paris, France

(Received 6 February 2009; revised 15 June 2009; accepted 27 July 2009)

Elastography applications require the use of efficient models to simulate the propagation of shear waves in soft media such as human tissues. These models are needed to improve understanding of the measured displacement field, to reconstruct the viscoelasticity of heterogeneous tissues, and to test inversion algorithms. This paper reports a numerical model based on a pseudospectral time domain method developed to simulate shear and compression wave propagation in an axisymmetric heterogeneous viscoelastic medium. This model was adapted to the study of soft tissues where the ratio between the compression and the shear wave velocity was about a thousand and validated in the homogeneous situation by comparison with an analytical model based on elastodynamic Green's functions. Displacements obtained experimentally using transient elastography are presented, compared with simulation results, and discussed.
 [DOI: 10.1121/1.3206585]

PACS number(s): 43.80.Qf, 43.20.Jr [PB] Pages: 2108–2116

I. INTRODUCTION

A. Background

Elastography is a branch of tissue characterization which encompasses a wide range of techniques used to measure or describe the elastic properties of tissues. This method emerged 20 years ago with the work of Krouskop *et al.*[1] and Lerner *et al.*[2] who used Doppler ultrasound to measure the displacements induced in soft tissues by external vibrations. Several techniques have been developed depending on the mechanical excitation of the tissues and the technique used to measure the mechanical response. Static elastography has been developed by Ophir *et al.*[3] since 1991. In this technique the tissues are insonified before and after slight compression, and the speckle distortion due to the compression is related to the stiffness of the medium. Dynamic elastography is based on the study of the propagation of low-frequency shear waves which are generated either monochromatically, as in the case of magnetic resonance elastography[4] (MRE) or transiently, as in the case of transient elastography. The impulse excitation used in transient elastography is generated either by an external vibrator or by a transient ultrasound radiation force. Several groups are currently working in the area of ultrasound radiation force.[5] Using acoustic radiation force impulse (ARFI) imaging, Nightingale *et al.*[6] generated a localized radiation force impulse in tissues using a conventional diagnostic ultrasound scanner. In supersonic shear imaging (SSI), shear waves are generated by moving the source created by focused ultrasound beams at a supersonic speed.[7]

In vivo measurements have been obtained using these different elastographic techniques on a variety of organs such as the breast, prostate, and liver. In the latter case, elastography has been clinically proven to be a relevant tool to evaluate liver fibrosis.[8] In their study, Sandrin *et al.* used a Fibroscan® device (Echosens, Paris, France) composed of a probe, a dedicated electronic system, and a control unit. A 3.5 MHz ultrasound transducer was mounted on the probe and used as a piston to generate a 50 Hz shear wave whose propagation in the liver was monitored at a repetition frequency of 6000 Hz. The shear wave velocity was then computed from strain rate images, assuming that the medium was purely elastic, isotropic, and homogeneous. Young's modulus can be estimated using the equation $E=3\rho c_s^2$.

B. Aims

The development of elastographic techniques is now being focused on heterogeneous organs. Many disorders such as cancers result in heterogeneous elasticity patterns or the presence of inclusions whose elasticity is different from the surrounding normal tissue. Using transient elastography, Sandrin *et al.*[9] showed that it is feasible to distinguish between two areas of differing elasticity *in vitro*, and Nightingale *et al.*[10] demonstrated that ARFI is able to image local variations in the mechanical response of tissues. In contrast to static elastography, several dynamic elastography techniques are quantitative, and the elasticity given by different modalities such as transient elastography, ARFI, SSI, and MRE should be the same, whatever the method of signal processing and reconstruction algorithm used. This is why there is considerable interest in modeling the propagation of shear waves in human tissues to test reconstruction algo-

[a]Also at Research and Development Department, Echosens, 153 avenue d'Italie, 75013 Paris, France. Electronic mail: cecile.bastard@echosens.com

rithms and to ensure that the image processing does not bias the elasticity value.

In transient elastography two elastic waves, i.e., a compression wave and a shear wave, are observed in soft tissues, their velocities being around 1500 and 1.5 m/s, respectively. Several techniques, including finite element and finite difference methods, have been used to investigate this problem. Palmeri *et al.*[11] applied a finite element method to study the displacements induced by an acoustic radiation force impulse. The focus of the study reported here was to develop a numerical model appropriate for modeling of the displacements induced by an external vibrator in transient elastography. The geometry of the source and the temporal profile of the excitation were taken into account and a pseudospectral time domain (PSTD) method was used to solve the elastic wave propagation equation. As in finite difference methods, the temporal increment of the numerical model is linked to the fastest wave velocity. In this paper, a technique to separate the shear and compression components of the source force is introduced. It allows to simulate the shear wave propagation independently from the compression wave, which simplifies the wave equation and allows to study the shear displacements without taking into account compression wave effects. The separation of the sources also permits to decrease the computation time due to the considerable discrepancy between the shear wave velocity and the compression wave velocity. The PSTD method chosen to achieve this simulation has the advantage to require only two cells per wavelength[12] compared to at least 10 for second order finite difference time domain (FDTD) methods. PSTD methods were first introduced by Kreiss and Oliger[13] and used in the field of electromagnetism as an alternative to FDTD methods.[14] The principle underlying PSTD methods is projection of the propagation equations in the Fourier domain in order to compute the spatial derivatives analytically. Fast Fourier transform (FFT) and inverse fast Fourier transform are used to shift between the spatial domain and the Fourier domain. Several PSTD methods have already been developed to simulate the propagation of elastic waves in a heterogeneous medium and mostly applied to ultrasonic waves.[15,16]

The first part of this paper presents a comparison between a complete numerical model and an analytical model in the case of a point source; the second part deals with the case of soft tissues and introduces a technique to separate shear and compression components of a force source. In the last part, experimental results are compared with simulated displacements.

II. COMPLETE MODEL

A. Numerical model

The propagation of elastic waves in a viscoelastic, linear, and isotropic medium is described by the following equation:

$$\rho \frac{\partial \vec{v}}{\partial t} = \vec{\nabla} \cdot (\bar{\bar{T}} + \bar{\bar{\tau}}) + \vec{f}, \tag{1}$$

where ρ, $\vec{v}$, $\bar{\bar{T}}$, $\bar{\bar{\tau}}$, and $\vec{f}$ are the density, the velocity, the stress tensor, the viscous stress tensor, and the force applied to the medium, respectively.

The model considered was axisymmetric, and all the derivatives as a function of θ were then equal to zero.

Stresses were obtained from strains using Hooke's law,

$$\bar{\bar{T}} = \lambda \, \mathrm{tr}(\bar{\bar{\epsilon}}) \bar{\bar{1}} + 2\mu \bar{\bar{\epsilon}}, \tag{2}$$

where tr stands for the trace and $\bar{\bar{1}}$ for the identity tensor.

The strain tensor $\bar{\bar{\epsilon}} = \frac{1}{2}(\vec{\nabla}\vec{u} + (\vec{\nabla}\vec{u})^T)$ is expressed in cylindrical coordinates as

$$\bar{\bar{\epsilon}} = \begin{pmatrix} \frac{\partial u_r}{\partial r} & 0 & \frac{1}{2}\left(\frac{\partial u_r}{\partial z} + \frac{\partial u_z}{\partial r}\right) \\ 0 & \frac{u_r}{r} & 0 \\ \frac{1}{2}\left(\frac{\partial u_r}{\partial z} + \frac{\partial u_z}{\partial r}\right) & 0 & \frac{\partial u_z}{\partial z} \end{pmatrix}. \tag{3}$$

Combining Eqs. (2) and (3), and taking the temporal derivative, yields

$$\begin{aligned} \frac{\partial T_{rr}}{\partial t} &= (\lambda + 2\mu)\frac{\partial v_r}{\partial r} + \lambda\left(\frac{v_r}{r} + \frac{\partial v_z}{\partial z}\right), \\ \frac{\partial T_{zz}}{\partial t} &= (\lambda + 2\mu)\frac{\partial v_z}{\partial z} + \lambda\left(\frac{v_r}{r} + \frac{\partial v_r}{\partial r}\right), \\ \frac{\partial T_{\theta\theta}}{\partial t} &= (\lambda + 2\mu)\frac{v_r}{r} + \lambda\left(\frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z}\right), \\ \frac{\partial T_{rz}}{\partial t} &= \mu\left(\frac{\partial v_r}{\partial z} + \frac{\partial v_z}{\partial r}\right). \end{aligned} \tag{4}$$

Introducing viscosity requires the choice of an appropriate rheological model. As in previous studies,[17,18] the Voigt model was chosen to introduce viscosity in the wave equations, leading to the following expression of the viscous stress tensor:

$$\bar{\bar{\tau}} = \eta_p \, \mathrm{tr}(\bar{\bar{S}}') \bar{\bar{1}} + 2\eta_s \bar{\bar{S}}', \tag{5}$$

where η_p, η_s, $\bar{\bar{1}}$, and $\bar{\bar{S}}'$ are the second viscosity coefficient, the shear viscosity, the identity tensor, and the deformation-rate tensor, respectively. The second viscosity coefficient was computed from the shear viscosity η_s and from the bulk viscosity η_v using the relation

$$\eta_p = \eta_v - \tfrac{2}{3}\eta_s. \tag{6}$$

Expressing $\bar{\bar{S}}' = \frac{1}{2}(\vec{\nabla}\vec{v} + (\vec{\nabla}\vec{v})^T)$ in cylindrical coordinates in Eq. (5) provides

$$\begin{aligned} \tau_{rr} &= (\eta_p + 2\eta_s)\frac{\partial v_r}{\partial r} + \eta_p\left(\frac{v_r}{r} + \frac{\partial v_z}{\partial z}\right), \\ \tau_{zz} &= (\eta_p + 2\eta_s)\frac{\partial v_z}{\partial z} + \eta_p\left(\frac{v_r}{r} + \frac{\partial v_r}{\partial r}\right), \end{aligned} \tag{7}$$

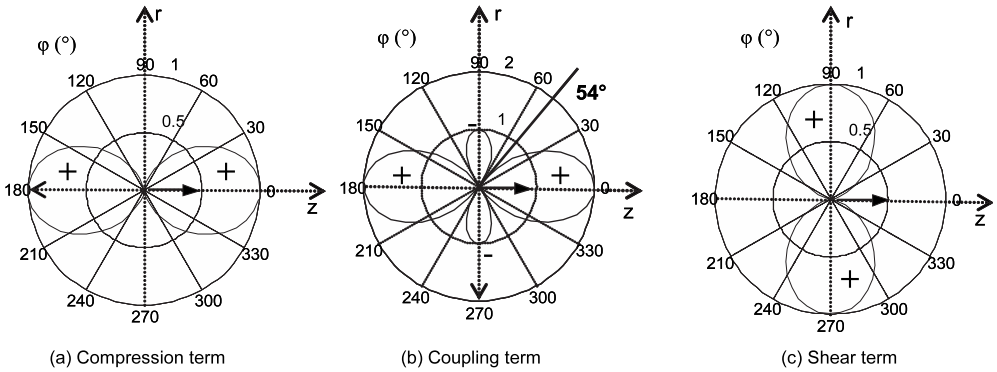


FIG. 1. Angular directivity of the z-component of the displacement for a point source applied in the $\vec{e}_z$ direction ($\varphi=0°$) in the plane $\theta=0°$ as a function of z and r.

$$\tau_{\theta\theta} = (\eta_p + 2\eta_s)\frac{v_r}{r} + \eta_p\left(\frac{\partial v_r}{\partial r} + \frac{\partial v_z}{\partial z}\right),$$

$$\tau_{rz} = \eta_s\left(\frac{\partial v_r}{\partial z} + \frac{\partial v_z}{\partial r}\right).$$

Taking into account the axisymmetric conditions, the divergence of the tensor $\bar{\bar{M}} = \bar{\bar{T}} + \bar{\bar{\tau}}$ in the referential (e_r, e_θ, e_z) is

$$\vec{\nabla}\cdot\bar{\bar{M}} = \begin{pmatrix} \dfrac{\partial M_{rr}}{\partial r} + \dfrac{1}{r}(M_{rr} - M_{\theta\theta}) + \dfrac{\partial M_{rz}}{\partial z} \\ 0 \\ \dfrac{\partial M_{rz}}{\partial r} + \dfrac{1}{r}M_{rz} + \dfrac{\partial M_{zz}}{\partial z} \end{pmatrix}. \tag{8}$$

The system to solve to compute the velocity is then

$$\frac{\partial v_r}{\partial t} = \frac{1}{\rho}\left(\frac{\partial M_{rr}}{\partial r} + \frac{\partial M_{rz}}{\partial z} + \frac{M_{rr} - M_{\theta\theta}}{r}\right) + \frac{f_r}{\rho},$$

$$\frac{\partial v_z}{\partial t} = \frac{1}{\rho}\left(\frac{\partial M_{rz}}{\partial r} + \frac{\partial M_{zz}}{\partial z} + \frac{M_{rz}}{r}\right) + \frac{f_z}{\rho}, \tag{9}$$

where $\vec{f} = f_r\vec{e}_r + f_z\vec{e}_z$.

The numerical implementation of the model was based on the discretization of Eqs. (4), (7), and (9). The spatial derivatives were computed in the Fourier domain. Matched absorbing boundary layers called perfectly matched layers[19] were used to avoid reflections on the numerical grid boundaries and to counter the wrap-around effect of the FFT. The temporal evolution of the model was computed with a staggered fourth order Adams–Bashforth method.[20]

B. Comparison with analytical models

To validate the numerical model proposed, its results were compared with the displacements given by an analytical model based on Green's functions used by Aki and Richards[21–23] and adapted to a viscoelastic situation.[17] In this model, the displacement induced in an infinite homogeneous isotropic medium was divided into three terms, i.e., a compression term, a shear term, and a coupling term.

The role of the coupling term was previously detailed by Sandrin *et al.*[22] The coupling term of the displacement has both shear and compression components. But while the shear and compression terms attenuate as r^{-1}, the behavior of the coupling term is more complex. It attenuates as r^{-1} in a very near field where $r<\frac{3}{4}\lambda_s$ and as r^{-2} for $r>\frac{3}{4}\lambda_s$, where λ_s is the shear wavelength. As r^{-1} becomes dominant over r^{-2} for $r\to\infty$, the shear and compression terms are called far-field terms. As r^{-2} is dominant over r^{-1} as $r\to 0$, the coupling term is called near-field term.

The radiation pattern of the component of these three terms along z for a force in the $\vec{e}_z$ direction is shown in Fig. 1. The directivity of the compression term is a function of $\cos^2\theta$, as shown in Fig. 1(a), whereas the shear term directivity is a function of $\sin^2\theta$ [Fig. 1(c)]. On the other hand, the magnitude of the coupling term is dependent on $(3\cos^2\theta - 1)$ and is equal to zero for $\theta=54°$ [Fig. 1(b)]. Figure 2 is a simulation of the response of the medium to a point source located in $O(0,0)$, the force being a 50 Hz apodized sine applied in the z direction. This figure shows the propagation of the compression and shear waves in a heterogeneous elastic medium. In order to observe both compression and shear waves, the shear wave speed and the compression wave speed were chosen to be very close to each other ($c_s = 1$ m/s and $c_p = 1.5$ m/s), which is not realistic for biological media. As expected, the shear wave amplitude was highest in the direction perpendicular to the excitation and the compression wave amplitude was highest in the direction of the excitation.

The amplitude of the displacement is represented as a

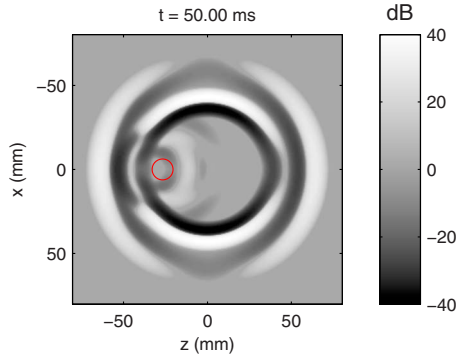


FIG. 2. (Color online) Component of the displacement along z computed at time $t=50$ ms in a purely elastic medium ($c_s=1$ m/s and $c_p=1.5$ m/s) containing a stiffer inclusion ($c_s=1.41$ m/s) represented by a circle. The source is a force applied in $O(0,0)$ in the z direction.

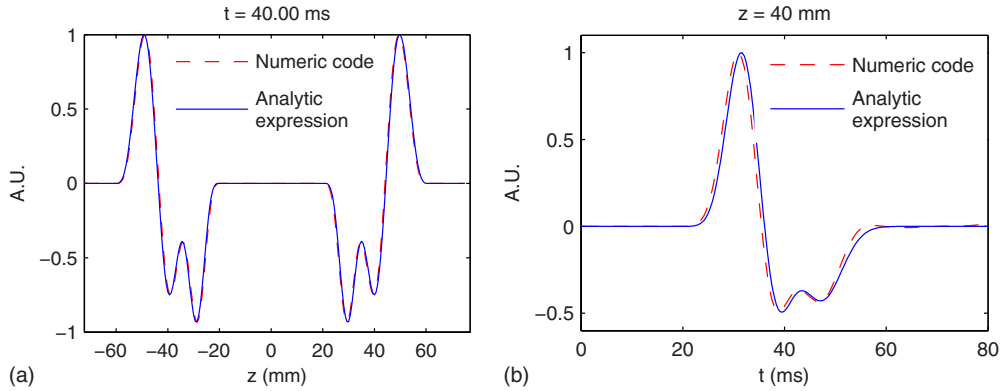


FIG. 3. (Color online) z-component of the displacement along z in a viscoelastic homogeneous medium with $c_s=1$ m/s, $c_p=1.5$ m/s, $\eta_s=0.5$ Pa s, and $\eta_p=0$ Pa s. The point source was located in $O(0,0)$ and the force was a 50 Hz sine applied in the z direction.

function of depth on the z-axis at time $t=40$ ms for a homogeneous viscoelastic medium with a shear viscosity of 0.5 Pa s and a second viscosity coefficient equal to zero [Fig. 3(a)]. The excitation is the same as that used in Fig. 2. Numerical and analytical simulations are in very good agreement, and at time $t=40$ ms the two waves are partly overlapping. The fastest wave, the compression wave, is clearly apparent at depth $z=\pm 60$ mm. The second wave is mainly due to the near-field term. Indeed, as the source is a point source, the shear displacement is equal to zero on the z-axis. The amplitude of the displacement as a function of time at location $A(r=0$ mm$,z=40$ mm) and the analytical displacement in Fig. 3(b) are overlaid, showing both near field and compression terms. In Fig. 3, the displacements have been normalized with respect to their spatio-temporal maximum.

C. Application to heterogeneous media

The medium represented in Fig. 2 was heterogeneous and contained a spherical inclusion (represented by a circle) centered at depth $z=-30$ mm where the shear modulus was higher than that of the surrounding medium ($c_s=1$ m/s in the background and $c_s=1.41$ m/s in the inclusion). The propagation of the compression wave was not disturbed by the inclusion because there was no change in the compression wave speed whereas the shear wave was distorted in the inclusion and reflected at its edges.

III. SOFT TISSUE SITUATION

A. Source separation

A model to simulate both shear wave and compression wave propagations was described in Sec. II. In this section the method used to separate the sources is presented. As the temporal increment of the numerical model is linked to the fastest wave speed, i.e., the compression wave speed, the separation of the sources permits to increase the temporal increment and thus to reduce the computational time to simulate shear wave propagation. The method used to achieve this separation was that introduced by Aki and Richards,[21] consisting of expressing the displacement and the body force with Helmholtz potentials:

$$\vec{f}=\vec{\nabla}\phi_f+\vec{\nabla}\times\vec{\psi}_f,\quad \vec{u}=\vec{\nabla}\phi_u+\vec{\nabla}\times\vec{\psi}_u, \tag{10}$$

where ϕ_f and $\vec{\psi}_f$ are the body force Helmholtz potentials and ϕ_u and $\vec{\psi}_u$ the displacement Helmholtz potentials.

The displacement was divided into two terms, $\vec{u}_c=\vec{\nabla}\phi_u$ which is irrotational ($\vec{\nabla}\times\vec{u}_c=\vec{0}$) and related to the compression wave, and $\vec{u}_s=\vec{\nabla}\times\vec{\psi}_u$ which is solenoidal ($\vec{\nabla}\cdot\vec{u}_s=0$) and related to the shear wave.

The source was also separated into a compression source ($\vec{f}_c=\vec{\nabla}\phi_f$) and a shear source ($\vec{f}_s=\vec{\nabla}\times\vec{\psi}_f$).

The Navier's equation was then rewritten as follows:

$$\rho\frac{\partial^2(\vec{u_c}+\vec{u_s})}{\partial t^2}=(\bar{\lambda}+2\bar{\mu})\vec{\nabla}\vec{\nabla}\cdot(\vec{u_c})-\bar{\mu}\vec{\nabla}\times\vec{\nabla}\times\vec{u_s}+\vec{f}_c+\vec{f}_s. \tag{11}$$

Decoupling the previous equation gives the system

$$\begin{aligned}\rho\frac{\partial^2\vec{u_c}}{\partial t^2}&=(\bar{\lambda}+2\bar{\mu})\vec{\nabla}(\vec{\nabla}\cdot\vec{u_c})+\vec{f}_c,\\ \rho\frac{\partial^2\vec{u_s}}{\partial t^2}&=-\bar{\mu}\vec{\nabla}\times\vec{\nabla}\times\vec{u_s}+\vec{f}_s.\end{aligned} \tag{12}$$

where $\bar{\lambda}=\lambda+\eta_p(\partial/\partial t)$ and $\bar{\mu}=\mu+\eta_s(\partial/\partial t)$.

It is interesting that, due to the irrotational properties of $\vec{u}_c$ and to the solenoidal properties of $\vec{u}_s$, the two equations of the previous system correspond to

$$\begin{aligned}&\frac{\partial^2\vec{u_i}}{\partial t^2}=v_i^2\Delta\vec{u_i}+\vec{f}_i/\rho\quad\text{with } i=c,s,\\ &v_s^2=\frac{\bar{\mu}}{\rho}\text{ and }v_p^2=\frac{\bar{\lambda}+2\bar{\mu}}{\rho}.\end{aligned} \tag{13}$$

This equation describes the propagation of a wave at a speed c_i in response to a force $\vec{f}_i$. The difference between the propagation of a compression and a shear wave therefore originates from the velocity of the wave but mainly from the spatial pattern of the source $\vec{f}_i$.

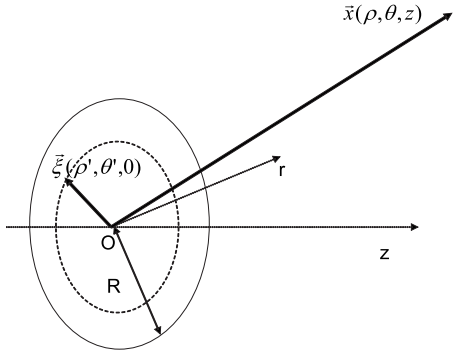


FIG. 4. Cylindrical coordinate system used to solve the Poisson equation. The sources are in $\vec{\xi}$ and the observation points are in $\vec{x}$.

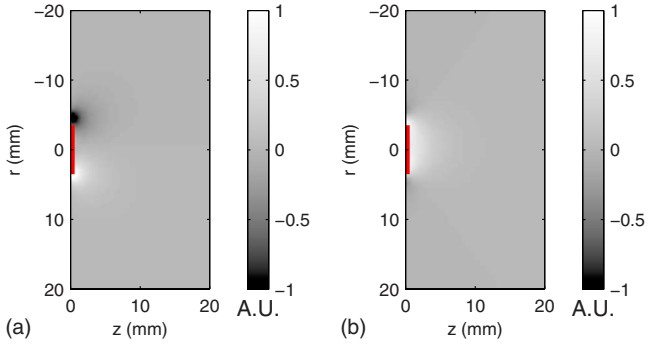


FIG. 5. (Color online) Shear force in a homogeneous medium. The source is a 9 mm vibrator centered in $O(0,0)$.

For a given source, it is then necessary to determine which part gives rise to shearing and which part gives rise to compression. In this situation, the solenoidal and irrotational parts of the source are computed.

If $\vec{W}$ is such that $\phi_f = \vec{\nabla}\cdot\vec{W}$ and $\vec{\psi}_f = -\vec{\nabla}\times\vec{W}$, then using Eq. (10), $\vec{W}$ is the solution of the Poisson equation

$$\vec{\nabla}^2\vec{W} = \vec{f}. \tag{14}$$

For a body force at the origin given by $\vec{f}(\vec{x},t) = F_0(t)\,\delta(\vec{x})\vec{e}_z$, the general solution of the equation is

$$\vec{W}(\vec{x},t) = -\frac{1}{4\pi}\int\int\int_V \frac{\vec{f}(\vec{\xi},t)dV(\xi)}{|\vec{x}-\vec{\xi}|}. \tag{15}$$

In the case of a point source, Eq. (15) becomes

$$\vec{W}(\vec{x},t) = -\frac{F_0(t)}{4\pi|\vec{x}-\vec{\xi}|}\vec{e}_z, \tag{16}$$

and in the case of a disk-shaped source (Fig. 4) the solution of Eq. (14) is

$$\vec{W}(\vec{x},t) = -\frac{1}{4\pi}\int_0^{2\pi}\int_0^R \frac{\vec{f}(\vec{\xi},t)\rho' d\rho' d\theta'}{\sqrt{\rho'^2+\rho^2-2\rho\rho'\cos(\theta-\theta')+z^2}}. \tag{17}$$

As $\vec{f}_s = -\vec{\nabla}\times(\vec{\nabla}\times\vec{W})$, in cylindrical coordinates, and taking into account the axisymmetry,

$$\vec{f}_s = \begin{cases} -\dfrac{\partial^2 W_z}{\partial r\partial z} \\ 0 \\ \dfrac{1}{r}\left(\dfrac{\partial W_z}{\partial r} + r\dfrac{\partial^2 W_z}{\partial r^2}\right). \end{cases} \tag{18}$$

Figures 5(a) and 5(b) represent the r-component and the z-component of the shear force for a 9 mm diameter source, respectively. The displacements are generated on the rims of the source.

B. Wave equations

In a viscoelastic isotropic medium,

$$\rho\frac{\partial^2\vec{u_s}}{\partial t^2} = -\bar{\mu}\vec{\nabla}\times\vec{\nabla}\times\vec{u_s} + \vec{f}_s, \tag{19}$$

with $\bar{\mu} = \mu + \eta_s(\partial/\partial t)$, where μ and η_s are the shear modulus and the shear viscosity, respectively.

If T^s_{rz} and τ^s_{rz} are defined as

$$\frac{\partial T^s_{rz}}{\partial t} = \mu\left(\frac{\partial v^s_z}{\partial r} - \frac{\partial v^s_r}{\partial z}\right) \quad \text{and} \quad \tau^s_{rz} = \eta_s\left(\frac{\partial v^s_z}{\partial r} - \frac{\partial v^s_r}{\partial z}\right), \tag{20}$$

in cylindrical coordinates, and taking into account axisymmetry,

$$-\bar{\mu}\vec{\nabla}\times(\vec{\nabla}\times\vec{u_s}) = \begin{cases} -\dfrac{\partial(T^s_{rz}+\tau^s_{rz})}{\partial z} \\ 0 \\ \dfrac{(T^s_{rz}+\tau^s_{rz})}{r} + \dfrac{\partial(T^s_{rz}+\tau^s_{rz})}{\partial r}. \end{cases} \tag{21}$$

The system to solve is then

$$\begin{aligned} \frac{\partial v^s_r}{\partial t} &= -\frac{1}{\rho}\frac{\partial(T^s_{rz}+\tau^s_{rz})}{\partial z} - \frac{1}{\rho}\frac{\partial^2 W_z}{\partial r\partial z}, \\ \frac{\partial v^s_z}{\partial t} &= \frac{1}{\rho}\left(\frac{(T^s_{rz}+\tau^s_{rz})}{r} + \frac{\partial(T^s_{rz}+\tau^s_{rz})}{\partial r}\right) + \frac{1}{\rho}\left(\frac{1}{r}\frac{\partial W_z}{\partial r} + \frac{\partial^2 W_z}{\partial r^2}\right), \\ \frac{\partial T^s_{rz}}{\partial t} &= \mu\left(\frac{\partial v^s_z}{\partial r} - \frac{\partial v^s_r}{\partial z}\right), \\ \tau^s_{rz} &= \eta_s\left(\frac{\partial v^s_z}{\partial r} - \frac{\partial v^s_r}{\partial z}\right). \end{aligned} \tag{22}$$

The same method can be applied for the compression wave.

C. Coupling term

The coupling term introduced in analytical models is composed of an irrotational part and a solenoidal part. The displacement can be written as $\vec{u} = \vec{\nabla}\phi_u + \vec{\nabla}\times\vec{\psi}_u$.

To calculate the displacement, Aki and Richards[21] computed that

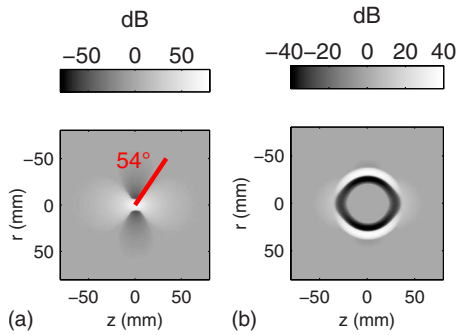


FIG. 6. (Color online) Component of the shear displacement along z computed at time $t=10$ ms (a) and $t=40$ ms (b) in a purely elastic homogeneous medium with $c_s=1$ m/s. The source is a force applied in $O(0,0)$ in the z direction.

$$\vec{\nabla}\Phi_u = -\frac{1}{4\pi\rho}\frac{3\gamma_i\gamma_j-\delta ij}{r^3}\int_0^{|r|/c_p}\tau F_0(t-\tau)d\tau + \frac{1}{4\pi\rho c_p^2}\frac{\gamma_i\gamma_j}{|r|}F_0\left(t-\frac{|r|}{c_p}\right) \quad (23)$$

and

$$\vec{\nabla}\times\vec{\Psi}_u = \frac{1}{4\pi\rho}\frac{3\gamma_i\gamma_j-\delta ij}{r^3}\int_0^{|r|/c_s}\tau F_0(t-\tau)d\tau - \frac{1}{4\pi\rho c_s^2}\frac{\gamma_i\gamma_j-\delta_{ij}}{|r|}F_0\left(t-\frac{|r|}{c_s}\right), \quad (24)$$

where $i,j=x,y,z$, $\gamma_i=\partial r/\partial i$ is the direction cosine and $F_0(t)$ is the body force applied in the z direction.

To introduce the coupling term, the two integral terms of Eqs. (23) and (24) were combined. The separation of the sources did not suppress the coupling term. Only one of the two integrals was preserved, the irrotational part or the solenoidal part. In soft media, the compression speed is a thousand times higher than the shear wave speed, which yielded $|r|/c_p \to 0$. In this case, the irrotational part of the coupling term was zero. Because of the speed of the compression wave, the part of the displacement due to this wave can be ignored. The propagation of an elastic wave in a soft medium can then be entirely described by Eq. (24). However, it should be noted that the coupling term gives rise to a displacement from time $t=0$ s although the shear wave has not crossed the medium at this time, which could appear physically surprising. The presence of displacements before the arrival of the shear wave is clearly apparent in Fig. 6. Their directivity is in agreement with the radiation pattern given by the Green's function in Fig. 1(b). This initial displacement was due to the fact that the medium had already been disturbed by the compression wave.

D. Theoretical validation

Figure 6 represents the propagation of a shear wave in a homogeneous medium where a 50 Hz apodized sine force is applied along direction z at position $r=0$ mm by a point source. The characteristics of the medium were $c_s=1$ m/s, $\eta_s=0.5$ Pa s and $\eta_p=0$ Pa s. The $(3\cos^2\theta-1)$ dependence of the displacement was clearly visible at time $t=10$ ms when the near-field term was predominant, and the $\sin^2\theta$ dependence was apparent at time $t=40$ ms when the shear wave was propagating. Comparison between the displacement computed using an analytical model and the displacement simulated using this numerical approach is shown in Fig. 7. Figure 7(a) displays the displacement as a function of depth on the axis $r=0$ mm at time $t=40$ ms and is superimposed with very good precision on the analytical expression. In Fig. 7(b), the amplitude of the displacement at location ($r=0$ mm, $z=30$ mm) is plotted as a function of time and compared with the analytical result. In this figure, the presence of a positive plateau before the arrival of the shear wave is conspicuous. The presence of this displacement before the arrival of the wave is due to the coupling term. It can be interpreted as a temporary disturbance of the medium by the compression wave which does not leave the medium in its original state.

E. Heterogeneous tissues

Figures 8 and 9 represent the propagation of a shear wave in a heterogeneous medium incorporating a spherical inclusion indicated by a circle. In Fig. 8 the shear wave speed is higher in the inclusion than in the surrounding medium ($c_s=1$ m/s in the background and $c_s=3$ m/s) in the inclusion). On the other hand, in Fig. 9, the shear wave speed is 0.5 m/s in the inclusion against 1 m/s in the background.

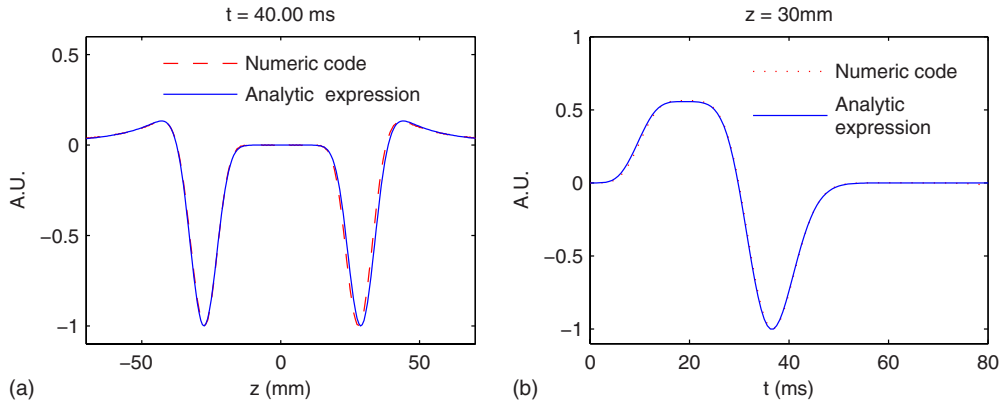


FIG. 7. (Color online) z-component of the shear displacement along z in a viscoelastic homogeneous medium with $c_s=1$ m/s, $\eta_s=0.5$ Pa s, and $\eta_p=0$ Pa s. The point source is located in $O(0,0)$ and the force is a 50 Hz sine applied in the z direction.

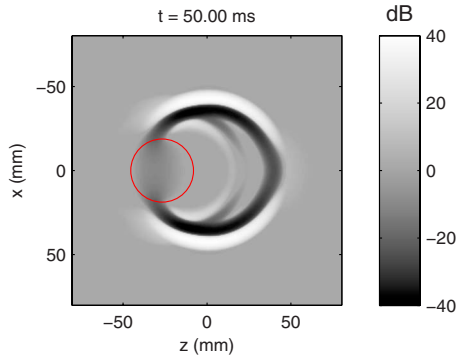


FIG. 8. (Color online) Component of the shear displacement along z computed at time $t=50$ ms in a purely elastic medium ($c_s=1$ m/s) containing a stiff inclusion ($c_s=3$ m/s) represented by a circle. The source is a force applied in $O(0,0)$ in the z direction.

In the first case, a decrease in shear wave amplitude can be seen in the inclusion. In the second case the shear wave velocity can be seen to decrease in the inclusion, and reflections of the wave can be seen at the edge of the sphere.

IV. EXPERIMENTAL RESULTS

A. Experimental setup

The measurements were performed on tissue-mimicking phantoms made of a mixture of styrene-ethylene/butylene-styrene (SEBS) copolymer and mineral oil. A 35–70 μm silica powder was used for acoustic scattering.[24] According to the Fibroscan®,[8] Young's modulus of the phantom was 4.8 kPa.

The electronic device was composed of a probe containing a low-frequency vibrator, an ultrasound transducer operating at 3.5 MHz, a dedicated electronic system, and a control unit. The sampling frequency of the signal was 50 MHz with a 14 bit resolution.[9] The technique used was transient elastography in transmit mode. In this configuration, an ultrasound transducer and a 9 mm diameter piston were aligned on each side of the medium. The piston was used to generate a 50 Hz transient vibration with a peak to peak amplitude of 2 mm and the transducer was used as an ultrasound emitter and receiver. The rf lines were acquired with a

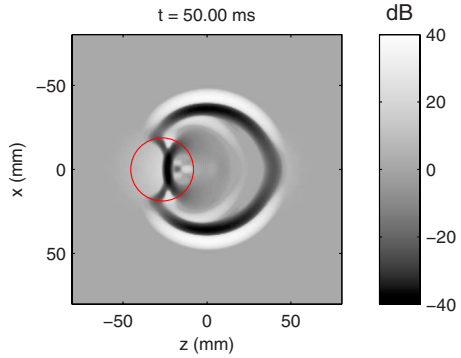


FIG. 9. (Color online) Component of the shear displacement along z computed at time $t=50$ ms in a purely elastic medium ($c_s=1$ m/s) containing a soft inclusion ($c_s=0.5$ m/s) represented by a circle. The source is a force applied in $O(0,0)$ in the z direction.

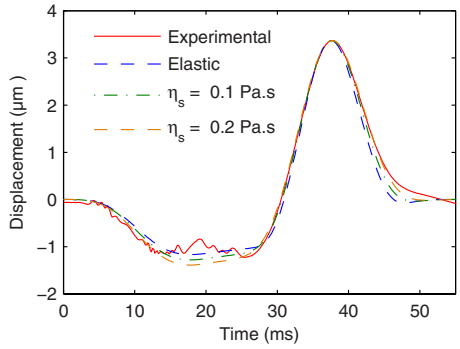


FIG. 10. (Color online) Comparison between the displacement measured in a SEBS phantom and the simulation for a 50 Hz transient excitation in a purely elastic medium, in a viscoelastic medium with $\eta_s=0.1$ Pa s, and in a viscoelastic medium with $\eta_s=0.2$ Pa s at a depth of 26 mm.

repetition frequency of 6000 Hz. The displacements of the axis of the vibrator as a function of time were known since the vibration system is servo-controlled.

B. Experimental displacement

The displacements induced in the medium were computed from the rf data using an autocorrelation method. With a 50 Hz, 2 mm peak to peak transient excitation, the maximum displacement measured in the phantom was in the range of 30–35 μm. Parameters such as the shear wave speed of the medium, the temporal evolution of the low-frequency transient excitation, and the diameter of the piston were used as inputs in the PSTD model to simulate the propagation of the shear wave in the SEBS phantom.

Figure 10 shows the displacements measured in the SEBS phantom displayed as a function of time at a depth of 26 mm. Their maximum amplitude was 3.7 μm. As predicted by the simulation, a negative plateau corresponding to the coupling term and a positive front corresponding to the shear wave can be seen. The polarity was different from that of the previous simulation because the sign of the excitation was inverted. However, some differences between the two curves can be seen and these will be discussed in Sec. V.

V. DISCUSSION AND CONCLUSION

A. Discussion

The aim of this study was to propose a model to simulate the propagation of shear waves in heterogeneous media. Two models of propagation of shear waves in viscoelastic media were compared and they showed results in good agreement with the analytical models. Such a model is a very useful tool to understand the displacements induced by an external vibrator using transient elastography. The importance of the source condition, the role of viscosity, and the influence of heterogeneities were investigated.

1. Source condition

Separation of the two parts of the source, i.e., irrotational and solenoidal, is essential for good modeling of shear wave propagation and is useful to study the displacements measured using transient elastography. Figure 6(a) shows

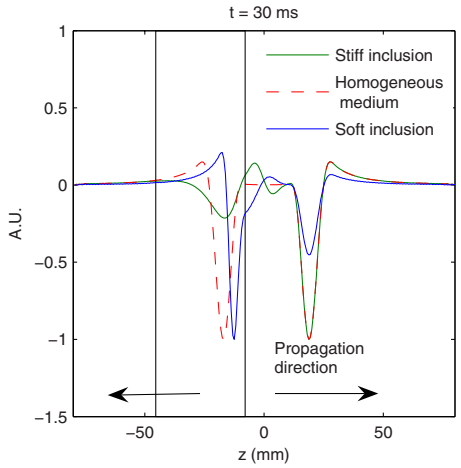


FIG. 11. (Color online) Comparison between the shear displacements in media with a stiff inclusion (c_s=3 m/s), with a soft inclusion (c_s=0.5 m/s), and in homogeneous medium (c_s=1 m/s). The boundaries of the spherical inclusion are denoted by the dashed line. The source is a 50 Hz sine force applied in $O(0,0)$ in the z direction.

that displacements are observable even at locations where the shear wave has not yet crossed the medium, and that the pattern of these displacements is in accordance with the predictions of the analytical model. Figure 1(c) shows the radiation patterns of the components of the shear displacement along z. The shear displacement is maximal at the rims of the source. However, in transient elastography the displacements are measured on the axis of the vibrator, which is not the location where the displacements are the largest in the far field. In fact, in the case of a perfect point source, no shear displacement would be measurable on the axis of the source. For instance, in Fig. 6, the maximum amplitude on the z-axis is 1.3 times higher than that on the r-axis in situation (a) (t =10 ms), when only the near-field term is apparent, and, on the other hand, once the shear wave is being propagated in situation (b) (t=40 ms), the maximum amplitude is 3.3 times higher on the r-axis than on the z-axis.

2. Influence of viscosity

The model developed also makes possible the investigation of viscosity effects on the shear wavefront. In Fig. 10, several curves are displayed representing displacements as a function of time. The first curve was computed using a purely elastic simulation, the second was obtained with a shear viscosity of 0.1 Pa s, and the third with a shear viscosity of 0.2 Pa s. The period of the shear wave seems to be higher in the experiment than in the purely elastic simulation. This shows that viscosity acts as a low-pass filter on the displacements, resulting in a widening of the wavefronts. In the case of the viscoelastic simulation, there was a better match between the periods in the experimental and simulated curves.

3. Heterogeneities

Figure 11 shows the amplitudes of displacement induced by a point source as a function of depth computed in three situations: first for a homogeneous medium with a shear wave speed of 1 m/s, then for the same medium including a spherical inclusion with a shear wave speed of 3 m/s, and finally for a medium with a soft inclusion with a shear wave speed of 0.5 m/s located at the same place as previously. As expected, the amplitude of the displacement was higher in the soft inclusion than in the stiff medium or in the background. The change in wavelength was also noticeable when the shear waves crossed the inclusion. The wavelength increased in the stiff inclusion and decreased in the soft one.

B. Conclusion

An axisymmetric model was developed to simulate the propagation of compression and shear waves in a viscoelastic medium and validated by comparison with an analytical model based on Green's functions. Two independent simulations were performed for the shear wave and for the compression wave. To achieve this separation, the source force was split into its solenoidal and irrotational parts, giving rise to a shear wave and a compression wave, respectively. This model was used to simulate the propagation of elastic waves in a heterogeneous soft tissue. The influences of viscosity and of heterogeneities on the amplitude of the shear displacement were studied, and the source condition and its influence on the displacements measured in the transient elastography configuration were investigated. The comparison between simulated displacements and experimental data showed good correspondence. Displacements measured in transient elastography before the arrival of the shear wave have been highlighted and explained by the repartition of the shear forces in the medium. Future improvements will include the development of a full three-dimensional simulation and the integration of other viscoelastic models. These simulations will also be used in transient elastography to test inverse problem algorithms for elasticity and viscosity mapping.

NOMENCLATURE

λ = Lamé's first parameter
μ = Shear modulus
η_s = Shear viscosity
η_v = Bulk viscosity
η_p = Second viscosity coefficient, $\eta_p = \eta_v - \frac{2}{3}\eta_s$
c_s = Shear wave speed, $c_s = \sqrt{\mu/\rho}$
c_p = Compression wave speed, $c_p = \sqrt{(\lambda+2\mu)/\rho}$
ρ = Density
$\vec{u}$ = Displacement vector
$\vec{v}$ = Velocity vector
$\bar{\bar{T}}$ = Stress tensor
$\bar{\bar{\tau}}$ = Viscous stress tensor
$\bar{\bar{\epsilon}}$ = Strain tensor
$\bar{\bar{1}}$ = Identity tensor
ϕ_f = Body force Helmholtz potential
$\vec{\psi}_f$ = Body force Helmholtz vector potential
ϕ_u = Displacement Helmholtz potential
$\vec{\psi}_u$ = Displacement Helmholtz vector potential
$\bar{\bar{S}}'$ = Deformation-rate tensor
$\vec{\nabla}$ = Nabla operator
tr = Trace

[1]T. A. Krouskop, D. R. Dougherty, and F. S. Vinson, "A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue," J. Rehabil. Res. Dev. **24**, 1–8 (1987).
[2]R. M. Lerner, K. J. Parker, J. Holen, R. Gramiak, and R. C. Waag, "Sono-elasticity: Medical elasticity images derived from ultrasound signals in mechanically vibrated targets," Acoust. Imaging **16**, 317327 (1987).
[3]J. Ophir, I. Cespedes, H. Ponnekanti, Y. Yazdi, and X. Li, "Elastography: A quantitative method for imaging the elasticity of biological tissues," Ultrason. Imaging **13**, 111–134 (1991).
[4]R. Muthupillai, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualization of propagating acoustic strain waves," Science **269**, 1854–1857 (1995).
[5]A. P. Sarvazyan, O. V. Rudenko, S. D. Swanson, J. B. Fowlkes, and S. Y. Emelianov, "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435 (1998).
[6]K. R. Nightingale, M. L. Palmeri, R. W. Nightingale, and G. E. Trahey, "On the feasibility of remote palpation using acoustic radiation force," J. Acoust. Soc. Am. **110**, 625–634 (2001).
[7]J. Bercoff, 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).
[8]L. Sandrin, B. Fourquet, J. M. Hasquenoph, S. Yon, C. Fournier, F. Mal, C. Christidis, M. Ziol, B. Poulet, and F. Kazemi, "Transient elastography: A new noninvasive method for assessment of hepatic fibrosis," Ultrasound Med. Biol. **29**, 1705–1713 (2003).
[9]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).
[10]K. Nightingale, M. S. Soo, R. Nightingale, and G. Trahey, "Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility," Ultrasound Med. Biol. **28**, 227–235 (2002).
[11]M. L. Palmeri, A. C. Sharma, R. R. Bouchard, R. W. Nightingale, and K. R. Nightingale, "A finite-element method model of soft tissue response to impulsive acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1699–1712 (2005).
[12]Q. H. Liu, "The pstd algorithm: A time-domain method requiring only two cells per wavelength," Microwave Opt. Technol. Lett. **15**, 158–165 (1997).
[13]H.-O. Kreiss and J. Oliger, "Comparison of accurate methods for the integration of hyperbolic equations," Tellus **24**, 199–215 (1972).
[14]Q. H. Liu and G. Zhao, "Review of pstd methods for transient electromagnetics," Int. J. Numer. Model. **17**, 299–323 (2004).
[15]O. Bou Matar, V. Preobrazhensky, and P. Pernod, "Two-dimensional axisymmetric numerical simulation of supercritical phase conjugation of ultrasound in active solid media," J. Acoust. Soc. Am. **118**, 2880–2890 (2005).
[16]E. Filoux, S. Calle, D. Certon, M. Lethiecq, and F. Levassort, "Modeling of piezoelectric transducers with combined pseudospectral and finite-difference methods," J. Acoust. Soc. Am. **123**, 4165–4173 (2008).
[17]J. Bercoff, M. Tanter, M. Muller, and M. Fink, "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).
[18]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).
[19]J. P. Berenger, "Three-dimensional perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **127**, 363–379 (1996).
[20]M. Ghrist, B. Fornberg, and T. A. Driscoll, "Staggered time integrators for wave equations," SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **38**, 718–741 (2000).
[21]K. Aki and P. G. Richards, *Quantitative Seismology*, 2nd ed. (University Science Books, Sausalito, 2002).
[22]L. Sandrin, D. Cassereau, and M. Fink, "The role of the coupling term in transient elastography," J. Acoust. Soc. Am. **115**, 73–83 (2004).
[23]S. Calle, J. P. Remenieras, O. Bou Matar, M. Elkateb, and F. Patat, "Temporal analysis of tissue displacement induced by a transient ultrasound radiation force," J. Acoust. Soc. Am. **118**, 2829–2840 (2005).
[24]J. Oudry, C. Bastard, V. Miette, R. Willinger, and L. Sandrin, "Copolymer-in-oil phantom materials for elastography," Ultrasound Med. Biol. **35**, 1185–1197 (2009).

Erratum: A new equation for the accurate calculation of sound speed in all oceans [J. Acoust. Soc. Am. 124, 2774–2783 (2008)]

Claude C. Leroy
IMM Vivaldi, 30 Chemin de la Baou, 83110 Sanary Sur Mer, France

Stephen P. Robinson and Mike J. Goldsmith
National Physical Laboratory, Teddington, Middlesex TW11 0LW, United Kingdom

(Received 16 April 2009; accepted 13 July 2009)
[DOI: 10.1121/1.3203669]

PACS number(s): 43.30.Es, 43.30.Pc, 43.10.Vx

A number of errors have come to light in the above paper with regard to units of salinity. In each example of the error, the unit for salinity has been stated as "‰" instead of "‰"; this erroneously represents the values as being in percent rather than in parts per thousand. The reader should note that the units for salinity throughout the paper are intended as parts per thousand. On p. 2776, line 15 of Sec. IIB, it should read "parts per thousand" rather than "percent."

The postal code for the National Physical Laboratory should read TW11 0LW.

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.*

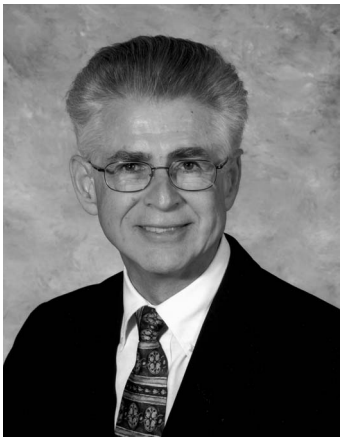
New Fellows of the Acoustical Society of America

Diane Dalecki—For contributions to the bioeffects of sound and ultrasound

Robert G. Leisure—For contributions to the acoustic characterization of condensed matter

David K. Mellinger—For contributions to the analysis of marine mammal calls

Susan Nittrouer—For contributions to developmental speech perception and production



Joe Posey—For leadership in aircraft noise control

Andrew J. Szeri—For contributions to the nonlinear dynamics of acoustic cavitation

President's report on the 157th meeting of the Acoustical Society of America held in Portland, Oregon

The 157th meeting of the Acoustical Society of America (ASA) was held 18–22 May 2009 at the Hilton Portland and Executive Tower in Portland, Oregon. This is the first time that the Society has met in this city and was the first "green meeting" held by ASA.

The meeting drew a total of 1470 registrants, including 216 nonmembers, 423 students and 155 registrants from outside North America. There were 28 from the United Kingdom, 23 from Japan, 20 from Germany, 17 from France, 14 from Korea, 11 from the Netherlands, 7 from Australia, 4 each from Russia, Spain, Sweden, and Taiwan, 3 each from Brazil and Israel, 2 each from Belgium and Norway, and 1 each from Argentina, Czech Republic, Hong Kong, Iceland, India, Italy, New Zealand, Portugal, and the Slovak Republic. North American countries, Canada, Mexico, and the United States, accounted for 53, 3, and 1259, respectively.

A total of 1091 papers, organized into 116 sessions, covered the areas of interest of all 13 Technical Committees and the Committee on Education in Acoustics. The Society is grateful to the individuals who organized the 75 special sessions that were part of the technical program. The meeting also included 19 meetings dealing with standards. The evening tutorial lecture series was continued by Ela and Leah Lamblin and Uwe Hansen with the

 0001-4966/2009/126(4)/2118/11/$25.00

tutorial titled "The Art and Science of Unique Musical Instruments." This tutorial was presented at the Newmark Theater and was open to the public. Over 400 people attended.

The Society's thirteen Technical Committees held open meetings during the Portland meeting where they made plans for special sessions at upcoming ASA meetings, discussed topics of interest to the attendees and held informal socials after the end of the official business. These are working, collegial meetings and all people attending Society meetings are encouraged to attend and to participate in the discussions. More information about Technical Committees, including minutes of meetings, can be found on the ASA Website ⟨http://asa.aip.org/committees.html⟩ and in the Acoustical News section of JASA in the September and November issues.

A short course titled "Outdoor Noise Estimation and Mapping" was given to a group of 20 students. The instructors were Ken Kaliski, Resource Systems Group, Robert Putnam, Siemens Energy Systems, Brigitte Schulte-Fortkamp, Technical University of Berlin, and Klaus Genuit, HEAD Acoustics.

Thomas L. Geers of the University of Chicago presented a Distinguished Lecture titled "A Residual-Potential Boundary for Time-Domain Problems in Computational Acoustics." A Student Design Competition, sponsored by the Technical Committee on Architectural Acoustics and the National Council of Acoustical Consultants was judged at the meeting by a panel of professional architects and acoustical consultants. The winners received awards of $1250 for the project judged "first honors," with $700 awarded to designs judged "commendation." Other events included a presentation on the new National Institutes of Health peer review procedures and a workshop on preparing JASA and JASA Express Letters articles.

The ASA Technical Committee on Noise in collaboration with the City of Portland's Noise Control Office presented a day-long symposium entitled "Urban Design with Soundscape in Mind: A Symposium on Urban Planning with the Consideration of Noise Impacts and the People Concerned" which was organized by Kerrie Standlee of Daly-Standlee & Associates, Inc. It brought together city and private planners, acousticians, architects, engineering, government officials, and the public to discuss urban planning and development in relation to the health and livability issues of noise pollution.

The Second ASA Special Workshop on Speech: Cross-Language Speech Perception and Variations in Linguistic Experience, was held 21 to 23 May, at the World Trade Center in Portland. The workshop, intended for basic and applied researchers and graduate students in psycholinguistics, linguistics, phonetics, speech sciences, languages and second language learning, drew a large attendance of about 200 people. The symposium organizers were Catherine Best, University of Western Sydney, Ann Bradlow, Northwestern University, Susan Guion, University of Oregon, and Linda Polka, McGill University.

The ASA Student Council sponsored a workshop for young investigators which focused on the mechanics of writing grant applications. A Student Reception was held with over 300 people in attendance. This reception, which is sponsored by the ASA and supported by the National Council of Acoustical Consultants, enabled students to meet with established members of the Acoustical Society of America. Several of the Technical Committees awarded Best Student Paper Awards or Young Presenter Awards to students and young professionals who presented papers at the meeting. The list of award recipients, as well as other information for students, can be found online at the ASA Student Zone website http://www.acosoc.org/student/.

Social events included the two social hours held on Tuesday and Thursday, an "icebreaker" and a reception for students, the Fellows Luncheon and the morning coffee breaks. A special program for students to meet one-on-one with members of the ASA over lunch, which is held at each meeting, was organized by the Committee on Education in Acoustics. These social events provided the settings for participants to meet in relaxed settings to encourage social exchange and informal discussions. The Women in Acoustics Luncheon was held on Wednesday afternoon with attendance of 100.

Members of the Society organized a "jam session" which was held in the hotel lounge on Tuesday evening from 9:00 p.m. to 12:00 p.m. A live band, which included members of the ASA and their friends and family, played for an enthusiastic audience of meeting attendees.

The plenary session included a business meeting of the Society, announcements, acknowledgment of the members and other volunteers who organized the meeting and the presentation of awards and certificates to newly elected Fellows.

FIG. 1. (Color online) ASA President Mark F. Hamilton (r) congratulates Martin Siderius, recipient of the 2009 Medwin Prize in Acoustical Oceanography.

ASA President Mark Hamilton presided over the Plenary Session and Awards Ceremony. Clark Penrod, Chair of the Fall 2009 meeting, addressed the audience and invited and encouraged them to attend the meeting to be held in San Antonio in October.

The 2009 Medwin Prize in Acoustical Oceanography was presented to Martin Siderius of Portland State University. Dr. Siderius presented the Acoustical Oceanography Prize Lecture titled "Seabed characterization and model based processing: Past, present, and future" earlier in the meeting (see Fig. 1).

The R. Bruce Lindsay Award was presented to Kelly J. Benoit-Bird, Oregon State University, "for contributions in marine ecological acoustics" (see Fig. 2). The Silver Medal in Speech Communication was presented to Winifred Strange of the Graduate School, City University of New York—Graduate School and University Center, "for contributions to understanding speech perception" (see Fig. 3). The Gold Medal was presented Thomas D. Rossing, Stanford University, "for contributions to musical acoustics, leadership in science education, and service to the Society" (see Fig. 4).

Election of fourteen members to Fellow grade was announced and fellowship certificates were presented. New fellows are: Russell E. Berger, II, Suzanne E. Boyce, Ann R. Bradlow, Douglas S. Brungart, Richard S. McGowan, Luc Mongeau, Patrick W. Moore, Trevor R. T. Nightingale, D. Lloyd Rice, Christine H. Shadle, Martin Siderius, Aaron M. Thode, Doug H. Whalen, Lisa M. Zurk (see Fig. 5).

FIG. 2. (Color online) ASA President Mark Hamilton presents the 2009 R. Bruce Lindsay Award to Kelly Benoit-Bird.

FIG. 3. (Color online) ASA President Mark Hamilton presents the Silver Medal in Speech Communication to Winifred Strange.

FIG. 4. (Color online) ASA President Mark Hamilton (r) presents the 2009 Gold Medal to Thomas D. Rossing (l).

ASA President Mark Hamilton introduced Lisa Zurk, Chair of the Portland meeting, (see Fig. 6) who expressed the Society's thanks to members of the Local Committee for the excellent execution of the meeting, which clearly evidenced meticulous planning, including Marjorie R. Leek, Technical Program Chair; James McNames and Brent Casaday, Audio-Visual; Cheryl Siderius, Food and Beverage/Social Events/Meeting Administrator; Jorge Quijano, Signs and Publicity; Michel Pinton, Accompanying Persons Program; Dawn Konrad-Martin, Local Outreach; and Barry Ma, Green Meeting Coordinator. She also expressed thanks to the members of the Technical Program Organizing Committee: Marjorie R. Leek, Technical Program Chair; Jeffrey A. Nystuen, Acoustical Oceanography; Holger Klinck, Animal Bioacoustics; David T. Bradley, Architectural Acoustics; Peter J. Kaczkowski, Azzdine Ammi, Biomedical Ultrasound/Bioresponse to Vibration; Thomas J. Matula, Education in Acoustics, Physical Acoustics,

FIG. 5. (Color online) New Fellows of the Acoustical Society of America with ASA President and Vice President (l to r): Mark Hamilton, Lisa Zurk, Doug Whalen, Martin Siderius, Christine Shadle, Lloyd Rice, Patrick Moore, Richard McGowan, Aaron Thode, Douglas Brungart, Ann Bradlow, Suzanne Boyce, Russell Berger, Victor Sparrow.

FIG. 6. (Color online) Lisa Zurk, Chair of the Portland meeting.

FIG. 7. (Color online) Diane Kewley-Port, ASA Vice President-Elect, presents the Vice President's gavel to Victor Sparrow, ASA Vice President.

FIG. 8. (Color online) Whitlow Au, ASA President-Elect (l), presents the ASA President's tuning fork to Mark Hamilton, ASA President (r).

and Structural Acoustics and Vibration; Thomas R. Howarth, Engineering Acoustics; Kerrie G. Standlee, Noise; Frederick Gallun, Psychological and Physiological Acoustics; Paul J. Hursky, Signal Processing in Acoustics; Susan G. Guion, Melissa R. Redford, Speech Communication; Thomas J. Matula and Jon La Follett, Structural Acoustics and Vibration; Dajun Tang, Underwater Acoustics.

The Plenary Session concluded with the presentation of the Vice President's gavel to Victor Sparrow and the President's Tuning Fork to Mark Hamilton, in recognition of their service to the Society during the past year (see Figs. 7 and 8).

The full technical program and award encomiums can be found in the printed meeting program or online for readers who wish to obtain further information about the Portland meeting (visit scitation.aip.org/jasa and select Volume 125, Issue 4, from the list of available volumes).

We hope that you will consider attending a future meeting of the Society to participate in the many interesting technical events and to meet with colleagues in both technical and social settings. Information about future meetings can be found in the *Journal* and on the ASA Home Page at http://asa.aip.org.

MARK F. HAMILTON

President 2008–2009

Calendar of Meetings and Congresses

	2009
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. 30th 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 SignalProcessing. Web: http://icassp2010.org
19–23 April	Baltimore, MD, USA. Joint meeting: 159th Meeting of the Acoustical Society ofAmerica and Noise Con 2010. Web: http://asa.aip.org/meetings.html
27–30 April	Ghent, Belgium. Institute of Acousts/Belgian Acoustical Association JointMeeting. 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 (Jointmeeting 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

Revision List

New Associates

Allen, Joyce, 1002 Valley Rd., Colorado Springs, CO 80904
Ammi, Azzdine, Oregon Health & Science Univ., Cardiovascular Medicine, 3181 SW Sam Jackson Park Rd., UHN-62, Portland, OR 97239
Anraku, Kazuhiko, Faculty of Fisheries, Kagoshima Univ., 4-50-20 Shimoarata Kagoshima, Japan
Assous, Said, British Geological Survey, Ultrasound Research Lab., Rm. K107, Keyworth Nottingham NG72 5AQ UK
Banning, Jeremy, 6304 Rancho Mission Rd. #136, San Diego, CA 92108
Barbar, Steve, LARES Associates/E-coustics, 30 Dunbarton Rd., Belmont, MA 02478
Bazin, Simon, Sealynx Automotive, Transieres, 27380 Charleval, France
Beache, Tristan, Proletariat Productions, LLC, P.O. Box 12968, Chicago, IL 60612
Berger, Thomas, 36382 Bird Song Ct., Winchester, CA 92596
Bigler, Kurt, 2953 Hillegass Ave., Berkeley, CA 94705
Blesser, Barry, Blesser Assoc., P.O. Box 155, Belmont, MA 02478
Boedts, Michael, ENT, Putkapelstraat 103A, 9051 Gent, Belgium
Boeker, Eric, USDOT Volpe NTSC, RVT-41, 55 Broadway, Cambridge, MA 02142
Bonamini, Gabriele, Studio Legno—Wood Consulting, Via Pietro Toselli, 164, I 56034 Firenze, Italy
Bonsi, Davide, Scuola di San Giorgio, Acoustics Lab., Isola di San Giorgio Maggiore, Venezia, Italy
Brooks, Laura, Victoria Univ. of Wellington, SGEES, P.O. Box 600, Wellington 6015, New Zealand
Brown, Lance, Glacier Bay Inc, Business Development, 2930 Faber St., Union City, CA 94587
Burton, Thomas, P.O. Box 232, Hakalau, HI 96710
Calantoni, Joseph, Naval Research Lab., Bldg. 1005, Rm. D-17D, Stennis Space Center, MS 39529
Caley, Michael, Savery & Associates Pty Ltd, P.O. Box 265, The Gap, 4061, Australia
Camp, Marc-Antoine, Dufourstrasse 42, CH-8702 Zollikon, Swizterland
Cartmill, John, 12200 Woodvale Ct, Herndon, VA 20170
Cazzanti, Luca, Univ. of Washington, Applied Physics Lab., BOX 355640, Seattle, WA 98105
Cheng, Rui, 1330 Charleston Rd., Mountain View, CA 94043
Choi, Jung-Woo, Samsung AIT, Multimedia Lab., Giheung-gu, Nongseo dong, Yongin-Si, Gyunggi, Korea
Cole, Brian, Gentex Corporation, Engineering, 2456 Brown Ave, Manchester, NH 03103
Coleman, Ronald, BBN Technologies, Sensor Systems, 10 Moulton St., Cambridge, MA 02138
Corbett, Scott, 11720 SW 28th Pl., Portland, OR 97219
Costello, Amy, Armstrong World Industries, Building Products Div., 2500 Columbia Ave., Lancaster, PA 17602
Cranford, Ted, 2674 Russmar Dr., 2674 Russmar Dr., San Diego, CA 92123-3422
Daddario, Nicholas, 866 Dukes Dr, Yardley, PA 19067
Darcy, Isabelle, Indiana Univ., Second Language Studies, 1021 E. Third St., Bloomington, IN 47405
Dauman, Rene, 47 Ave. Victor Hugo, Villas de Bouscat, Villa u 23, Le Bouscat 33110, France
Davies, Max, 18 Highfield Crescent, Southampton, SO17 1SF, UK
Davis, Alison, 43 King St, Apt. 1, Falmouth, MA 02540
Degertekin, Levent, Georgia Inst. of Technology, School of Mechanical Engineering, 801 Ferst Dr., NW, Atlanta, GA 30332
Delgado, Rafael, Intelligent Hearing Systems, 6860 SW 81st St., Miami, FL 33143
Despaux, Gilles, Univ.Montpellier2, IES CC082, Place E Bataillon, 34095 Montpellier, France
Dib, Tarek, Illingworth & Rodkin, Inc, 505 Petaluma Blvd. South, Petaluma, CA 94952
Dicecco, John, 95 Kings Ridge Rd., Wakefield, RI 02879
Dick, Elliott, 14217 Arbre Ln. North, Hugo, MN 55038
Digerness, Joseph, Arup, Acoustics, 560 Mission St., Ste. 700, San Francisco, CA 94105
Dries, John, Dries Engineering, 1509 Edgwood Place, Louisville, KY 40205
Duble, Albert, 16905 NE Kings Grade, Newberg, OR 97132
Dugan, Joshua, Acoustics and Light, 3846 Calle Cita, Santa Barbara, CA 93110
Dunlap, Kristen, P.O. Box 20872, Juneau, AK 99802
Duraiswami, Ramani, Univ. of Maryland, College Park, MD 20742
Ebenezer, D, NPOL, Transducer Group, Thrikkakara, Kochi, Kerala 682021, India
Edson, Patrick, 99 Perimeter Road, Nashua, NH 03063
Eliasson, Veronica, Caltech, GALCIT, 1200 E. California Blvd, Pasadena, CA 91125
Fadum, Ole, Scanmar AS, R&D, Asgaardstrandveien 359, 3179, Norway
Farion, Rob, 3749 43A Ave., Ste. 1NO1, Red Deer Alberta T4N 3G2 Canada
Feinstein, Steven, MIT Lincoln Lab., Group 62, 244 Wood St., Lexington, MA 02420
Fischer, Wayne, 943 Clithero Dr., Boise, ID 83703
Foreman, Terry, 6804 Deer Run Dr., Alexandria, VA 22306-1103
Fornshell, John, 6911 Quander Rd., Alexandria, VA 22307
Fox, Peter, Kongsberg Nesotech Ltd., Engineering, 1598 Kebet Way, Port Coquitlam, BC V3C 5M5, Canada
Frangetto, Polyana, Kirkegaard Associates, Rm. Acoustics, 801 W. Adams St., 8th Fl., Chicago, IL 60607
Garai, Massimo, Univ. of Bologna, Dienca, Viale Risorgimento, 2, Bologna 4136, Italy
Garinis, Angie, 3039 South Court St., Seattle, WA 98144
Girvin, Raquel, FAA, Office of Environment and Energy, 800 Independence Ave. SW, Washington, DC 20591
Gladden, Joseph, Univ. of Mississippi, Physics/NCPA, 1 Coliseum Dr., University, MS 38677
Gosselin, Francois, Decibel Consultants, 265 Hymus, Bureau 2500, Pointe-Claire, QC H9R 1G6, Canada
Greenwood, Margaret, Pacific Northwest National Lab., MS K5-26, P.O. Box 999, Richland, WA 99352
Groenaas, Halvor, WesternGeco/Schlumberger, Marine, Solbraaveien 23, 1166 Asker, Norway
Guarro, Joan, Major 52, 25007 Lleida, Spain
Guralnik, Zachary, 176 N. Columbus St, Arlington, VA 22203
Guthrie, Priscilla, 123 Gresham Pl., Falls Church, VA 22046
Hammon, William, BAI, LLC, 7527 Jason St., Houston, TX 77074
Handa, Rajash, Anatomy and Cell Biology, Indiana Univ. School of Medicine, 635 Barnhill Dr, MS510, Indianapolis, IN 46202
Hanser, Sean, Evolution & Ecology, Univ. of California, 1 Shields Ave., Davis, CA 95616
Harisadhan, Patra, Boys Town National Research Hospital, Research, 555 N. 30th St., Omaha, NE 68131
Harrison, Seth, KJWW Engineering Consultants, 623 26th Ave., Rock Island, IL 61201
Hart, George, 16158 NW Church Rd., Seabeck, WA 98380
Haynes, John, American Inst. of Physics, 2 Huntington Quadrangle, Ste. 1NO1, Melville, NY 11747
Hermes, Michael, Harley-Davidson Motor Company, 11800 W. Capitol Dr., Wauwatosa, WI 53222
Hirata, Yoshihiro, Hokkai-Gakuen Univ., Dept. of Engineering, 1-1, Nishi 11, Minami 26, Chuou-ku, Sapporo, 0640926, Japan
Hoffman, Mark, PS&S, Environmental, 67B Mountain Boulevard Extn, Warren, NJ 07059
Holland, Julian, 3C All Saints Road, Clifton, Bristol, BS8 2JG, UK
Homma, Kenji, 225 Mountain Breeze Dr., Blacksburg, VA 24060

Hopkins, Lawrence, Huntair Inc., R&D, 11555 SW Myslony St., Tualatin, OR 97062
Houston, Derek, Indiana Univ School of Med, Otolaryngology, 699 West Dr., RR044, Indianapolis, IN 46202
Ikeda, Teiichiro, Hitachi Ltd., Central Research Lab., 1-280 Higashi-Koigakubo, Kokubunji-shi, Tokyo, 185-8601, Japan
In-Kyung, Song, ARUP, Acoustics, 12777 West Jefferson Blvd., Ste. 200, Los Angeles, CA 90066
Isono, Sam, 4968 Yonge St. unit 3706, Toronto, ON M2N 7G9, Canada
Janata, Petr, Univ. of California, Davis, Center for Mind and Brain, 267 Costeau Pl., Davis, CA 95618
Javitch, Ronald, P.O. Box 67-Station H, Montreal, QC H3G 2K5, Canada
Jedrzejczak, Wiktor, Institute of Physiology and Pa, ul. Zgrupowania AK Kampinos 1, 01-943, Warsaw, Poland
Jeffries, Aileen, Energy Resources, P.O. Box 601, Winthrop, WA 98862
Jones, Theodore, Code 6795, U.S. Naval Research Lab., 4555 Overlook Ave., SW, Washington, DC 20375
Jones, Jeffery, Wilfrid Laurier Univ., Psychology, 75 Univ. Ave. W., Waterloo, ON N2L 3C5, Canada
Kastelein, Ronald, SEAMARCO, Julianalaan 46, 3843 CC Harderwijk, The Netherlands
Kelly, Kane, PBK Architects, Technology, 4141 Rosemeade Pkwy., Apt. 3304, Dallas, TX 75287
Kirby, Edmund, Geocomp Corporation, 118A Fulton St., Ste. 316, New York, NY 10038
Kirn, Larry, 2106 Kenwood, Austin, TX 78704
Klinck, Holger, Oregon State Univ., CIMRS, 2030 SE Marine Science Dr., Newport, OR 97365
Koenig, Axel, 4345 NE Failing St., Portland, OR 97213
Lau, Siu-Kit, Charles W. Durham School of Architecture, Engineering and Construction, 203C Peter Kiewit Institute, 1110 S. 67th St.-C, Omaha, NE 68182-0681
Laux, Peter, 6654 scenic shores Dr., Holland, MI 49423
Lavely, Eugene, BAE Systems AIT, Signal Understanding & Network, 6 New England Executive Park, Burlington, MA 01803
Lawson, Aaron, 11729 Route 365, Remsen, NY 13438
Lawson, Gareth, Woods Hole Oceanographic Inst, 266 Woods Hole Rd., Woods Hole, MA 02543
Lee, Won Sang, Korea Polar Research Institute, Polar Earth-System Sciences, Songdo TechnoPark 7-50, Songdo-dong, Yeonsu-gu, Incheon, 406-840, Korea
Lehman, Michael, American Micro Industries Inc., Acoustic Div., 440 Ramsey Ave., Chambersburg, PA 17201
Lepper, Paul, 215 Ratcliffe Rd., Sileby, Leicestershire LE12 7PY, UK
Liu, Xiazhou, Institute of Acoustics, No. 22, Hankou Rd., Nanjing Tiangsu 210093, China
Loftis, Charles, 1003 Blossom Ln., Redondo Beach, CA 90278
Lopez-Soto, Teresa, Univ. of Seville, Dpto. Lengua Inglesa, Palos de la Frontera, s-n, 41004 Sevilla, Spain
Martinelli, Sheri, 55 Hunter Ave., Newport, RI 02840
Mceth, Michael, 204 Indian Springs Rd., Williamsburg, VA 23185
Mckee, Andrew, 53 W.111th St., Apt. 2E, New York, NY 10026
Mckenzie, David, P.O. Box 133, Callaghan, NSW 2308, Australia
Melamed, Sarah, Portland VAMC, National Ctr. for Rehabilitation Auditory Research, 3710 SW US Veterans Hospital Rd., Portland, OR 97239
Merimaa, Juha, 945 Crane St., Apt. F, Menlo Park, CA 94025
Misol, Malte, German Aerospace Center, Inst. of composite structures, Lilienthalplatz 7, 38108 Braunschweig, Germany
Moore, Philip, 2857 Altura Ave., La Crescenta, CA 91214
Morrill, Mark, 4 Russell St., Apt. 3L, Brooklyn, NY 11222
Munger, Lisa, Univ. of Calif San Diego, Scripps Inst of Oceanography, 9500 Gilman Dr., Mailcode 0205, La Jolla, CA 92093-0205
Murphy, David, 593 Vauxhall St. Ext., Waterford, CT 06385
Murphy, Chad, Amirix Systems Inc, 211 Horseshoe Lake Dr., Halifax, NS B3S0B9, Canada
Musso-Escude, Laura, Wash. State Dept. of Trans., Acoustic, Air Quality, Energy, 15700 Dayton Ave. N., MS138, P.O. Box 330310, Shoreline, WA 98133-9710
Nam, Hosung, Haskins Laboratories, 300 George St., Ste. 900, New Haven, CT 06511
Nilsson, Andreas, WS Atkins andPartners Overseas, Acoustics, WS Atkins and Partners Overseas, P.O. Box 5620, Dubai, United Arab Emirates
Noble, Erik, 13199 E. Coyote Well Dr., Vail, AZ 85641
O'Regan, Stephen, Ships Signatures, NSWC Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20874
Ogden, Jason, Parsons, PTG, 100 W. Walnut St., Pasadena, CA 91124
Oglesbee, Eric, Bethel College, Modern Language Studies, 1001 Bethel Cir., Mishawaka, IN 46545
Okada, Kazuhide, Motor Corporation, Vehicle Control System Dept., 1200, Mishuku, Sussono, Shizuoka, 410-1193, Japan
Orme, Richard, 8 Portwey Close, Weymouth, DT4 8RF, UK
Papenberg, Cord, IFM-GEOMAR, Dynamics of the Ocean Floor, Wischhofstrasse 1-3, 24148 Kiel, Germany
Parisuttikhan, Thepparit, 1,33Moo4, Poochaosamingprai Rd., Bang Hua Sua, Prapradang, Samutprakarn 10130, Thailand
Pees, Edward, 50 Mill Pond Rd., Exeter, RI 02822
Peot, Mark, 800 Highgrove Dr., Chapel Hill, NC 27516
Petculescu, Petre, Ovidius Univ., Physics, 124 Mamaia Blvd, 900527 Constanta, Romania
Pettersen, Inge, Verven 18B—L0508, 4014 Stavanger, 4014 Stavanger, Norway
Purkait, Bobby, Nivasonix, 6420 Via Real, Ste.1, Carpinteria, CA 93013
Raghavan, Raghu, 4203 Somerset Place, Baltimore, MD 21210
Ramani, Deepak, Northrop Grumman Corporation, Systems Engineering, P.O. Box 1488, MS 9910, Annapolis, MD 21401
Rao, Mohan, Michigan Technological Univers, ME-EM, 1400 Townsend Dr., Houghton, MI 49931
Ratnam, Rama, Univ of Texas at San Antonio, Biology, One UTSA Cir., San Antonio, TX 78249
Rauchholz, Daniel, UGN, Inc., 38555 Hills Tech Dr., Farmington Hills, MI 48331
Rawlings, Samantha, Veneklasen Associates, Acoustics, 1711 16th St., Santa Monica, CA 90404
Regmi, Sundar, Olympus NDT, 48 Woerd Ave., Waltham, MA 02453
Reichmuth, Colleen, UC Santa Cruz, Institute of Marine Sciences, Long Marine Lab., 100 Shaffer Road, Santa Cruz, CA 95060
Rick Boyce, Rick, Memtech Inc., Acoustical, 9033 General Dr., Plymouth, MI 48170
Ritcey, James, Univ. of Washington, Electrical Engineering, Box 352500, Seattle, WA 98195
Rodriguez, Ivan, 201 E. South Temple #223, Salt Lake City, UT 84111
Rogers, Daniel, 3M, Infection Prevention Div., 3M Center, Bldg. 0270-02-N-03, St. Paul, MN 55144-1000
Romblom, David, Sennheiser Research, 3239 El Camino Real, 3rd Fl., Palo Alto, CA 94306
Rosenblum, Steven, Corning Inc., Strategic Growth, SP-FR-02-12, Corning, NY 14850
Rosenhouse, Judith, 89 Hagalil St., Haifa 32684 Israel
Rosenhouse, Giora, 89 Hagalil Str., Haifa 32684 Israel
Rountree, Rodney, Marine Ecology and Tech. Appl, 178 Teaticket Hwy, Ste. 101, East Falmouth, MA 02536
Runzer, Tim, 625, East 9th St., North Vancouver, BC V7L2B8, Canada
Russell, Daniel, Kettering Univ., Physics Dept., 1700 W. Univ. Ave., Flint, MI 48504
Rutter, Ben, Univ. of Oklahoma Health Sciences Ctr., Communication Sciences & Disorders, 1200 N. Stonewall Ave., Oklahoma City, OK 73126-0901
Ryan, Jim, Target Finders, Unit6B Grays Farm Road, Orpington, BR5 3BD, UK
Saviotti, Massimo, Noisetronic Sas, Via Ballanti Graziani, 20, 48018 Faenza, Italy
Schlaefer, David, 3M, EAR, 3M Center, Building 235-2E-57, Saint Paul, MN 55144
Schmith, Jean, Gomes Freire de Andrade, 299, Canoas, Brazil
Scholer, Dave, MED-EL, Clinical Research Dept., Fuerstenweg 77a, 6020 Innsbruck, Austria
Seiner, John, National Center for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677
Serry, Mohamed, Kamigyo-ku Nishimaruta-cho 173, Melodyheim-Nijojo #303, Kyoto-Fu, Kyoto-Shi, 602-8148, Japan

Sevilla, Andrew Michael, Kinetic Audio Technologies, 22 Tacloban St, ALab.ang Hills, Village, Muntinlupa City Mm, Philippines
Singleton, Sabrina, 7454 Blue Bird Dr., Marion, MS 39342
Smaragdis, Paris, 101 Monmouth St., Apt. #307, Brookline, MA 02446
Smirnov, Andrey, Acoustic Group Ukraine, k.14, 8-9, Str. Gaitsana, 01010 Kiev, Ukraine
Smith, Greg, IAC, Engineering, IAC House, Moorside Road, Winchester, SO23 7US, UK
Smith, Sonya, Howard Univ., Mechanical Engineering, 2300 6th St. NW, Rm. 2032, Washington, DC 20059
Solbeck, Jason, Sound Innovations Inc., 55 Railroad Row, Ste. 307, White River Junction, VT 05001
Stewart, Catherine, USACHPPM, Operational Noise Program, 5158 Blackhawk Rd, Apg, MD 21010
Sun, Guiqing, Insitute of Acoustics, CAS, Beisihuan west road 21, 100190 Beijing, P.R. China
Swift, Reni, Sea Mammal Research Unit, Scottish Oceans Institute, Univ. of St. Andrews, St. Andrews, Fife, KY16 8LB, UK
Takeuchi, Yasuhito, Kagoshima Univ., Information and Computer Science, 1-21-40 Korimoto, Kagohima, 8900065, Japan
Takeyasu, Hajime, Japan, 4-6-5-308 Yahata-cho, Nada-ku, Kobe-Shi, Hyogo, 657-0051, Japan
Tamayo, Thomas, 84 Whitewood Dr., Wakefield, RI 02879
Tevlin, Maksim, Nachal Dolev 29 A, flat 7, 99630 Beyt Shemesh, Israel
Thompson, Christian, 345 86th St., Apt. 120, Brooklyn, NY 11209
Thompson, Paul, Univ. of Aberdeen, Lighthouse Field Station, George St., Cromarty IV11 8YJ, UK
Thomson, Ron, Brock Univ., Dept of Applied Linguistics, 500 Glenridge Ave., St. Catharines, ON L2S 3A1, Canada
Townsend, John, 1430 Capri Dr., Campbell, CA 95008
Tyson, Terence, Kirkegaard Associates, 801 W. Adams, Chicago, IL 60607
Vasconcelos, Ivan, c. Mir 5, 1o 2a, 08172 St Cugat Del Valles, Spain
Veau, Nicolas, Neurelec, Research Development, 2720 Chemin Saint Bernard, 06224 Vallauris, France
Versluis, Michel, Univ. of Twente, Physics of Fluids Group, P.O. Box 217, 7500 AE Enschede, The Netherlands
Vestergaard, Martin, Univ. of Cambridge, Physiology Dept., Downing St., Cambridge CB2 3EG, UK
Voncannon, David, D & E Associates, 4054 Tilden Ave., Culver City, CA 90232
Vonohlsen, Jon, Quest Product Development, 4900 Iris St., Wheat Ridge, CO 80033
Walker, Steven, Harbor Technology & Acoustics, 15191 Willowwood Cir., Grand Haven, MI 49417
Walker, Shane, Scripps Inst. of Oceanography, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238
Walter, Mary Ann, 2410 Happy Hollow Rd., E4, West Lafayette, IN 47906
Wang, John, Volvo Construction Equipment, 312 Volvo Way, Shippensburg, PA 17257
Warner, Brian, Vibra-Tech Engineers, Structural Dynamics/Engineer, 2700 Holloway Rd., Ste. 113, Louisville, KY 40299
Webb, Jacqueline, Univ. Rhode Island, Biological Sciences/CBLS, 120 Flagg Rd., Kingston, RI 02881
Westin, Magnus, Ostervagen 22B 1tr, 16953 Solna, Sweden
White Sjolander, Peta, Royal Inst. Technology (KTH), School of Technology & Health, Alfred Nobels alle 10, 14152 Huddinge, Stockholm, Sweden
Wilcock, William, Univ. of Washington, School of Oceanography, Box 357940 Seattle, WA 98195
Williams, George, Webb Institute, 298 Crescent Beach Road, Glen Cove, NY 11542-1398
Wilson, Willsingh, Motzstr. 10, 10777 Berlin, Germany
Winkel, Eric, Design Research Engineering, 46475 DeSoto Ct, Novi, MI 48377
Wysoczanski, Walter, 4155 Texas St. #11, San Diego, CA 92104-1649
Yadzinski, Edward, SUNY-Buffalo, Music, 487 Berryman Dr., Amherst, NY 14226
Yardim, Caglar, UCSD, Scripps Inst. of Oceanography, MPL/SIO-0238 9500 Gilman Dr., La Jolla, CA 92093-0238
Zhang, Yu, Univ. of Wisconsin-Madiso, Surgery, 1300 Univ. Ave., Madison, WI 53705

New Students

Ahmadi, Mahnaz, 4040 N. High St., Apt. 41, Columbus, OH 43214
Al-Hamad, Nanu, 718 W. 18th St. 2nd Fl., Chicago, IL 60616
Alexander, Erika, Brown Univ., Psychology, 89 Waterman St., Campus Box 1853, Providence, RI 02906
Amano-Kusumoto, Akiko, 1138 SW Kiley Way #132, Beaverton, OR 97006
Anderson, Christian, Washington Univ., Physics, 1 Brookings Dr., Campus Box 1105, Saint Louis, MO 63130
Atagi, Eriko, 3420 E. Covenanter Dr., Bloomington, IN 47401
Aughenbaugh, Ryan, Columbia College, 410 S Morgan St., Apt. 214, Chicago, IL 60607
Ballinger, Thomas, 2009 Skyline Dr., Wenatchee, WA 98801
Barias, Arsenia, 82-02 Rockaway Beach Blvd, Rockaway Beach, NY 11693
Barnes, Paul, 3A Spurway Parade, Woodford Ave., Ilford, IG2 6UU, UK
Barnobi, Christopher, 408 Clay St SW, Blacksburg, VA 24060
Barroso, Celia, 3556 Maplewood Ave., Los Angeles, CA 90066
Barto-Sisamout, Karen, Univ. of Arizona, SLAT PhD Program, 3338 N. Pebble Rapids Pl., Tucson, AZ 85712
Berezina, Maria Andrey, 235 Albany St., Apt. 1077, Cambridge, MA 02139
Bernstein, Alan, ARL-UT, P.O. Box 8029, Austin, TX 78713
Blom, Philip, 2216 Church St., Apt. 31, Oxford, MS 38655
Bockman, Alexander, Rensselaer Polytechnic, Architectural Sciences, 110 8th St., Troy, NY 12180
Borucki, Ewa, 16300 Mt. Baden Powell St., Fountain Valley, CA 92708
Botts, Jonathan, 2148 5th Ave., Troy, NY 12180
Brady, Michael, Indiana Univ., Cognitive Science, 838 Eigenmann, 1910 E. 10th St., Bloomington, IN 47408
Brenner, Daniel, 1510 E 9th St., Apt. 223, Tucson, AZ 85719
Brouwer, Susanne, MPI for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands
Camp, Margaret, 3133 E. Pima St., Tucson, AZ 85716
Campbell, Richard, 909 SW 12th Ave., # 411, Portland, OR 97205
Cao, Rui, 321 Univ. Village #8, Gainesville, FL 32603
Carty, Michael, Gentex Corp., Engineering, 2456 Brown Ave., Manchester, NH 03103
Carvell, Kelsey, 160F Univ. Village, Ames, IA 50010
Cascao, Irma, IMAR, Univ. of the Azores, Cais de Santa Cruz, 9901-862 Horta, Portugal
Castonguay, Paula, 715 Walker Road, Windsor, ON N8Y 2N2, Canada
Cavaro, Matthieu, CEA Cadarache, DEN/DTN/STPA/LTTS, Batiment 201, 13100 St Paul Lez Durance, France
Chanethom, Vincent, New York Univ., Linguistics, 726 Broadway, 7th Fl., New York, NY 10003
Chen, Ying, Univ. of Oregon, Dept. of Linguistics, 1290 Univ. of Oregon, Eugene, OR 97403-1290
Chen, Yu-Wen, Univ. of Minnesota, Speech-Language-Hearing Science, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455
Chen, Hong, Univ. of Washington, Applied Physics Lab., 1013 NE 40th St, APL, Seattle, WA 98105
Chen, Xiaojun, RSMAS/U Miami, AMP, 4600 Rickenbacker Cswy., Miami, FL 33149
Chiu, Chenhao, 2111 Lower Mall, Rm# 1092 SJC, Vancouver, BC V6T 1Z4, Canada
Chu, Chun Kau, Univ. of New South Wales, School of Psychology, Mathews Building, Univ.of New South Wales, Sydney, NSW 2052, Australia
Collins, Zachary, 441 N. 1220 W., Provo, UT 84601
Cooper, Angela, Simon Fraser Univ., Linguistics, RCB 6203, 8888 Univ. Dr., Burnaby, BC V5A 1S6, Canada
Counselman, David, 130 Northbrook Ln., #213, State College, PA 16803
Crabb, Michael, 6510 Pine St, Omaha, NE 68114
Daily, David, 2671 Waverly Way, Livermore, CA 94551
Darvishnarenjbon, Shahin, 217, McElwain, Univeristy Park, PA 16802
De Ryck, Tawny, Kapelstraat 163 0101, 8450 Bredene, Belgium
Demaison, Laura, 9631 W. Moonlight Bay Ln., Oak Harbor, OH 43449
Derrick, Donald, Univ. of British Columbia, Linguistics, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada
Diaz, Miriam, 340 N. Main Ave., Tucson, AZ 85701
Dickout, Verona, #1003-10303 105 St., Edmonton, AB T5J 5G3, Canada

Dietz, Mathias, Universitaet Oldenburg, Institute of Physics, Carl-von-Ossietzky Str. 9-11, 26111 Oldenburg, Germany
Echols, Eric, 1500 Bolton St., Baltimore, MD 21217
Ellis, Timothy, 7397 Caneberry Cr. NE, Keizer, OR 97303
Ellison, John, 60 Jefferson St., 1st Fl., Troy, NY 12180
Fairchild, Mark, 50 Country Meadows Ln., Monticello, KY 42633
Faragher, Stuart, St. John's College, St. Giles, Oxford, OX1 3JP, UK
Fazi, Filippo, Via Sant″Osvaldo 12, Bolzano 39100 Italy
Fonseca, William, Fed. Univ. of Santa Catarina, Noise and Vibration, Rua Douglas Seabra Levier, 163307C—Trindade, Florianopolis—SC, 88040-410, Brazil
Frank, Austin, 86 Castleman Rd., Rochester, NY 14620
Garcia, Kristen, Univ. at Buffalo, SUNY, Psychology, 206 Park Hall, Buffalo, NY 14260
Garcia, Paula, 3424 Kingsbridge Ave., Apt. 1A, New York, NY 10463
Garza-Prisby, Emmylou, Wayne State Univ, Communication Sciences & Disorders, 10705 Elgin Ave., Huntington Woods, MI 48070
Gebbie, John, 6215 NE 37th Ave., Portland, OR 97211
Ghoul, Asila, UC Santa Cruz, Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060
Giuliano, Vincent, 320 Forest Ln., Schaumburg, IL 60193
Goettlinger, Christian, Klematisstr. 10, Munich, Germany
Gogoi, Divya, 322 Univ. Village,. # 08, Gainvesville, FL 32603
Gong, Yanjun, Boston Univ., Mechanical Engineering, 110 Cummington St., Boston, MA 02215
Goodin-Mayeda, Carrie, 4411 College St., Bellingham, WA 98229
Goodwick, Jesse, 37 Burr Rd., Bloomfield, CT 06002
Gormley, Andrea, 69 Pamilla St., Ottawa, ON K1S 3K7, Canada
Greenhow, Danielle, 140 7th Ave S, Saint Petersburg, FL 33701
Harrison, Richard, 63 Southtown Road, Great Yarmouth, NR310DY, UK
Hart, Noel, 2403 21st St., Apt. 5, Troy, NY 12180
Haslam, Mara, 4981 Viewmont Cir., Holladay, UT 84117
Hayes, John, 11 N. Restin Rd., Greenwood, IN 46142
Heald, Shannon L., Univ. of Chicago, Dept of Psychology, 5848 S. Univ. Ave., B402, Chicago, IL 60637
Helms Tillery, Kate, Arizona State Univ., Speech and Hearing Science, Coor Hall 2211, P.O. Box 870102, Tempe, AZ 85287
Henry, Samuel, 1323 SW Curry St., Portland, OR 97237
Henry, Molly, Bowling Green State Univ., Dept. of Psychology, Bowling Green, OH 43403
Ho, Julio, Purdue Univ., Biomedical Engineering, 206 S. Martin Jischke Dr., Mailbox #33, West Lafayette, IN 47906
Hodgson, John, 4111 Elledge Dr., Roeland Park, KS 66205
Hollis, Harold, 2155 Purcell Ln., Alpharetta, GA 30004
Hornick, Heather, WHOI, AOPE, Co-op 202, MS 16, Woods Hole, MA 02169
Hossein Abadi, Shima, Univ. of Michigan, Mechanical Engineering, 3625 GreenBrier Blvd., Apt.160C, Ann Arbor, MI 48105
Hughes, William, 101 Carbonera Dr., Santa Cruz, CA 95060
Hyslop, Gwendolyn, Dept. of Linguistics, 1290 Univ. of Oregon, Eugene, OR 97403
Jacobs, Peter, Portland VA Medical Center, NCRAR, 3710 SW US Veterans Hospital R, Portland, OR 97239
Jangjamras, Jirapat, 1727 SW 42nd Ave., Gainesville, FL 32608
Ji, Sunjing, Univ. of Arizona, Dept. of Linguistics, Rm 200E, Douglass Building, Tucson, AZ 85719
Juergens, Tim, Medical Physics, Carl-von-Ossietzky Univ., Marie-Curie-Str.2, 26129 Oldenburg Germany
Kainada, Evia, Univ. of Edinburgh, Dept. of Linguistics, 14 1F2 Moncrieff Ter., Edinburgh, EH9 1NA, UK
Kaiser, Eden, Univ. of Minnesota, ILES, 214 Nolte Center, 315 Pillsbury Dr. SE, Minneapolis, MN 55455
Kamiyama, Takeki, ILPGA—Paris 3, LPP, 19, rue des Bernardins, 75005 Paris, France
Karpfinger, Florian, Curtin Univ. of Tech, Exploration Geophysics, GPO Box U 1987, Perth, WST 6845, Australia
Kelley, Dakota, 1607 N 49th St., Omaha, NE 68104
Keum, Kyungae, 1603 W 15th St., 001-C, Lawrence, KS 66044
Kiburz, Jace, 6105 Washington, Downers Grove, IL 60516
Kim, Eun, San Jose State Univ., Linguistics, 10116 Imperial Ave., Cupertino, CA 95014
Kirby, James, 2742 W. Chicago Ave., Chicago, IL 60622
Kniffin, Gabriel, 4500 Nicholson Rd., Apt. E34 Vancouver, WA 98661
Koch, David, 12622 NW Hamel Dr., Portland, OR 97229
Kollengode Subramanian, Ananthakrishnan, 5 Caulfield Crescent, Paralowi, Adelaide, SA 5108, Australia
Kondo, Atsumi, Boston Univ., Mechanical Engineering, 110 Cummington St., Boston, MA 02215
Kong, Lingqiang, Boston Univ., Cognitive and Neural Systems, 677 Beacon St., Boston, MA 02215
Kusch, Mark, 2027 Busse Hwy., Des Plaines, IL 60016
Lammert, Adam, Univ. of Southern California, Computer Science, 3740 McClintock Ave., EEB 400, Los Angeles, CA 90089
Leary, Del, 318 Island View Dr., Boutiliers Point, NS B3Z 1R6, Canada
Lemke, Shannon, 11008—88th Ave., Ste. 33, Edmonton, AB T6G 0Z2, Canada
Lemkes, Jennifer, 272 Greenhouse Road, Mills River, NC 28759
Lemonnier, Yann, Universiti de Sherbrooke, Ginie Micanique, 2500 Boul. Universiti, Sherbrooke, QC J1K2R1, Canada
Leone, Dorothy, 100 Cedar St., #A45, Dobbs Ferry, NY 10522
Li, Jingyu, Peking Univ., Speech and Hearing Research Sci., 2212#, No.2 Science Building, Peking Univ., Beijing, 100871 Beijing, P.R. China
Li, Zizheng, 1912 6th Ave, SW, Rm. 1412, Portland, OR 97201
Lin, Susan, 312 E. Jefferson St., Apt. 6, Ann Arbor, MI 48104
Lin, Robin, 59 E. Van Buren St., #2603, Chicago, IL 60605
Lind, Amanda, 833 West College Ave, State College, PA 16801
Littrell, Robert, 1322 S. Forest Ave., Apt. 2, Ann Arbor, MI 48104
Liu, Sheng, 1901 Union St., Apt. 321, Lafayette, IN 47904
Liu, Ching-Ju, Univ. of Wisconsin-Madison, Communicative Disorders, 1975 Willow Dr., Madison, WI 53706
Lo Forte, Dan, 155 S. Pleasant Grove Blvd., #30, Pleasant Grove, UT 84062
Luque, Jenna, 1575 Oak Ave., Apt. 13, Evanston, IL 60201
Ma, Guilin, Technical Univ. of Denmar, Electrical Engineering, Xrsteds Plads, Building 352, Rm. 111, 2800 Lyngby, Denmark
MacfarLn., Rebecca, 3-128 Fairleigh Ave., S., Hamilton, ON L8M 2K3, Canada
Maeda, Masako, 4233 Eagle Lake Dr., Bel Aire, KS 67220-1717
Matlack, Kathryn, GA Institute of Technology, Mechanical Engineering, Graduate Box 380, Woodruff School of Mech. Eng., Atlanta, GA 30332
Matteson, Robyn, 104 COAS Admin. Bldg., Oregon State Univ., Corvallis, OR 97331
Meissen, Emily, 6099 SE Frances St., Hillsboro, OR 97123
Miao, Qi, Oregon Health & Science Univ., CSLU, 20000 NW Walker Rd., Beaverton, OR 97006
Miller, Rachel, Univ. of California, Riverside, Psychology, 900 Univ. Ave., Riverside, CA 92521
Mitra, Vikramjit, Univ. of Maryland, Electrical and Computer Eng., 5105 Crossfield Ct, Unit 6, Rockville, MD 20852
Moazenahmadi, Alireza, No.2 Hamilton Pl., Mawson Lakes, Adelaide, 5095, Australia
Monteleone, Marisa, 401-240 Simcoe St., Victoria, BC V8V 1K8, Canada
Muthusamy, Hariharan, Universiti Malaysia Perlis, School of Mechatronic Engg, Block A, Jejawi, Arau 02600 Kangar, Malaysia
Muzi, Lanfranco, 7391 NE Cherry Dr., Hillsboro, OR 97124
Myrbeck, Shane, 160 1st St., 2nd Fl., Troy, NY 12180
Mysore, Gautham, 344 Olmsted Rd., Apt. 220, Stanford, CA 94305
Nam, Kyung, 245 SW Lincoln St., #3302, Portland, OR 97201
Nash, Juliette, 3051 Meade Ave., #3, San Diego, CA 92116
Natarajan, Logesh Kumar, Wayne State Univ., Mechanical Engineering, 5050 Anthony Wayne Dr., Rm. 2124, Eng. Building, Detroit, MI 48202
Nava, Emily, Univ. of Southern California, Dept. of Linguistics, GFS 307, Los Angeles, CA 90089
Oberg, Martin, 2822 W. King Edward Ave. W, Vancouver BC V6K 1T9 Canada
Oesterlein, Tobias, 801 SW Broadway Dr., Portland, OR 97201
Ogden, George, P.O. Box 873385, Vancouver, WA 98687
Oh, Grace, 110 Westbrook Way, Eugene, OR 97405
Oliveira, Claudia, DOP—Universidade dos Agores, Dept. Oceanografia e Pescas, Cais de Santa Cruz 9901-862 Horta, Portugal
Orme, Wesley, 1681 W Sandalwood Dr., Meridian, ID 83646
Phillips, Max, 3801 Clinton Pkwy, Apt. L3, Lawrence, KS 66047
Piccinini, Page, 2146 Sherman Ave., Apt. 1D, Evanston, IL 60201

Powell, Eric, 9881 S 161 St., Omaha, NE 68136
Rajoriya, Dinesh K., S-1, G.P.Hostel, I.I.T.Campus Roorkee, Roorkee 247667, India
Ramanarayanan, Vikram, Univ. of Southern California, Electrical Engineering, 3740 McClintock Ave., EEB 421, Los Angeles, CA 90089
Reed, Rebecca, JJ 315 Coventry Manor, Lawrence, KS 66049
Reese, Marc, 170 Northbrook Ln., #113, State College, PA 16803
Ripplinger, David, 897 E 300 S, Provo, UT 84606
Rohena-Madrazo, Marcos, New York Univ., Linguistics, 726 Broadway, 7th Fl., New York, NY 10003
Rosas, Jason, 1780 West 3rd St., Apt. 5D, Brooklyn, NY 11223
Russert, Jeff, 59 E. Van Buren St., Chicago, IL 60605
Russo, Mladen, Univ. of Split, FESB, Rudjera Boskovica bb., Split, Croatia
Sagers, Jason, 11411 Research Blvd., Apt. 1622, Austin, TX 78759
Saitis, Harry, McGill Univ., Music Research, Strathcona Music Building, 555 Sherbrooke St. West, Montreal, QC H3A 1E3, Canada
Salles, Ashley, 15609 Pamela Dr., Silver Spring, MD 20905
Salton, Alexandria, 23 Elm St., #4, Cambridge, MA 02139
Saracino, Anthony, Illinois Institute of Tech., 4103 N. Damen Ave., 2, Chicago, IL 60618
Sargent, Valerie, 817 Ferry St., Apt. #3, Lafayette, IN 47901
Schaerer, Zora, Konvikt. 18, 79098 Freiburg, Germany
Schultheis, Hannah, 4004 Echoridge Dr., Carrollton, TX 75007
Scott, John, Indiana Univ., Germanic Studies, 1020 East Kirkwood Ave, Ballantine Hall Rm. 644, Bloomington, IN 47405
Sharma, Nabin, UMass Dartmouth, ECE, 285 Old Westport Rd., North Dartmouth, MA 02747
Sheehan, Conor, 284 Pawling Ave, Troy, NY 12180
Sims, Michelle, #806—10020—115 St., Edmonton, AB T5K 1T2, Canada
Slatcher, Neil, 3 Brindle Heights, Chorley, PR6 8YA, UK
Son, Semyung, Unit 801, 1101 St. Paul St., Baltimore, MD 12101
Spargo, David, Unit 9, 110-116 Reserve Rd., Artarmon, NSW 2064, Australia
Srinivas, Sushma, Biomedical Engineering, 28640 Blue Pond Trail, Solon, OH 44139
Stanley, Raymond, Georgia Institute of Technology, Psychology, 654 Cherry St, Atlanta, GA 30332-0170
Sukovich, Jonathan, Boston Univ., Mechanical Engineering, 110 Cummington St., Rm. 101, Boston, MA 02215
Sung, Min, POSTECH/VATrans Lab., Mechanical Engineering, PIRO 416, POSTECH San 31 Hyoja-dong Namgu, Pohang, 790-784, Korea
Talkington, William, West Virginia Univ., Physiology and Pharmacology, One Medical Center Dr., P.O. Box 9229, Health Sciences, Morgantown, WV 26506-9229
Tausig, Benjamin, New York Univ., Music, 24 Waverly Place, Rm. 268, New York, NY 10003
Terrell, Michael, Queen Mary UL, Center for Digital Music, Electronic Engineering, QMUL Mile End Road, London, E1 4NS, UK
Tessel, Carol, 207 E. 37th St., Apt. 4J, New York, NY 10016
Tisch, Erich, 53 Creighton St., Apt. 1, Boston, MA 02130
Toltin, Abigail, 4 Cardinal Road, Sandwich, MA 02563
Trippe, Julia, 3343 Admiral St., Eugene, OR 97404
Trivedi, Uday, 185 Vandenburgh Pl., Unit A, Troy, NY 12180
Ueki, Kaori, Univ. of Hawaii at Manoa, Linguistics, 569 Moore, 1890 East-West Rd., Honolulu, HI 96822
Umile, Michael, 29 Brunswick Rd., Apt. 1, Troy, NY 12180
Utami, Sentagi, Univ. of Michigan, Architecture, 2000 Bonisteel Blvd., TCAUP R1244, Ann Arbor, MI 48105
Vasconcelos, Raquel, R. Bartolomeu Gusmao n11, 1Esq, 2780 Oeiras, Portugal
Veglio, Andrea, Univ. of Torino, Str Prov 142 Km 3.95, III Floor Div Mol Angiogenesis 10060 Candiolo, Italy
Verhulst, Sarah, CAHR, Electrical engineering, Oersteds plads Building 352, 2800 Kgs. Lyngby, Denmark
Vinke, Louis, Bowling Green State Univ., Dept. of Psychology, Bowling Green, OH 43403
Vitt, Nikolas, Pennsylvania State Univ., Acoustics, 3813 SW 34th St, Apt. 55, Gainesville, FL 32608
Walker, Katie, 1716 Jennings Dr., Bel Air, MD 21015-4800
Wall, Carrie, Univ. of South Florida, College of Marine Science, 140 7th Ave. South KRC 2121, St. Petersburg, FL 33701
Wall, Alan, 371 E 720 S, Orem, UT 84058
Walsh, Daniel, 40 East 9th St., Apt. 307, Chicago, IL 60605
Wang, Jun, 645 S. 17th St., Apt. 402, Lincoln, NE 68508
Wells, Adam, 261 N. Aberdeen Ave, Wayne, PA 19087
Wensveen, Paul, 70E Peddie St., Dundee, DD1 5LY, UK
Whitehurst, James, 24 Lancelot Rd., Stapleton, Bristol, BS16 1WG, UK
Wilkes, Daniel, Curtin Univ./CMST, Imaging and Applied Physics, Building 301 Rm.136, GPO Box U1987, Perth, WST 6845, Australia
Wong, Simpson Wai Lap, Univ. of Oxford, St John's College, Oxford, OX1 3JP UK
Wong, Devin, 1060 Wickham Dr., Moraga, CA 94556
Woodall, Ashley, 4014 Baywatch Dr., Rowlett, TX 75088
Wu, Jiang, 209 Rano Blvd, Vestal, NY 13850
Yang, Hongseok, Univ. of Sheffield, School of Architecture, 51, Park Grange Mount Norfolk, Sheffield, S2 3SP, UK
Yip, Jonathan, Univ. of Michigan, Linguistics, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48104
Yoshida, Kenji, Indiana Univ., Dept. of Linguistics, Memorial Hall 322, 1021 E 3rd St., Bloomington, IN 47405-7005
Zhang, Likun, 1580 NE Merman Dr., Apt. 310, Pullman, WA 99163
Zheng, Xinliang, Univ. of Washington, Bioengineering, 1705 NE Pacific St., Rm. N210, Seattle, WA 98105
Zhu, Na, Wayne State Univ., Dept. of Mechanical Engg, 5050 Anthony Wayne Dr., Detroit, MI 48202
Zhu, Xiaoli, 22 Old Westport Road, North Dartmouth, MA 02747
Zissimos, Ioannis, 14 Grig. Kydonion str, 17123 Nea Smyrni, Greece
Zitelli, Jeffrey, Univ. of Texas at Austin, ICES, 1172 Navasota St., Austin, TX 78702

Members Elected Fellows

D. L. Adams, M. Fatemi, M. F. Insana, R. G. Leisure, P. J. Loughlin, R. T. Muehleisen, S. N. Nittrouer, D. A. Paoletti, O. A. Sapozhnikov, K. B. Smith, A. J. Szeri,

Associates Elected Fellows

M. A. Averkiou, D. S. Brungart, C. C. Coussios, D.Dalecki, Kullervo H. Hynynen, B. Katsnelson, P. C. Loizou, D. A. Mann, D. K. Mellinger, S-I. Umemura,

Associates Elected Members

S. M. Alam, A. Alizrad, S. L. Beamer, A. Blanc, B. Blesser, P. S. Brown, J. R. Buck, J. P. Chambers, C. G. Clopper, J. M. Dalby, L. Degertekin, K. J. DeJong, G. S. Donaldson, A. A. El-Aassar, W. J. Elliot, R. A. Fleming, S. A. Frisch, D. A. Greenberg, N. O. Handegard, S. Harvey, S. In-Kyung, Y. J. Kang, R. J. Kloser, X. Liu, R. W. Lee, R. C. Maher, K. T. McNally, G. R. Mellema, K. F. McKelvie, J. R. Nickell, M. L. O'Donnell, T. M. Ortkiese, H. Patra, C. D. Petersen, L. Polka, A. E. Queenan, J. Rathsam, K. R. Richardson, T. Robinson, C. L. Rogers, T. G. Schmidt, M. A. Schwob, P. D. Thorne, M. VanDam, A. C. van der Veeke, P. Wong, C-M. Wu, R. C. Zwiebel

Student to Associate

P. Wong

Associates to Students

C. L. Abercrombie, S. L. Denes, A. M. Jaramillo

Resigned

R. Teranishi, G. S. K. Wong—*Fellows*
S. E. Forsythe, H. Herzel, T. Idogawa, E. A. Karlow, T. Maeda, L. Marraccini, T. Sasakura, S. N. Trevino, Y. Umeda—*Members*
J. R. M. da Silva, M. Greaves, K. Kishi, D. Kupersztoch, P. LafranceJ. Lekner, F. Lipsett, J. Marque, R. Narasimhan, D. Ranson, K. Scannell, J. Stratman—*Associates*
M. Cassidy, A. Costantino, S. Hettler, J. Thomas, R. L. Taylor, T. Treichel, D. Vasquez Velez—*Students*

Deceased

J. J. Faran, I. D. Groves, R. P. Hellman, D. Middleton, S. N. Wolf—*Fellows*
J. R. Ballentine, H. W. Frings, M. C. Goody, W. C. Green, J. P. Legouix, K. Mitsuhashi—*Members*
S. J. Bezuszka, E. M. Gates—*Associates*

Dropped

Robert P. Chapman, Elizabeth A. Cohen, Martin Hirschorn, Michael M. Merzenich, Bertel Mohl, Jean R. Nicolas, Robert P. Porter, Barbara J. Sotirin, Dianne J. Van Tasell, Ben T. Zinn, *Fellows*

Peter J. Alfonso, William P. Arnott, Frank E. Barber, Robert A. Berkovitz, Ryan B. Biziorek, Ian R. Brewe, Acacia J. Brunett, Gary L. Bullock, Dominique Cathignol, Angela C. Chen, Yvon Chevalier, Ya-Cherng Chu, John J. Cornyn, Otto E. Crenwelge, Huibert J. Dane, R Prakash Dixit, John F. Duda, David J. Erickson, Hee Joon Eun, Daniel J. Fedor, Vladimir N. Fokin, Margartia S. Fokina, Kyoji Fujiwara, Kelly J. Garcia, William J. Gruner, Nathan C. Hall, Ning-Ji He, Henry E. Heffner, Donald Henderson, Heinrich J. Hesselfeld, Wai-Lun W. Ho, Elaine Y. Hsieh, Richard R. Illingworth, Gary P. Jacobson, John B. Kedzierski, Liat Kishon-Rabin, Richard A. Kolano, Nicholas Kremenic, Patrick C. Kujawa, William M. Leach, James P. Lee, Ming-xuan Li, Davide Marandino, Alex I. Marash, Ken I. McAnally, Therese J. McGee, Alan W. Meyer, G Kirby Miller, Wendell R. Miller, David J. Moretti, Sam M. Mukdadi, Todd W. Murray, Rudolf M. Nielsen, Richard S. Pascucci, Hans Pasterkamp, Alan R. Reich, Stephanie C. Renaud, Ian A. Rice, L R. Roberts, Ian Roebuck, Lawrence Shotland, Isaac Shreiber, Jacob J. Skubal, Justin D. Smith, Karen B. Snell, George Spano, Richard C. Stiffler, Mark W. Sutterlin, Stephen W. Tehon, Peter R S. Terroux, Dennis Thurmond, George C. Tibbetts, Michael P. Valoski, Martin Veidt, Richard D. Vogelsong, Joseph P. Walton, Yi Kevin. Wang, Harper J. Whitehouse, Terry L. Wiley, Kenneth W. Winkler, Adriene A. Wohler, Tsung-Tsong Wu, Wen Xu, Gang Yuan, Alex Zaharakis, Daniel R. von Recklinghausen, *Members*

Nassima Abdelli-Beruh, Marianne Abramson, Jingo Adachi, James L. Adcock, Ashraf Alkhairy, Curtis l. Alpha, Jacob E. Ament, Jim N. Anderson, Peter L. Anson, Mariko Aoki, Youssef Ya. Attalla, Alyson J. Azzara, Kenneth C. Baldwin, John Y. Bang, Sarah P. Bant, Michael Barboza, Mike Barker, Craig A. Bateman, Jason R. Baughman, Ya-Juan J. Bemman, Sofia Ben Jebara, Nathan Bentall, Eric J. Berkenpas, Steven S. Bishop, Christian Bjerrum-Niese, Barbara R. Blankenship, Daniel J. Bodony, Wilco Boek, Thomas Bohlen, Joseph E. Bondaryk, Christopher B. Braun, Melanie M. Brimer, Christopher A. Brown, Daniel Brown, Chris MS. Bull, Peter N. Burns, Thomas P. Cahill, Michael Cappiello, Hatim Carim, Samuel Carter, Teresa Cervera-Crespo, Sergej Cerviakov, C David Chadwell, Chak Bun. Chan, Sang Yee. Cheon, Darron A. Chin-Quee, Mike G. Clark, Mark P. Clarke, Robert J. Coleman, John W. Coltman, Barbara K. Cone-Wesson, Maxwell T. Cotter, Katherine M. Crosswhite, John Daniel, Sorasak Danworaphong, Nicholas M. Dawes, Diane N. de Kerckhove, Thomas K. DeLillo, Caroline M. DeLong, David A. Demer, Gregoire P. Derveaux, Eric Desart, Laure Desutter-Grandcolas, Kenneth J. Doyle, John J. Dyckmans, Catharine H. Echols, Hendrik Eden, Joseph R. Edwards, Robert H. Ess, Andres Fernandez, Jill B. Firszt, Anders K. Friberg, Kenji Fujii, Richard Furse, Jakub Gadek, Stuart H. Gage, Michael A. Galloway, William F. Ganong, Oliver Gauci, Linda A. Gedemer, Cassandra A. Gentry, Alessandro Ghiotto, Buddie P. Gilstrap, Jose C. Giner, Stephen D. Goldinger, David W. Gow, David L. Groves, Judson M. Gudgel, Sandra J. Guzman, Azhal Izuan Halim, Jin-Ha Han, Ji-Yeon Han, Emad Ibrahim Elsaghir. Hanna, Ross Harwell, Elizabeth Hester, Takafumi Hikichi, Bernard R. Hill, Sara S. Holcomb, Stephen D. Holmes, Sean S. Hotchkiss, Wen-Jeng Hsueh, Chan Chun Huang, Vern A. Hult, Paul B. Hultz, Randy E. Hunt, Robert A. Hunt, Clare M. Hurtgen, Gerald R. Iannelli, Dennis Ingrisano, John W. Irza, Narciso J. Jacobo, Seung-Ho Jang, Joseph C. Jankovsky, Alexandra Jesse, Andrew M. Jessop, Roger Johnson, David J. Jones, Jayaraj Karuppiah, Kosuke Kato, John C. Keiffer, Mathieu Kemp, E Dale. Kennedy, Nick D. Kettman, Cheol-An Kim, Jaehwan Kim, Sang-Myeong Kim, Seongil Kim, Sunga Kim, Ole Kirkeby, Anssi Klapuri, Claudia Kloss, Barbara J. Kraft, Alan H. Krakauer, Richard J. Krantz, Ashok K. Krishnamurthy, Arun Kumar, Seto Kunisato, Peter Lagergren, Sam J. Lalk, Frans-Peter A. Lam, Lawrence A. Langenberg, Jeffrey B. Larsen, Richard R. Laverty, Dorothy K. Lawrence, Brian M. Leff, Maria Eva Lerin Borao, Russ Lewis, Hock Siong. Lim, Jin-Yuan Liu, Ke Liu, Sheng Liu, Samuel Lopez, David C. Lowery, Peter T. Madsen, Toshiki Maetani, Daniel J. Maguire, Jerry Mahabub, Daneh Manouchehri, Leonard C. Manzara, Marc Marc Sabat, Jean-Martial Mari, Richard J. Marini, David H. Marlin, Svein-Erik Masoy, Jerome Mathieu, Yoshinari Matsumoto, Karen J. Mattock, Larry McCracken, Deborah McDonnell, Ian J. McKnight, Neil M. McLachlan, Antonio Maria G M. Meireles, Luis Melecio, Joseph Mendoza, Jorge Mendoza L¿pez, Randy B. Metcalf, David K. Miller, Joel D. Miller, Sarah Mimran, Mark A. Moehring, Nader Mohammadi, Marcel M. Molenaar, Julia A. Mossbridge, Paulin Buaka. Muanke, Robert A. Muir, Jordi F. Mulet, James J. Murray, Rupert A G. Murray, Janice Lynn. Musfeldt, Emily B. Myers, Ron Nagar, Masahiro Nakama, Yang-Hee Nam, Keivan Negahbani, John M. Neill, Silvana Neves, Eddie YF. Ng, Bhattad Nilesh Bhattad, Jim R. Nokes, Timothy F. Noonan, Susan J. Norton, Yang Ki Oh, Bernard Olschewski, Taner Onsay, Wasim A. Orfali, Daniel L. Osborne, Soonkwon Paik, Insu Paek, William Pagliuca, Caroline Palmer, Sally J. Pardue, Joo-Bae Park, Seok Tae. Park, Frederic J. Patat, James D. Perea, Gary E. Perren, David R. Peterson, Harry W. Peterson, Phillip P. Peterson, Vadavalli V. Phani Kiran, Andrzej Pietrzyk, Stephen P. Pitt, Victor H. Poblete, Jeffrey B. Pride, Jessica L. Pristera-Goodson, Marios Psaras, Damon C. Quesenberry, James L. Rae, David B. Rankin, Janice L. Reed, Maria Ribeiro, John M. Rice, Yongrae Roh, Vladimir E. Rok, Alfred B. Roney, Jason P. Rudzinsky, Philip Sabransky, Ramy S. Sadek, Timothy E. Sandrik, Rebecca A. Scarborough, Edward K. Scheer, Steven L. Schmidt, Peter L. Schuck, Anthony A. Scott, Eric Heinz Seifert, Hakan A. Serafettinoglu, Amee P. Shah, Ryota Shimokura, Arne Sieber, Christian M. Skinner, Philip W. Smith, Phil Solomon, Limin Song, Cary Sorenson, Edmund A. Spencer, Jaclyn B. Spitzer, K. Srinivasan, Erik A. Steiner, Gerald Steiner, Timothy A. Stiles, Emily T. Stoneham, Alan J. Subkey, Yang Sun, Noriko Suzuki, Lawrence P. Swist, Matthew E. Sylvest, Joseph L. Tai, Dongling Tao, David J. Thomas, Takeshi Tokashiki, Chiu-Yu Tseng, Ted Uzzle, Luc Van Immerseel, Sander J. van Wijngaarden, Deborah A. Vickers, Steve J. Volaric, Ted J. Vornbrock, Makoto Wada, Steven J. Walker, Edward P. Walters, Tongan Wang, Jeremy Webster, Robert Webster, Benson Wenceslas, Ashley Whicker Harkrider, Michael Wild, Eric Wolfe, Kwangsee A. Woo, Philip M. Wort, Jie Xiao, Huiyu Yang, Liyong Yang, Carlos L. Yapura, Imai Yasutaka, Kiyoko Yoneyama, Taeseok Yoon, Tim Young, Gongqiang Yu, Michelle Yuen, Nikolai P. Zagrai, Mary L. Zampini, Paulo H T. Zannin, Zhiewei Zhang, Karin Zimmer, *Associates*

Jad Abouzeid, Rudi Agius, Abdulrahman A. Al-Wohaibi, Marilee A. Andrew, Kofi A. Anguah, Jason P. Anick, Ashli J. Armstrong, John R. Ballard, Adam Bauer, Michael J. Berkowitz, Magdalena Blaszak, Sandra H. Blumenrath, Anais Bodson, Sandra E. Bohn, Mark V. Boyle, Edmund Brekke, Kathleen S. Brennan, Julie N. Buchan, Kristi A. Buckley, David R. Burd, Deborah L. Burleson, William A. Burnett, Lisa J. Burton, Lynnika Butler, Stacie K. Byrne, Michael C. Cain, Michael S. Caldwell, Suzanne P. Carr, Brandon M. Casper, Nancy F. Chen, Xi Chen, Zhixin Chen, Susie Choi, Marc Choiniere, Christiana Christodoulou, Mehgan Clark, Mandy L H. Cook, Huw R. Cooper, Kimberly A. Cordray, Jennifer L. Cornish, Andrew R. Covato, Timothy J. Crone, Sean A. Davidson, Ben Davis, Jeff J. Defoe, Alex A. Del Giudice, Steve J. Derezinsk, Jamie L. Desjardins, Sasha Devore, Christopher R. Dillon, Bianca Dimulescu, Stephen B. Dobie, Hugo Douville, James Drechsel, Nancy A. Driscoll, Connor R. Duke, Mikila Duke, Mary Dungan, Yumiko Enyo, Sarah A. Epps, Zhiliang Fan, Kimberly Farinella-Bocian, Chelsea D. Fenush, Daniel C. Finfer, Etan Z. Fisher, Adrian Garcia-Sierra, Mysore J. Gautham, Bruno Gauthier, Jean-Philippe Gauthier, Bradley E. Giardiello, Maria L. Gomes, Juan M. Gonzalez, Samantha R. Gonzalez, Tomoya Goto, David R. Grandall, Sherry M. Grazer, Jerry Gregoire, Dakota Gundy, Ryan A. Henry, Tracy Hickman, Eric W. Howell, Renee Hung, Hyekyung Hwang, Danielle Inverso, Ali Israr, Michelle A. Jarick, Jaehoon Jeong, Changmo Jeung, Cheng Jin, Stefan Johann, Stacy L. Johner, John Jones, Thomas Kannanaikkel John, Han Kao, Brian A. Kappus, Daniel M. Khader, Daniel Khalaf, Matthew D. Kleffner, Michael L. Kuntzman, Puxiang Lai, Jeremy A. Lambert, Janet M. Lane, Gregor Langer, Adrian M. Larbi-Cherif, Kent C. Lau, Duy Nam. Le, Nicolas LeDantec, Hye-Sook Lee, Jae Hee. Lee, Ji-Yeoun Lee, Kyogu Lee, Mark L. Leger, Diane F. Lesley-Neuman, Emmanuel E L. Levitte, Phoitack Lew, Chi Nin Herman. Li, Yao-ju Lin, Ruoshui Liu, James D. Lloyd, Joel B. Lonzaga, Leonard V. Lopes, Kritsakorn Luangvilai, Stephen D. Lynch, Ashley N. Makarowski,

Nicholas J. Manzi, Deborah Markel, Margaret S. Martei, Yuka Matsugu, Randy T. May, Annette G. Mazevski, Scott D. McDermott, Kathryn F. McGee, Ryan D. Mcmullan, Antje Mefferd, Joseph J. Mendis, Jonathan Miller, Robert D. Miller, Kay C. Minn, Taniya Mishra, Andrew F. Moore, Kristy A. Moore, Karen E. Morian, Mark W. Muller, Molly F. Murphy, Stephen R. Murray, Brian M. Napoletano, Akshay Navaladi, Margaret C. Ney, Doug C. Niemann, Isabelle C. Noirot, Pelham D. Norville, Steve P. O'Connor, Meredith A. Oakey, Jorge Oberreuter, Stefano Oberti, Jason W. Ogden, Anthony O. Okobi, Justin J. Oldham, Aybike Ongel, Edward Paes, Jonathan Page, Elizabeth Panzer, Jason M. Pappafotis, Iain M. Parnum, Andrew M. Patterson, Anthony E. Petrites, Scott M. Philips, Rafal W. Pietruch, Joseph M. Pisano, Jay M. Pleckham, David M. Pope, Peter S. Popolo, Anne Pycha, Luke Rackers, Luke N. Rasmus, Kitti Rattandait, Ashley Ringer, Sarah A. Rowland, Janine Sadaka, Melissa L. Santerre, Theresa M. Scarlotta, Justin M. Schaefer, William L. Schattner, Lauren G. Schneider, Troy D. Schultz, Daniel M. Sebastian, Samantha E. Sefton, Molly S. Shilman, Augusto Silva, David J. Silver, Alexandra M. Sinclair, Dustin Sirny, Ahmad T. Smadi, Adam R. Smith, Kathryn L. Starke, Nicholas L. Statzer, Thomas L. Steen, Clara Suied, Elizabeth A. Sukara, Olimpia M. Szczepaniak, Megumi Takenaka, Ryota Tanaka, Kristen A. Taylor, Kuen-Shian Tsai, Meng Vang, Heike I. Vester, Lori E. Vidal, Emile Vrijdags, Thomas c. Waite, Xin Wang, Yuko Watanabe, Samuel N. Weinberg, Charles E. White, Nicole L. Wiessner, Sonia L. Wilkie, E. Courtenay Wilson, Claire K. Winchell-Manning, Sylvie M. Wodzinski, Alan C. Worf, John V. Xavier, Jinshan Xu, Min Yang, David Yeh, KunLung Yen, Aleksandra Zaba, Franck B. Zagadou, Benjamin R. Zendel, Yuxuan Zhang, Yixin Zhao, Zhuo Zheng, *Students*

Fellows	923
Members	2068
Associates	3373
Students	1128
	7492

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.

A Breviary of Seismic Tomography: Imaging the Interior of the Earth and the Sun

Guust Nolet

Cambridge University Press, Cambridge, 2008. xiv+344 pp. Price $65.00 (hardcover). ISBN 978-0-521-88244-6

"The Earth is a laboratory, but one that is very different from those in experimental physics, where we are taught to carefully design an experiment so that we have full control."

This quote, taken from Chap. 14 of this book, reflects one of the main challenges seismologists are faced with in their quest for imaging Earth's interior using uncontrolled seismic waves observed by a sparse and irregular network of seismometers. Professor Guust Nolet, a leading scientist in the field of global seismology, has succeeded in writing a very readable treatise on the many aspects of imaging the interior of Earth and, albeit with less emphasis, the Sun. The book covers the underlying theory of seismic wave propagation and scattering, the practical aspects of seismic observations, and, last but not least, the theory and practice of modern seismic inversion and imaging methodologies. The existing approaches to imaging Earth's interior range from normal mode analysis to reflection and transmission tomography. The emphasis in this treatise is on seismic transmission tomography, a methodology related to medical x-ray and ultrasonic tomography, ocean acoustic tomography, and cross-well tomographic imaging in exploration geophysics. Some of the specific challenges of seismic tomography are related to the fact that the interior of Earth is inhomogeneous at many scales and to the already mentioned aspect of the "uncontrolled experiment." The author addresses these challenges with the authority of more than 25 years research experience in seismic tomography.

The book is written for seismic practitioners and students. The author has chosen for a somewhat informal style, with much more emphasis on physical explanation than on mathematical rigor. This is not to say that the mathematics is compromised, but only that the author has decided to limit the treatment of the underlying mathematics to what is needed for understanding the principles of seismic tomography. Readers interested in more extensive mathematical derivations are guided to the relevant literature. The physical explanations are often chosen from everyday life: once you have read the explanation why the sky is blue and clouds are white you will never forget the essential difference between Rayleigh and Mie scattering. What I consider very valuable in this book are the many exercises for students and, in particular, the references to software repositories, which help students improving their understanding of tomography and practitioners analyzing their data.

Let me further substantiate my appreciation of this book with a brief chapter-by-chapter discussion. The introductory chapter mentions a presentation by Keiti Aki at an AGU Conference in 1974, in which arrival times of seismic waves were interpreted in terms of an image, as the starting point of modern seismic tomography. This chapter further summarizes the early successes of seismic tomography with respect to geophysical discoveries, such as images of slabs subducting into the lower mantle and evidence of major low velocity regions near the core-mantle boundary, so-called superplumes. It concludes by mentioning that further improvement of the images requires a finite-frequency approach to tomography, accounting for the fact that arrival times are sensitive to perturbations within Fresnel zones around the rays.

Chapter 2 starts with the fundamentals of acoustic and elastodynamic wave theory and derives the differential equations for ray tracing in inhomogeneous isotropic Earth. It also contains a section on the corrections that are required for ray tracing in the Sun, where the flow of the medium plays a role. I like the fact that the author does not complicate things more than needed. For example, instead of giving a formal definition of correlation distance, he mentions in a footnote that "for the purpose of this book it is sufficient to define the correlation distance as the distance over which a parameter still looks smooth." Chapter 3 introduces several approaches to ray tracing, i.e., solving the differential equations derived in Chap. 2, and discusses aspects such as convergence and stability over long distances. The author alternates theory with everyday life examples, e.g., by explaining the shortest path method as finding the shortest route between road signs in a town. Chapters 4 and 5 conclude the fundamentals of the theory for acoustic and elastodyamic body waves with a discussion of first order wave scattering (the Born approximation) and an overview of amplitude effects, such as geometrical spreading, anelastic damping, and scattering loss.

Chapter 6 discusses the various aspects of travel time observation, such as the determination of onset times, wavelet estimation, and determination of differential times via cross-correlation. It also includes a brief discussion on the sliding transition point between signal and noise and the recent developments of retrieving Green's functions from the cross-correlation of ambient noise. The interpretation of travel times is discussed in Chap. 7. After an introduction on wave front healing due to finite frequencies, it is shown that travel times are affected by heterogeneities near the ray. The sensitivity to these heterogeneities is quantified by the so-called Fréchet kernel. Chapter 8 discusses the observation and interpretation of amplitudes along similar lines as the previous two chapters, including discussions on amplitude healing and amplitude Fréchet kernels.

Chapter 9 is dedicated to normal mode analysis. Since Earth is a finite body with a stress-free surface, the solution of the wave equation yields a discrete spectrum of eigenfrequencies. For very low frequencies, below 10 mHz, these frequencies are separated well enough to be measured and analyzed, leading to an estimate of the very long wavelength features of the Earth's interior. The models obtained by transmission tomography should match these features at the low frequency end of their spectrum.

The treatment of surface waves, the most prominent arrivals on a seismogram, is complicated because surface waves do not have clearly defined wave fronts, like body waves, nor do they exhibit a discrete spectrum, like normal modes. In Chaps. 10 and 11, the author deals with this "ray-mode duality," discusses how to measure the dispersion, and develops Fréchet kernels for finite frequency surface waves.

The following chapters address the various aspects related to the tomographic inversion itself. Chapter 12 deals with the art of model *parametrization*, i.e., finding the delicate balance between not imposing too much smoothness on the model, while allowing the data misfit to be sufficiently small and at the same time avoiding too much *overparametrization*. Parameters such as the ellipticity of Earth, surface topography, crustal thickness, and instrument response cannot be estimated from the data. Chapter 13 discusses how these parameters can be included as model corrections prior to inversion, assuming of course that *a priori* information on these parameters is available. Chapter 14 discusses the actual linear inversion. It reviews the basic aspects of maximum likelihood estimation, singular value decomposition, Bayesian inversion, and information theory. An inversion result only makes sense if one knows its limitations. Therefore, a chapter is dedicated to resolution and error analysis (Chap. 15).

In the first 15 chapters, Earth and the Sun were assumed to be isotropic. Chapter 16 discusses some of the modifications that are required to take anisotropy into account. In the final chapter (Chap. 17), the author looks ahead at promising new developments in imaging and observation, such as the analysis of multiple scattering, the application of non-linear inversion, and the strive toward global coverage of seismic sensors with programs like U.S. Array. The latter program involves a dense regular array of temporary seismic stations, moving across the continent of the United States from 2005 to 2015. This program constitutes an important step in the direction of "the Earth as a carefully designed experimental physics laboratory," and will contribute to obtaining sharper and more reliable images of the interior of our planet.

 0001-4966/2009/126(4)/2130/1/$25.00

Being an exploration geophysicist, I was only moderately familiar with the methodologies applied in global seismology. This very readable and inspiring book has brought my knowledge and understanding of this very interesting research field to a higher level. To my opinion, this book should be on the shelves of every seismology student and every practitioner of global seismic tomography. Moreover, it may be a very useful source for researchers in other disciplines employing tomography, such as ultrasonics, ocean acoustics, and exploration geophysics.

KEES WAPENAAR
Professor of Applied Geophysics,
Department of Geotechnology,
Delft University of Technolog,
y, Stevinweg 1,
2628 CN Delft, The Netherlands

Springer Handbook of Speech Processing

Jacob Benesty, Mohan M. Sondhi, and Yiteng Huang

Springer, 2008. 1176 pp. Price: $199.00 (hardcover). ISBN: 978-3-540-49125-5

Speech technology has made great strides over the past 3 decades. Automatic speech recognition (ASR), text-to-speech synthesis, voice and language recognition, speech enhancement, and auditory prostheses have all matured during this interval. The *Springer Handbook of Speech Processing* describes many of the methods and practices used to accomplish these advances. It is a highly technical tome, with most of the 53 chapters laden with equations, illustrations, and tables. The quality of writing is excellent. The prose is clear, simple to understand, and succinct. The illustrations are professionally drawn and uniform in format and appearance. The volume has the look and feel of a professional textbook, an impressive feat given that 85 authors were involved. The Handbook would be well suited as a textbook for an advanced graduate-level course in speech technology and engineering.

A typical chapter begins by briefly summarizing its contents and providing a brief historical overview. More technical (i.e., mathematical) material follows. Many chapters conclude with a discussion of commercial applications as well as a brief summary. In short, the volume strives to strike a balance between the practical and the theoretical, and it usually succeeds.

The Handbook consists of nine sections: (1) Production, Perception and Modeling of Speech, (2) Signal Processing in Speech, (3) Speech Coding, (4) Text-to-Speech Synthesis, (5) Speech Recognition, (6) Speaker Recognition, (7) Language Recognition, (8) Speech Enhancement, and (9) Multichannel Speech Processing. A complete listing of chapter titles can be found at http://www.springer.com/engineering/signals/book/978-3-540-49125-5?detailsPage=toc. A particularly useful feature is the accompanying DVD, which contains a fully searchable electronic version (PDF format) of the Handbook. Its interface allows the reader to traverse the text in a highly intuitive way, making the book a pleasure to read in electronic form.

As with any book of this length and scope, some of the chapters are more successful than others in conveying the essence of a field. Particularly detailed and useful are the chapters on pitch extraction, speech synthesis, automatic speech recognition, and environmental robustness; many of these are definitive treatments and should prove useful for years to come. Particularly helpful is the detailed discussion of experimental and computational data, which serves to clarify and enhance the theoretical sections through concrete examples.

Perhaps the volume's greatest topical weakness is its scanty treatment of speech perception and production. An additional chapter or two on the neuroscience and cognition of spoken language (and its visual analog) would have been welcome. The chapter on commercial applications of automatic speech recognition is dated due to significant changes in the industry over the past 4 years.

Another weakness of the Handbook is its focus on the past and present, and relative neglect of the future. Only a few chapters discuss future trends in a meaningful way, the most notable example being "Towards Superhuman Speech Recognition," which provides a superb description of how ASR systems may function 10–20 years hence. Also lacking is a concerted attempt by the editors to link the chapters into an overarching theoretical framework. Given the book's broad scope, this is understandable. However, many readers will wonder what ties the chapters together other than a focus on speech technology.

Many of the authors (as well as the editors) have had a professional association with Bell Laboratories at some stage of his/her career. Fortunately, this slant generally enhances the Handbook's utility based on Bell Labs' distinguished research record. However, approaches pioneered at other institutions are also well represented.

The bibliography accompanying each chapter is extensive and comprehensive. Each citation in the electronic version is linked to the appropriate reference in the text, greatly facilitating its use. Occasionally, a chapter's bibliography is overly selective, reflecting the authors' particular point of view (e.g., the chapters on speech perception and nonlinear cochlear processing).

In summary, *The Springer Handbook of Speech Processing* is a first-rate production, providing a definitive treatment of the methods and techniques used in contemporary speech technology. Although its cost is high, the Handbook's superb quality and comprehensive treatment of highly technical material should prove an attractive investment for advanced students and speech engineering professionals.

STEVEN GREENBERG
Silicon Speech,
Santa Venetia, CA 94903

REVIEWS OF ACOUSTICAL PATENTS

Sean A. Fulop
Dept. of Linguistics, PB92
California State University Fresno
5245 N. Backer Ave., Fresno, California 93740

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ANGELO CAMPANELLA, *3201 Ridgewood Drive, Hilliard, Ohio 43026-2453*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prairie, Minnesota 55344*
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*

7,520,330

43.20.Tb SYSTEM AND METHOD FOR LIMITING VORTEX-INDUCED VIBRATIONS ON AN OFFSHORE PRODUCTION RISER

Cédric Le Cunff *et al.*, assignors to Institut Francais du Petrole
21 April 2009 (Class 166/350); filed in France 16 November 2001

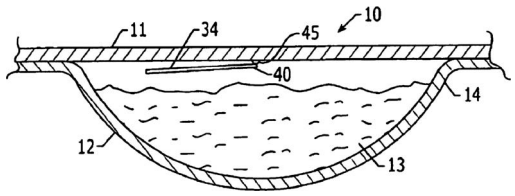
This patent describes a method of reducing fatigue in an oil-drilling platform riser (piling) by actively driving vibrations in it. The inventors mention that fatigue due to vibrational modes excited by vortex shedding in ocean currents is a significant factor in reducing the service lifetime of oil-drilling platforms. Their goal is to reduce the amplitude of such vibrations, and they claim two novel approaches to this end. First, they claim that driving the top (platform) end of the riser at a frequency off resonance can slave the oscillation to the new drive frequency and establish it at a lower amplitude than if left to freely interact with the current. Second, they claim that by use of active control in the form of magnetorheological dampers at the top end connecting to the platform they can effectively impose a fixed-velocity boundary condition at the top end of the riser, which also shifts the frequency and amplitude of the oscillations. These are pretty unusual claims, and just how a damper effects a constant-velocity condition is unclear, but they do show some data (from simulations of the dynamics) to back up their claims.—JAH

7,429,127

43.35.Rw MAGNETOACOUSTIC SENSOR SYSTEM AND ASSOCIATED METHOD FOR SENSING ENVIRONMENTAL CONDITIONS

Dwight Sherod Walker and Michael Bernard James, assignors to Glaxo Group Limited
30 September 2008 (Class 374/109); filed 23 April 2003

A magnetoacoustic pharmaceutical blister pack 11-12 environmental sensor 10 is claimed. Magnetic strip 40 coated with an environmentally sensitive (e.g., to moisture) material 0.005–10 μm thick, is suspended by one end 45. An interrogating signal is fed to a proximate coil (not shown) that will determine the mechanical resonance frequency (typically 40–50 kHz). Calibration is achieved by precise control of the physical dimensions of strip 40 and its coating thickness.—AJC

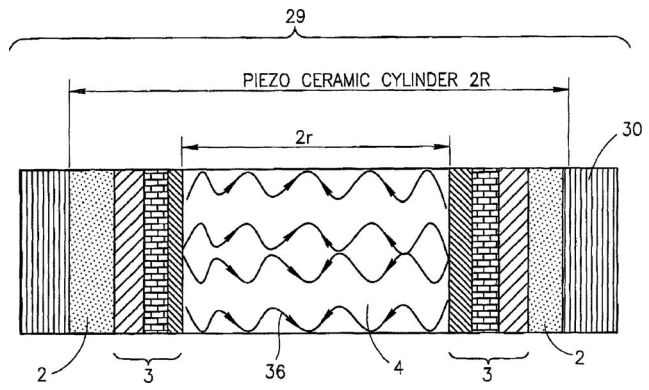


7,431,892

43.35.Wa APPARATUS FOR STERILIZING A LIQUID WITH FOCUSED ACOUSTIC STANDING WAVES

Jona Zumeris *et al.*, assignors to Piezo Top Limited
7 October 2008 (Class 422/128); filed 25 September 2002

Sonic transducer 3300 for sterilization of flowing material 3312-3302-3304-3314 is claimed. Piezoelectric element 3306 is driven in several vibration modes simultaneously to produce sound pressure waves of several atmospheres' amplitude in passing material 3302-3304. Such waves at a few megahertz (3306 thickness mode) will break up microorganisms. Other resonance modes of 3306 are longitudinal (kilohertz range) and ring mode (20–50 kHz).—AJC



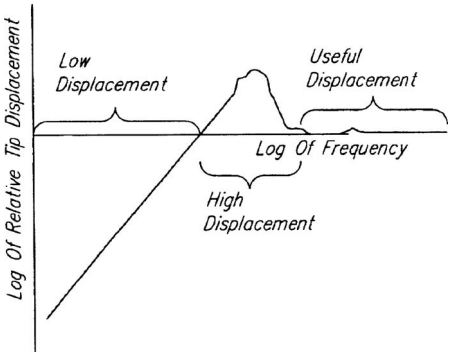
 0001-4966/2009/126(4)/2131/12/$25.00

7,521,841

43.38.Ar STRAIN ENERGY SHUTTLE APPARATUS AND METHOD FOR VIBRATION ENERGY HARVESTING

Dan J. Clingman and Robert T. Ruggeri, assignors to The Boeing Company
21 April 2009 (Class 310/339); filed 1 February 2006

In this patent, the authors disclose a method of coupling a bistable mechanically resonant structure to a piezoelectric bar transducer so as to increase the frequency range of mechanical to electrical energy conversion of the structure. This "energy shuttle apparatus" is claimed to be able to harvest vibrational energy efficiently in either of two stable states which differ in operating frequency, so as to increase the range of optimal energy harvesting performance of the device. This is all supposed to go on without any control inputs, as a passive operation of the device. There do not seem to be any calculations or data to back up the claims that this results in an improvement in performance, and one wonders how the system is ever supposed to switch autonomously from one state to the other in order to optimize its behavior. Why is this an improvement over operating the two piezoelectric resonators separately and combining their outputs with diodes?—JAH

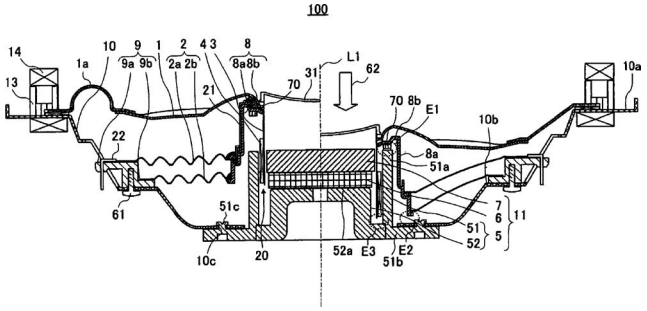


7,499,563

43.38.Dv SPEAKER DEVICE

Tomoyuki Watanabe, assignor to Pioneer Corporation
3 March 2009 (Class 381/412); filed in Japan 19 March 2004

Buffer member 70 is affixed to bent portion 8b of loudspeaker 100 to reduce the impact when the vibrating assembly is driven hard against part 52 of yoke 5 by an excessive driving signal. Buffer member 70 can also be mounted in areas E2 or E3. In the right portion of the figure, one can see the extreme stress in surround la and spiders 2 compared to the rest position of same shown in the left part of the figure. What can happen when the speaker moves in the other direction is left to the reader to imagine.—NAS

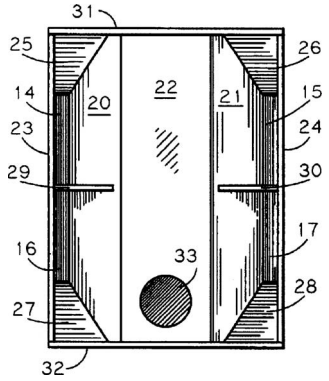


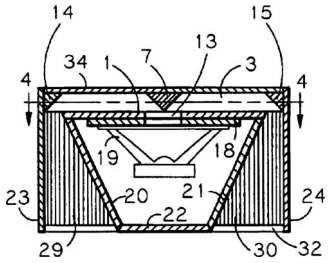
7,520,368

43.38.Ja HORIZONTALLY FOLDED REFLEX-PORTED BASS HORN ENCLOSURE

Dana A. Moore, Bothell, Washington
21 April 2009 (Class 181/155); filed 8 June 2007

This is the third in a series of simple folded horn designs from the same inventor. In this case, the footprint is rectangular and the rear chamber can be vented.—GLA



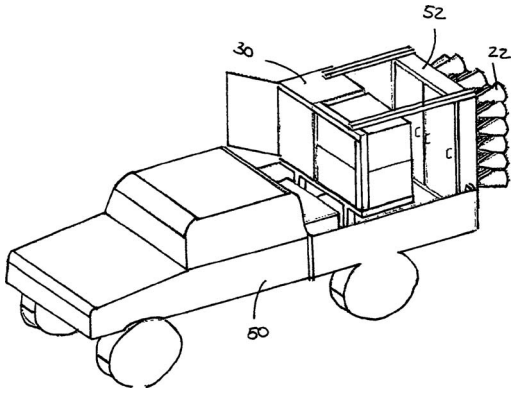


7,522,473

43.38.Ja CONTROLLED ACOUSTIC BEAM GENERATOR FOR CROWD CONTROL

Zvi Zlotnik and Eitan Zeiler, assignors to Electro-Optics Research and Development Limited
21 April 2009 (Class 367/139); filed in Israel 11 November 2003

This amusing patent discloses the design of an acoustic beam generator for mounting on various vehicles to stun and amaze crowds into submission. In one embodiment, what appears to be a rocket truck is equipped with an array of hydraulically-steered horn loudspeakers that can be played across the crowd. There is also a nautical variant (for crowds of swimmers) that mounts on a boat.—JAH

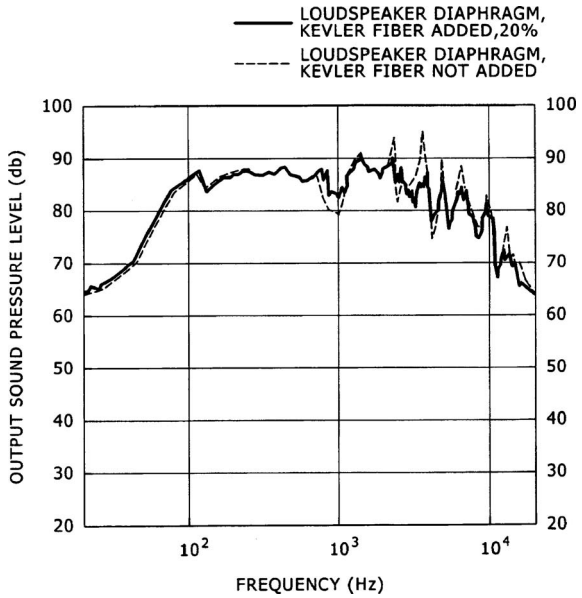


7,527,124

43.38.Ja LOUDSPEAKER DIAPHRAGM

Kunihiko Tokura *et al.*, assignors to Sony Corporation
5 May 2009 (Class 181/169); filed in Japan 10 August 2005

This patent discloses an improved formula and method for making a composite polypropylene loudspeaker diaphragm with enhanced internal losses. Polyamide fibers are dispersed in a direction perpendicular to the resin flow direction whereby the internal loss is improved. Comparative response curves are included.—GLA

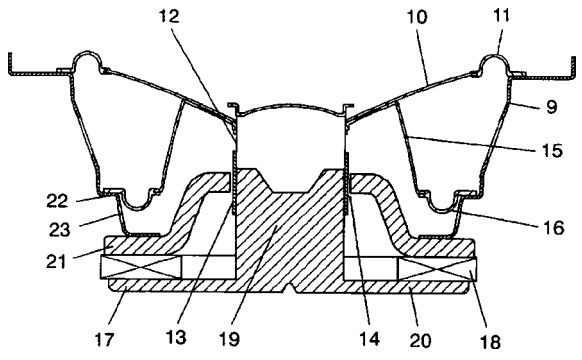


7,532,736

43.38.Ja SPEAKER

Osamu Funahashi, assignor to Panasonic Corporation
12 May 2009 (Class 381/412); filed in Japan 27 August 2004

Instead of a conventional centering spider, this loudspeaker design employs a second half-roll suspension 16. The goal is to improve linearity at large cone excursions. The patent text emphasizes that the symmetrical arrangement of suspensions 11 and 16 reduces distortion, but this feature is not included in the patent claims. Although not mentioned anywhere in the patent, the geometry shown would also be expected to counteract rocking modes.—GLA

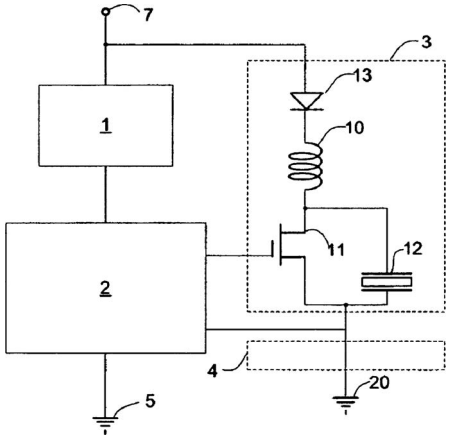


7,505,600

43.38.Lc PROCESSOR CONTROL OF AN AUDIO TRANSDUCER

Joseph E. Dryer, assignor to Floyd Bell, Incorporated
17 March 2009 (Class 381/111); filed 30 March 2005

An audio system consisting of regulator block 1, controller block 2, driver block 3, and feedback block 4 offers users of piezoelectric transducers a system that can generate a variety of audio tones. Block 1 is a voltage regulator, such as a 78L05. Block 2 is a programmable controller, such as a PIC12C671. Block 3 is as illustrated but can have several topologies, including those that provide a bipolar drive by mirroring the diode-inductor field-effect transistor and adding a second drive output from block 2. Block 4 can be configured to provide a feedback path to block 2.—NAS

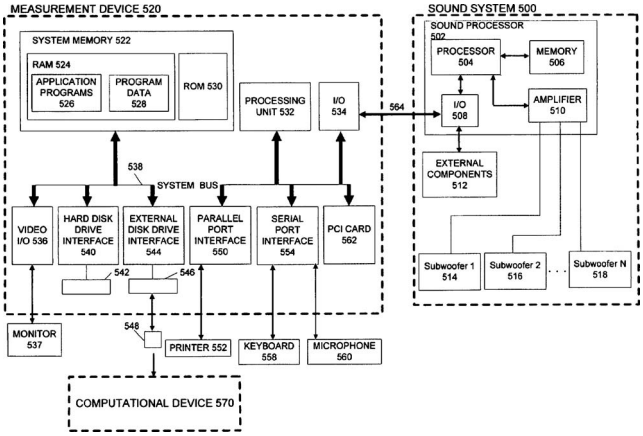


7,526,093

43.38.Lc SYSTEM FOR CONFIGURING AUDIO SYSTEM

Allan O. Devantier and Todd S. Welti, assignors to Harman International Industries, Incorporated
28 April 2009 (Class 381/59); filed 10 October 2003

This fairly long patent includes 42 illustrations. It is one of a group of three related patent applications filed on the same day. All three deal with the problem of optimizing low frequency reproduction in a listening room. In the method proposed here, one or more types of loudspeakers are tested at one or more possible locations in relation to one or more listening locations.



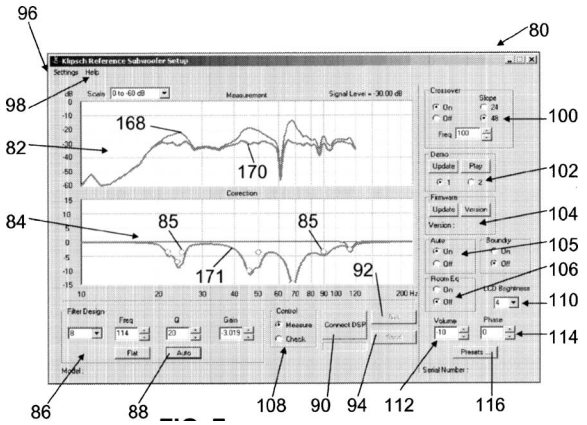
A computer program then performs some kind of statistical analysis to determine the optimum loudspeaker types, quantities, and locations, along with optimum equalization and delay for each loudspeaker. The 66 patent claims are not much more specific than the preceding two sentences, but the concept is interesting and the patent is easy to follow.—GLA

7,529,377

43.38.Lc LOUDSPEAKER WITH AUTOMATIC CALIBRATION AND ROOM EQUALIZATION

Fawad Nackvi and Jon Douglas Zenor, assignors to Klipsch LLC
5 May 2009 (Class 381/103); filed 28 July 2006

This is another patent dealing with the process of equalizing the low frequency response of loudspeakers in listening rooms. The method described makes sense and the patent is easy to follow. The equalization procedure compares measured response with a reference response and then



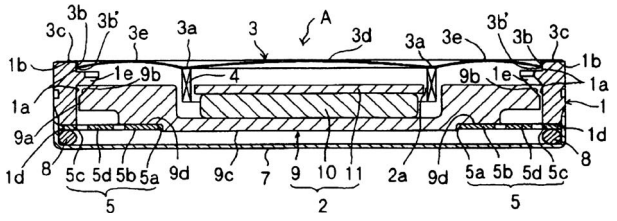
attenuates peaks, with no attempt to fill in narrow dips—a well tested strategy. The problem for this reviewer is that every one of the steps set forth in the 35 patent claims seems to be familiar prior art.—GLA

7,529,380

43.38.Si MULTIFUNCTION-TYPE VIBRATION ACTUATOR AND MOBILE TERMINAL DEVICE

Minoru Ueda *et al.*, assignors to Namiki Seimitsu Houseki
5 May 2009 (Class 381/398); filed in Japan 22 March 2004

This is a combination loudspeaker and vibrator driven by a single coil 4. At voice frequencies, it operates as a typical moving coil loudspeaker. Diaphragm 3 is fixed at its perimeter so that portion 3e acts as the outer suspension, which is familiar prior art. The novel feature is the little step 3b, which is formed as part of the diaphragm. At this point some curious assertions are made. "This rising portion 3b serves as a corrugation to decrease



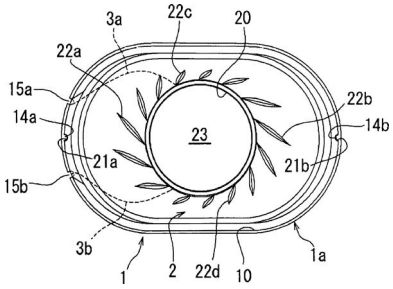
the lowest resonance frequency…" Is a right-angle bend inherently more compliant than a shallower angle or a groove? Even more curious is the statement that extended portion 3c, which is bonded to mounting surface 1b, "…enlarges the entire radius of the diaphragm 3 and improves the acoustic characteristics."—GLA

7,529,381

43.38.Si MULTIFUNCTION-TYPE VIBRATION ACTUATOR AND PORTABLE COMMUNICATION EQUIPMENT

Shoichi Kaneda and Minoru Ueda, assignors to Namiki Seimitsu Houseki Kabushiki
5 May 2009 (Class 381/403); filed in Japan 27 February 2004

The basic mechanism of this device and its inventors are shared by companion U.S. Patent 7,529,380. In this case, the stepped diaphragm edge



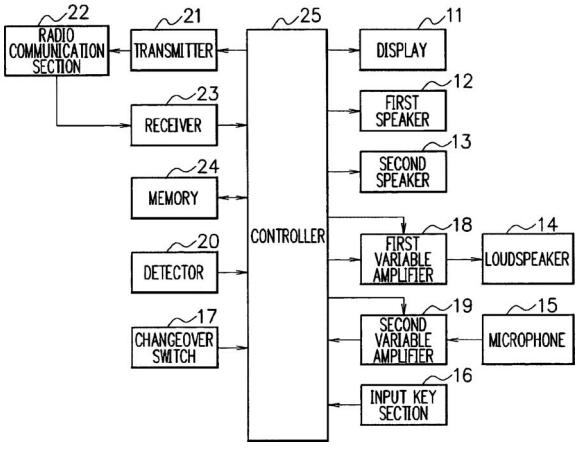
detail is not used. Here, the novel feature is an elongated diaphragm 2 with a dome-shaped center section 23.—GLA

7,529,570

43.38.Si FOLDING CELLULAR PHONE

Hirobumi Shirota, assignor to NEC Corporation
5 May 2009 (Class 455/575.3); filed in Japan 20 April 2004

It is not unusual to find that a patent abstract describes a great deal more than is actually covered by the patent claims. In this case the opposite is true. According to the abstract, the microphone sensitivity of a folding cellular phone is varied depending on whether the case is open or closed. However, the patent claims also include sensing call start instructions, determining whether a call exists, and adjusting send and receive gains based on combinations of these various factors.—GLA

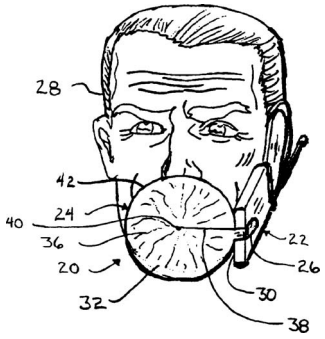


7,532,719

43.38.Si VOICE ABSORBER FOR PORTABLE TELEPHONIC DEVICES

Robert M Snodgrass, Spring Lake, Michigan
12 May 2009 (Class 379/441); filed 22 March 2005

The invention disclosed here is a little tent of absorptive material that clips onto a cellular phone. It is intended to act as an acoustical black hole, enabling its user "…to engage in a more private conversation and limit the disruption to individuals nearby the user during the conversation."—GLA

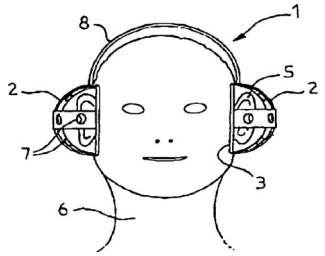


7,532,734

43.38.Si HEADPHONE FOR SPATIAL SOUND REPRODUCTION

Hong Cong Tuyên Phan, F-94460 Valenton and Ambroise Recht, F-75015 Paris, both of France
12 May 2009 (Class 381/370); filed in France 29 April 2003

This surround sound headphone design, like a number of earlier patents, takes the form of a pair of miniature listening rooms clamped over the user's ears. In this instance each earphone includes at least five speakers. The speakers are all non-directional—not too difficult for something the size



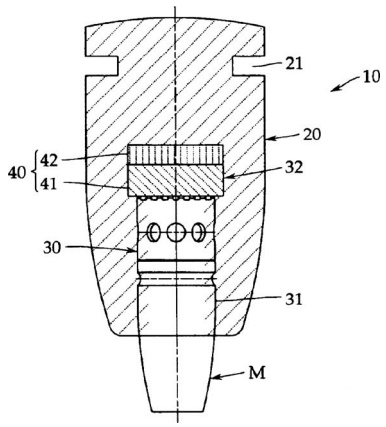
of an aspirin tablet—so that a spatial quality of sound can be produced "…by application of a Huygens Fresnel principle." The intent is to re-create sound waves moving across the outer ear as if they had been generated by individual sound sources in a listening room.—GLA

7,496,208

43.38.Tj WIND SHIELD AND MICROPHONE

Satoshi Uchimura, assignor to Kabushiki Kaisha Audio-Technica
24 February 2009 (Class 381/359); filed in Japan 2 June 2004

Shield body 20 is made of open cell foam. Microphone M fits in cavity 30. Pop filter 40, also made of open cell foam but having a different density than the foam used for body 20, can be placed in cavity 32, which is for use



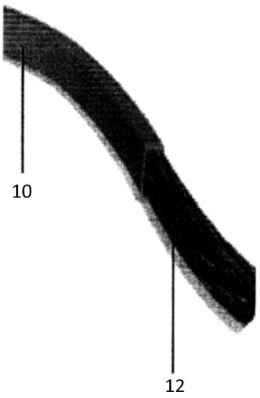
when the sound source is at 0° relative to the microphone axis. Optional ring shaped notch 21 forms an air cavity that can also reduce pop sound.—NAS

7,504,588

43.38.Tj ACOUSTICALLY TRANSPARENT STRANDED CABLE

Keith Robberding, Daly City, California
17 March 2009 (Class 174/128.1); filed 25 June 2007

A cable similar in construction to a tinsel type cable is described in prosaic terms and subjective performance claims. One such statement, in a section called "Critical Understanding and A Speculative Theory Explaining Why and How It Works," is that the finer the constituent strands, which can be as fine as 60 gauge (0.000 309 in. in diameter—which is said to be the theoretically finest wire that can be drawn), the "high frequencies are smoother, less harsh and more detailed, low frequencies are faster and pitch definition is much more apparent…" Another example is a cable comprised of 3100 strands of 56 gauge wire. How these are terminated is not clear.—NAS



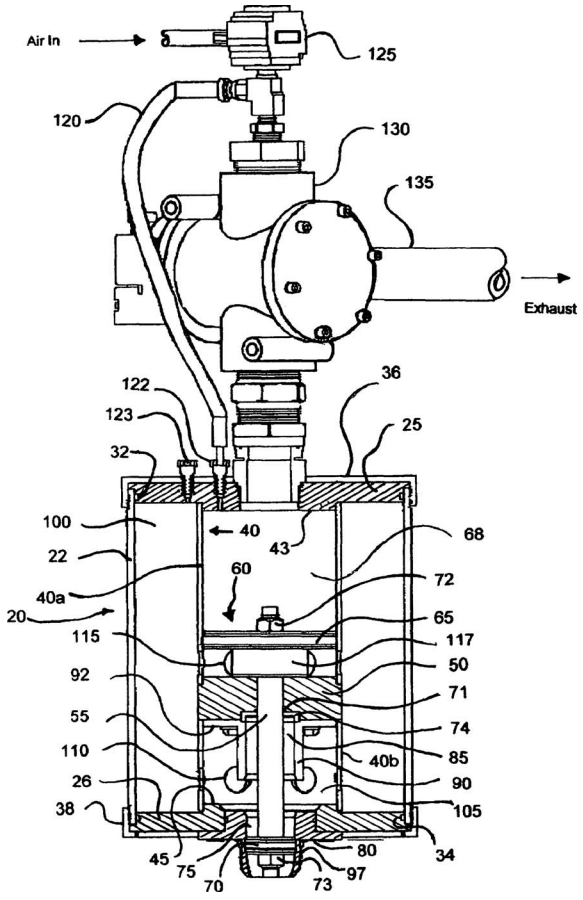
7,434,988

43.38.Yn LOW PRESSURE ACOUSTIC PYROMETER SIGNAL GENERATOR

George Kychakoff *et al.*, assignors to Enertechnix, Incorporated
14 October 2008 (Class 374/117); filed 17 May 2006

A sonic shock pulse generator for cement kiln pyrometry is claimed. Compressed air at 85–125 psi is supplied through valve 125 through hose 120 and nozzle 122 into chamber 68. Leakage at wall 40a connections will also admit compressed air into chamber 100, into chamber 117 through holes 115, and into chamber 105 through holes 110 so that eventually, all chambers of assembly 20 are charged with air at the supply pressure. Triggering of the pulse (to be emitted from orifice 97) commences by opening an exhaust valve (not shown) which decompresses chamber 68. Internal pres-

sure in cavity 117 immediately propels piston assembly 65-55-75 upward removing piston 75 from orifice 80-97 releasing a burst of air that creates an external shock wave. The released air and shock are guided by a pipe (not shown) from orifice 97 to an orifice in the cement kiln chamber wall. Presumably, the transit time of this shock across the kiln chamber becomes a measure of the average temperature of the hot kiln gasses (ergo "pyrometry").—AJC

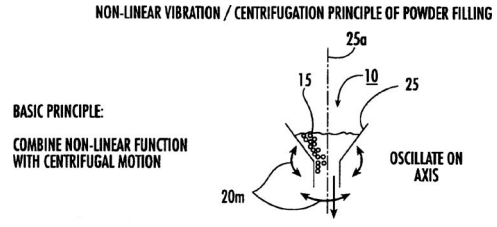


7,428,446

43.40.Ga DRY POWDER DOSE FILLING SYSTEMS AND RELATED METHODS

Timothy M. Crowder and Anthony J. Hickey, assignors to Oriel Therapeutics, Incorporated
23 September 2008 (Class 700/240); filed 12 July 2005

Several embodiments of nonlinear vibratory means for accurately dispensing small quantities of dry pharmaceutical powder 15 are claimed, one of which is shown, where rotation and oscillation 20m of the dispensing orifice is applied.—AJC

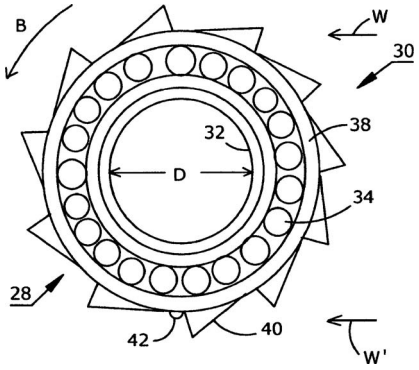


7,494,420

43.40.Le SPORTS SWING AID

James P. Whalen, Toms River, New Jersey
24 February 2009 (Class 473/228); filed 14 April 2007

To provide feedback about swing smoothness and strength, swing aid 28 is mounted using a mounting block of split conical design, near club head on the shaft of the club shaft, with *D* being large enough to accommodate the club grip width. Single weight 42 mounted to outer race 38 creates an eccentricity. When the golfer swings the club, vanes 40 cause the device to rotate. "The degree of vibration is directly associated with the swing movement, providing a direct feedback of the swing rhythm to the player." How this feedback works is left unsaid, but two embodiments are described.—NAS

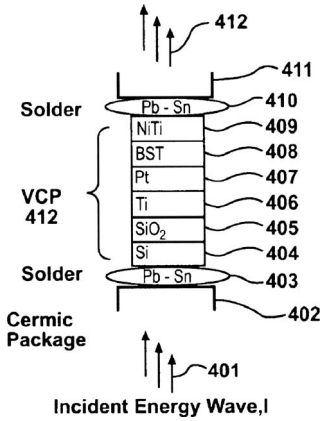


7,529,154

43.40.Sk HYBRID THIN FILM HETEROSTRUCTURE MODULAR VIBRATION CONTROL APPARATUS AND METHODS FOR FABRICATION THEREOF

Melanie W. Cole and William Nothwang, assignors to The United States of America as represented by the Secretary of the Army
5 May 2009 (Class 367/162); filed 28 September 2007

This curious patent discloses the invention of a micromachined "vibration control pedestal" whose function is to isolate from external vibrations certain vibration-sensitive microelectronic mechanical sensors, such as gyros. The pedestal shown in the figure is composed of seven different materials in eight layers. The authors claim that vibrations are reflected from the various layers as illustrated in the figure, but how this is accomplished in micron-thick layers at frequencies "between 3 and 20 kHz" is not clear. It is also claimed that the piezoelectric layer of $BaSrTiO_3$ (second from the top in the figure) is important to the operation of the device, but how this can happen is also unexplained. A very complicated pedestal, indeed.—JAH

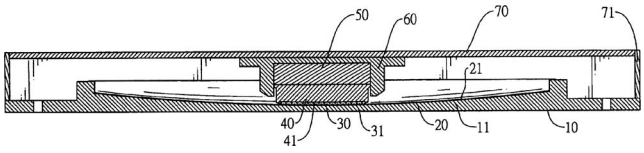


7,530,555

43.40.Tm DAMPING ASSEMBLY FOR A CONSTRUCTION

Kuo-Jung Chuang, Taipei, Taiwan
12 May 2009 (Class 267/136); filed in Taiwan 18 April 2006

This patent describes a means for protecting a building or other structure from earthquake-induced vibrations. The supporting base 10, which is to be attached atop the building's foundation, has a spherically curved top face 11 on which there rests a similarly shaped stainless steel layer 20 with a slippery top surface 21. The building to be protected is constructed atop the supporting plate 70, which rests on a round plug arrangement consisting of a holder 60 that contains a rubber pad 50 and a stainless steel layer 40. A thin slippery layer 30 of polytetrafluoroethylene is shaped to conform to the spherically curved surface 11. Around the circumference there is a shield 71 to keep dirt out. Movement of the foundation in the horizontal direction causes the building to move along the spherical surface, and thus to vibrate as if it were suspended on a pendulum whose length corresponds to the radius of curvature of the spherical surface. Thus, the building in effect is mounted on an isolation system with a very low natural frequency. Vertical vibration isolation results from compression of the rubber element 50.—EEU



7,528,525

43.40.Vn DAMPING AND STABILIZATION FOR LINEAR MOTOR STAGE

James F. Smith *et al.*, assignors to Anorad Corporation
5 May 2009 (Class 310/317); filed 26 October 2007

This patent relates to suppressing the vibrations of extended structures, such as gantry beams that are used to position tools over particular work surfaces. According to this patent, this suppression is accomplished by measuring the vibrations of the structures and applying forces that oppose these vibrations by means of piezoelectric elements. The control arrangement is described only in general outline.—EEU

7,520,111

43.40.Yq STONE DETECTION METHOD AND APPARATUS FOR A HARVESTER

John G. Berger *et al.*, assignors to CNH America LLC
21 April 2009 (Class 56/10.2J); filed 24 February 2006

Harvesters pick up crops in an intake header and convey them into an array of blades where they are cut and subjected to other processing. Since stones or other hard items that may be included with the crop materials can damage the equipment, it is desired to remove these items before they reach the processing stage. This patent describes a means for detecting stones that are conveyed along an intake header together with the crop materials. The vibrations produced as rocks impact on protrusions on the header surface are detected, band-pass filtered, and compared to a preset threshold value. A trapdoor is actuated if the signal exceeds the threshold, allowing the rock to drop out of the header. In some embodiments, a second signal channel is used to assess the impacts produced by the crop materials, so that the machine's operator can adjust the threshold value for optimum performance.—EEU

7,520,162

43.40.Yq VIBRATION METER AND METHOD OF MEASURING A VISCOSITY OF A FLUID

Alfred Wenger *et al.*, assignors to Endress + Hauser Flowtec AG
21 April 2009 (Class 73/54.41); filed in Germany 27 April 2000

If a section of pipe through which a fluid of interest flows is made to vibrate flexurally in a given plane, Coriolis forces cause the pipe to exhibit a secondary out-of-plane motion, which is a measure of the fluid's mass flow rate. The fluid-conveying pipe's torsional motion in response to a given excitation provides a measure of the fluid's viscosity. This patent describes sensor and signal processing arrangements for determination of the desired data, based on the aforementioned principles.—EEU

7,523,663

43.40.Yq MICROMECHANICAL ROTATION RATE SENSOR HAVING ERROR SUPPRESSION

Rainer Willing *et al.*, assignors to Robert Bosch GmbH
28 April 2009 (Class 73/504.12); filed in Germany 22 December 2004

This sensor consists of a seismic mass that is made to oscillate along a given axis. As this axis is rotated, the Coriolis force causes the mass to vibrate also in a direction that is orthogonal to the aforementioned axis. The magnitude of this secondary vibration is sensed and provides a measure of the rotation rate. Compensation arrangements are provided to minimize the contributions to the rotation rate determination from spurious off-axis vibrations.—EEU

7,523,667

43.40.Yq DIAGNOSTICS OF IMPULSE PIPING IN AN INDUSTRIAL PROCESS

Gregory C. Brown and Mark S. Schumacher, assignors to Rosemount Incorporated
28 April 2009 (Class 73/592); filed 23 December 2003

By impulse piping here is meant a section of pipe that includes a measuring device, such as a flow meter, orifice, or pitot tube, which tends to cause turbulence and a pressure drop in the flow. These phenomena can cause flow constrictions, plugging of the line, and malfunction of the measuring device. Since disassembly and inspection of the impulse piping are difficult, this patent describes a means for checking the flow by periodically generating a "check pulse" vibration signal on one side of the measuring device, sensing the resulting vibration on the other side of the device, filtering the sensed signal, and comparing it to a threshold value.—EEU

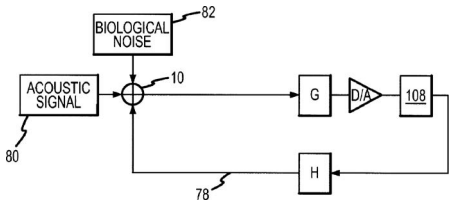
7,532,118

43.40.Yq VIBRATION SENSOR FOR BOUNDARY FENCES

Asaf Gitelis, Moshav Bnei Atarot, Israel
12 May 2009 (Class 340/564); filed in Israel 2 August 2004

Sensors on security fences need to provide reliable real-time indications of attempts to scale or cut through the fence, and they need to give years of maintenance-free service in all weathers. A typical sensor consists of a conductive ball that rests on three mutually isolated contact supports, arranged so that a sudden disturbance will displace the ball and cause a momentary interruption in the circuit. Electronic means can filter out random noise and the effects of wind and birds, and analysis of the pulses and their durations can provide an indication of the probable cause. The sensors described in this patent consist of planar arrangements of six balls, each resting on three supports, to provide greater reliability and making it more

difficult to defeat the system by wiring around it. These sensor configurations also include improvements over earlier design to reduce wear and corrosion of the contact points, to keep birds away, and to enable more convenient installation.—EEU

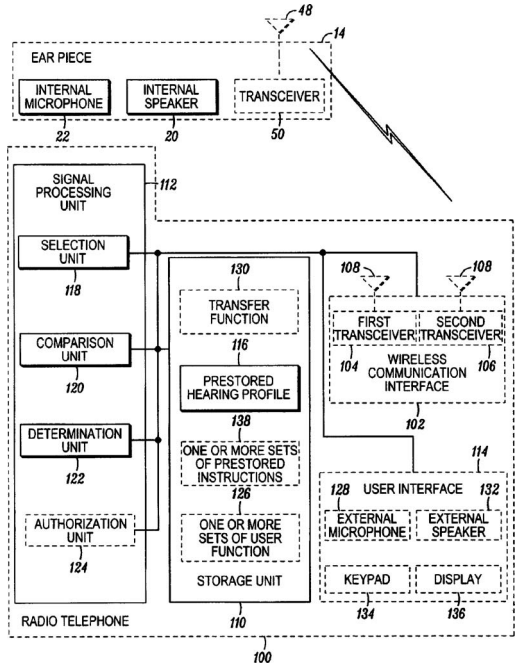


7,529,379

43.64.Ha SYSTEM AND METHOD FOR DETERMINING AN IN-EAR ACOUSTIC RESPONSE FOR CONFIRMING THE IDENTITY OF A USER

Robert A. Zurek *et al.*, assignors to Motorola, Incorporated
5 May 2009 (Class 381/328); filed 4 January 2005

Using an earpiece containing a microphone and speaker, the identity of a person is confirmed by comparing at multiple frequencies a previously stored ear canal signal for that person to the ear canal signal that is produced by the speaker. The detected signal may contain distortion product otoacoustic emission tones.—DAP



7,522,738

43.66.Ts DUAL FEEDBACK CONTROL SYSTEM FOR IMPLANTABLE HEARING INSTRUMENT

Scott Allan Miller III, assignor to Otologics, LLC
21 April 2009 (Class 381/318); filed 30 November 2006

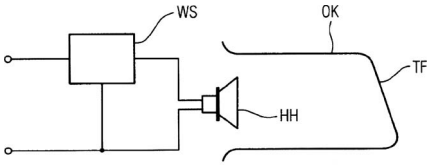
At least one motion sensor is connected to an implantable support member to differentiate between desired ambient sounds and undesirable vibratory-generated sounds such as the wearer's chewing and own voice. The axis of sensitivity of the motion sensor is aligned with a principal movement direction of the implanted microphone diaphragm. After frequency shaping and/or phase shifting, a weighted motion sensor output is subtracted in multiple bands from the microphone output. In low ambient noise environments, mechanical feedback in the microphone caused by the output transducer is reduced with a second control loop.—DAP

7,522,739

43.66.Ts HEARING AID WITH A SWITCHING DEVICE FOR SWITCHING ON AND OFF AND CORRESPONDING METHOD

Uwe Rass and Riku Sinikallio, assignors to Siemens Audiologische Technik GmbH
21 April 2009 (Class 381/323); filed in Germany 11 May 2004

Insertion and removal of the hearing aid in a wearer's ear canal are detected with changes in temperature, pressure, load resistance (representing acoustical volume), or acoustic level. High and low thresholds are used at



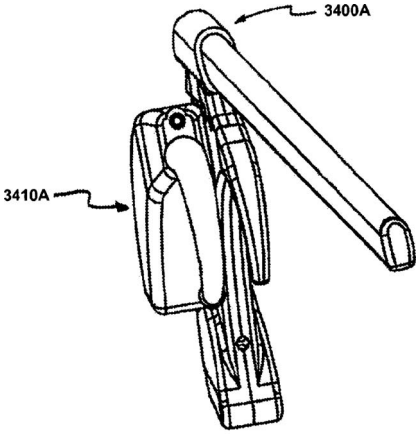
the sensor output to automatically switch the hearing aid on and off, or from operating to stand-by modes, respectively. Alternately, the hearing aid may be switched on-off by a remote control signal.—DAP

7,522,740

43.66.Ts MULTI-COIL COUPLING SYSTEM FOR HEARING AID APPLICATIONS

Stephen D. Julstrom *et al.*, assignors to Etymotic Research, Incorporated
21 April 2009 (Class 381/331); filed 1 February 2006

A headworn hearing accessory device that may have an array of directional microphones filters amplifies and converts the output signal into a magnetic field using several transmitting inductors having different orientations for improved coupling to the telecoil in a hearing aid. The magnetic



Right Ear Assembly

field orientation selected is dependent on whether coupling is to a behind-the-ear or in-the-ear hearing aid.—DAP

7,522,961

43.66.Ts INNER HAIR CELL STIMULATION MODEL FOR THE USE BY AN INTRA-COCHLEAR IMPLANT

Gene Y. Fridman and Leonid M. Litvak, assignors to Advanced Bionics, LLC
21 April 2009 (Class 607/56); filed 17 November 2004

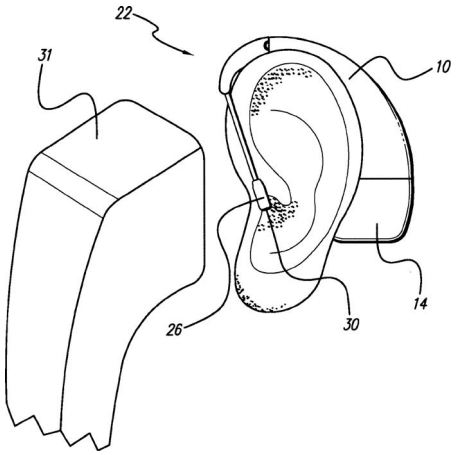
The onsets of sound in an acoustic signal associated with a particular frequency band are detected by the circuitry in a cochlear implant so they can be emphasized to improve sound clarity and speech recognition. The onset associated with the frequency band is used to determine a forcing voltage and a transmitting factor as a function of the quantity of neurotransmitter available at the time that modulates the frequency, amplitude, and decay rate of the acoustic signal to generate an output signal.—DAP

7,526,096

43.66.Ts IN THE EAR AUXILIARY MICROPHONE FOR BEHIND THE EAR HEARING PROSTHETIC

William Vanbrooks Harrison *et al.*, assignor to Advanced Bionics, LLC
28 April 2009 (Class 381/330); filed in the World Intellectual Property Organization 8 August 2002

An external microphone, attached to the earhook of a behind-the-ear sound processor for a cochlear implant, is placed in the concha. When a telephone handset is brought to the ear, the intended result is better pickup of low frequency telephone output signals.—DAP

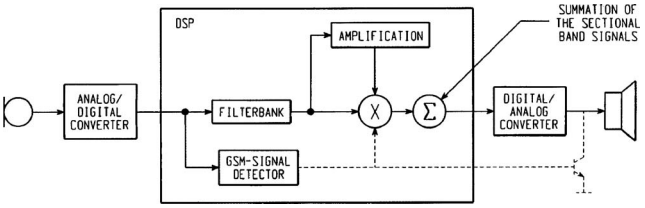


7,529,378

43.66.Ts FILTER FOR INTERFERING SIGNALS IN HEARING DEVICES

Hans-Ueli Roeck and Stefan Daniel Menzl, assignors to Phonak AG
5 May 2009 (Class 381/318); filed 12 November 2004

The repetition rate and duty cycle of well-defined interfering signals present in the hearing aid input signal, such as pulses from Global System Mobile cellular phones, are determined by the hearing aid digital circuitry. An automatically-adjustable volume control in the hearing aid suppresses



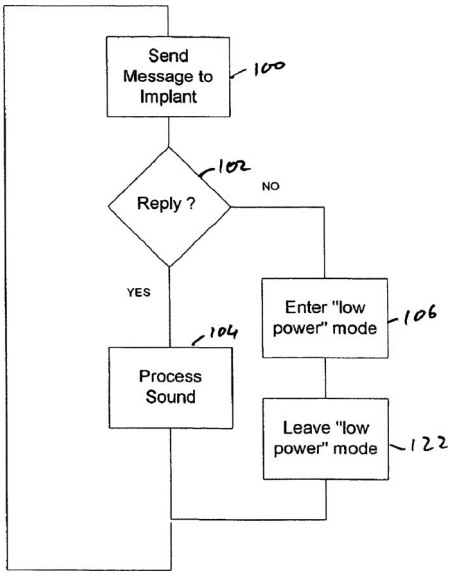
the interference in the output signal only within the limited frequency range of the interference signal spectrum.—DAP

7,529,587

43.66.Ts EXTERNAL SPEECH PROCESSOR UNIT FOR AN AUDITORY PROSTHESIS

Peter Scott Single, assignor to Cochlear Limited
5 May 2009 (Class 607/57); filed in Australia 13 October 2003

The external speech processor for a cochlear implant periodically monitors via a telemetry command whether the internal antenna coil is in proximity and in communication with the external antenna. The external



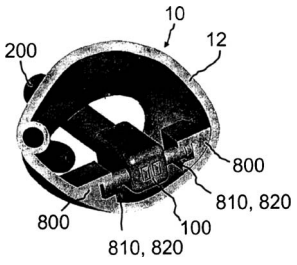
speech processor is placed into an idle state with reduced power consumption and disabled memory and transmission ability if no response or an inadequate response is received from the internal coil. The advantage is no separate power on-off switch is required in the external unit.—DAP

7,532,733

43.66.Ts FEEDBACK REDUCING RECEIVER MOUNT AND ASSEMBLY

Oleg Saltykov, assignor to Siemens Hearing Instruments, Incorporated
12 May 2009 (Class 381/324); filed 21 September 2004

A hearing aid receiver, suspended on a sound output tube, is further stabilized and protected against damage from shock by tethering it on op-



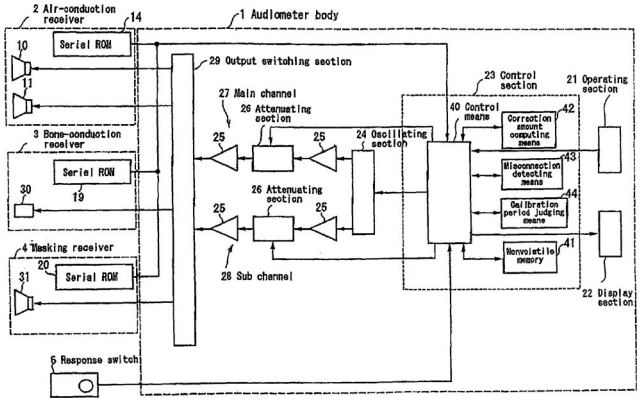
posite sides with studs on a pair of flexible mounting elements to matching receptacles in the hearing aid housing.—DAP

7,526,958

43.66.Yw AUDIOMETER RECEIVER AND AUDIOMETER

Yasuo Nojima, assignor to Rion Company, Limited
5 May 2009 (Class 73/579); filed in Japan 11 May 2006

To avoid having to perform an audiometer calibration each time an output transducer is replaced, air-conduction, bone-conduction, and masking receiver calibration and identification information are stored within each receiver in serial ROM memory. A misconnection detector determines whether a receiver has been improperly plugged into the audiometer.—DAP

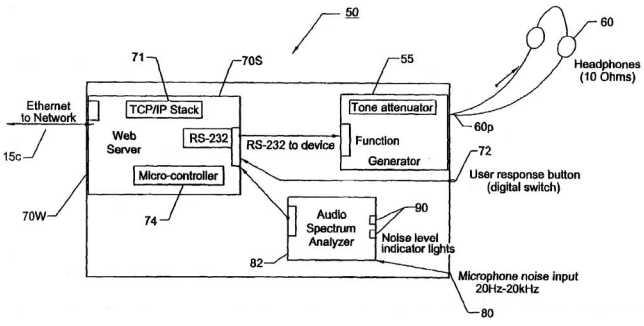


7,530,957

43.66.Yw SYSTEMS, METHODS AND PRODUCTS FOR DIAGNOSTIC HEARING ASSESSMENTS DISTRIBUTED VIA THE USE OF A COMPUTER NETWORK

Gregg D. Givens *et al.*, assignors to East Carolina University
12 May 2009 (Class 600/559); filed 25 April 2005

Under the control of a clinician at a remote site, diagnostic middle ear, compliance, and distortion product emission hearing tests are conducted simultaneously on multiple patients at separate physical sites via a computer network.—DAP



7,521,622

43.72.Ar NOISE-RESISTANT DETECTION OF HARMONIC SEGMENTS OF AUDIO SIGNALS

Tong Zhang, assignor to Hewlett-Packard Development Company, L.P.
21 April 2009 (Class 84/609); filed 16 February 2007

The goal of the patented system is to be able to classify an audio signal as one of speech, music, noise, or silence, or any combination of these. This would be done by performing a pitch analysis followed by a harmonic content analysis on each frame of the signal. An interesting method of pitch analysis is described, consisting of a weighted combination of time-domain and fast Fourier transform-based frequency-domain autocorrelations. The classification would apparently be performed by a hand-crafted decision tree technique, described in some detail in the patent text. Interestingly, the claims merely state the general goals and approach, but none of the details of how the task would actually be done.—DLR

7,526,361

43.72.Ar ROBOTICS VISUAL AND AUDITORY SYSTEM

Kazuhiro Nakadai *et al.*, assignors to Honda Motor Company, Limited
28 April 2009 (Class 700/245); filed in Japan 1 March 2002

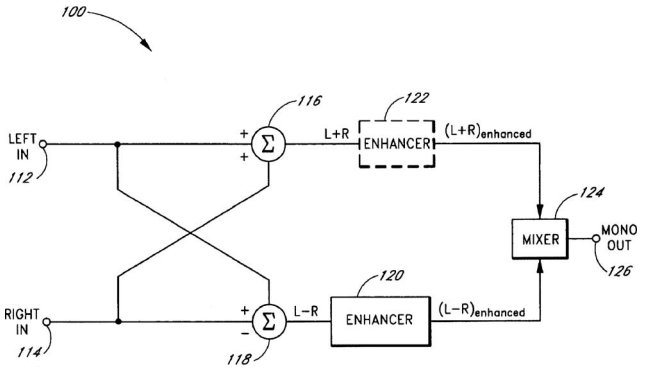
The patent describes a software system suitable for a robot to be able to identify the direction of signal arrival based on both auditory and visual sensory information. In fact, the patent text is devoted almost exclusively to the auditory task, with very little being said about visual analysis. The binaural auditory system would compute interaural differences in both time and spectral density. Reference to internal models of the head-related transfer functions for both ears, along with head-turn information from neck-control motors, would provide sound source location in a body-based coordinate system. A provision is included for an inverse transform to generate a single, clean version of the original sound.—DLR

7,522,733

43.72.Gy SYSTEMS AND METHODS OF SPATIAL IMAGE ENHANCEMENT OF A SOUND SOURCE

Alan Kraemer and Richard J. Oliver, assignors to SRS Labs, Incorporated
21 April 2009 (Class 381/1); filed 12 December 2003

The motivation is to reduce cancellation of out-of-phase summed stereo sound signals to enhance their spatial characteristics. A portion of left and right stereo audio input signals are combined to isolate sum and difference information between different stereo channels. Taking into account speaker characteristics, the difference information is enhanced with appropriate amplitude and phase changes and combined with enhanced sum information to produce an enhanced monophonic output with preserved audio information that would otherwise be canceled.—DAP



7,523,038

43.72.Ne VOICE CONTROLLED SYSTEM AND METHOD

Arie Ariav, 79304 Doar-Na Hof Ashkelon, Israel
21 April 2009 (Class 704/275); filed 24 July 2003

The system described in this patent barely qualifies as a speech recog-

nizer, as suggested by the assigned PACS code. It is a bare-bones speech analyzer intended to be able to distinguish among the inputs "yes," "no," or "stop," with a bare minimum of electronic hardware. The device would consist of a microphone and high-pass and low-pass filters, followed by a microprocessor connected so as to be able to read the averaged output of each of the filters and compare those outputs to predetermined threshold levels. If both filters produce an output, then if the high output came first, the word was stop. If low came first, it was yes. And if only the low-pass filter produced a signal, the word was no.—DLR

7,526,073

43.72.Ne IVR TO SMS TEXT MESSENGER

John Patrick Romeo, assignor to AT&T Intellectual Property L.L.P.
28 April 2009 (Class 379/88.18); filed 1 July 2005

Instead of typing destination and a text message on a wireless phone keypad using a short message service (SMS), users select the interactive voice response system (IVR) option, which provides prompts for speaking both the desired telephone number (or name) and the message. The IVR translates the message to text and speaks it back using a text-to-speech converter for user confirmation. After any necessary corrections are made, the message is forwarded to the SMS messaging gateway. The method may also be adapted for use on public switched telephone networks and wireline networks.—DAP

7,529,665

43.72.Ne TWO STAGE UTTERANCE VERIFICATION DEVICE AND METHOD THEREOF IN SPEECH RECOGNITION SYSTEM

Sanghun Kim and YoungJik Lee, assignors to Electronics and Telecommunications Research Institute
5 May 2009 (Class 704/236); filed in Republic of Korea 21 December 2004

The intent of this speech recognition procedure is to improve on the results of a typical but unspecified speech recognition system. Following a normal recognition that passes by the typical system, a support vector machine analysis is done based on the log likelihood score, the *n*-best score, and the word duration. If this results in a suitably favorable outcome, then a table of information is collected reflecting the results of the recognition process. This information would include the dialect, gender, and age of the speaker and, for each word, the recognized phonetic sequence, the number of syllables, the number of phonemes, the speaking rate, average pitch, word duration, and several other such items. This information is used to execute a decision tree trained using the classification and regression tree procedure. If this results in a satisfactory outcome, the recognition result is deemed correct.—DLR

7,533,020

43.72.Ne METHOD AND APPARATUS FOR PERFORMING RELATIONAL SPEECH RECOGNITION

James F. Arnold *et al.*, assignors to Nuance Communications, Incorporated
12 May 2009 (Class 704/257); filed 23 February 2005

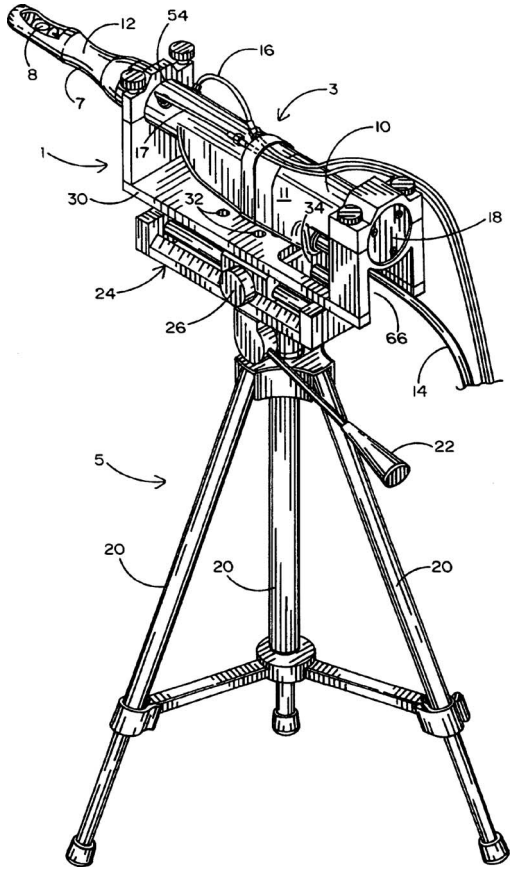
Recognition of a speech utterance involving street address information can be a very difficult task, especially when the speaker is at the wheel of an automobile on the road. In order to improve the success rate, this recognition system uses semantic relationships among the recognized words and numbers to help narrow the search space. For example, if a zip code is recognized, even partially, it can greatly help narrow the range of street names and/or city district names which may also be present in the utterance. This recognizer appears to implement such semantic constraints by guiding the construction in real time of the language model, in the form of the usual context free word sequence grammars. Some details of this process are disclosed in the patent text.—DLR

7,524,288

43.80.Vj HOLDER FOR A HIGH INTENSITY FOCUSED ULTRASOUND PROBE

Douglas O. Chinn, Arcadia, California
28 April 2009 (Class 600/439); filed 22 February 2005

This holder consists of a base that can be attached to a conventional tripod and an apparatus into which a high-intensity focused ultrasound probe is mounted.—RCW



7,524,289

43.80.Vj RESOLUTION OPTICAL AND ULTRASOUND DEVICES FOR IMAGING AND TREATMENT OF BODY LUMENS

Jay A. Lenker, Laguna Beach, California
28 April 2009 (Class 600/466); filed 19 May 2003

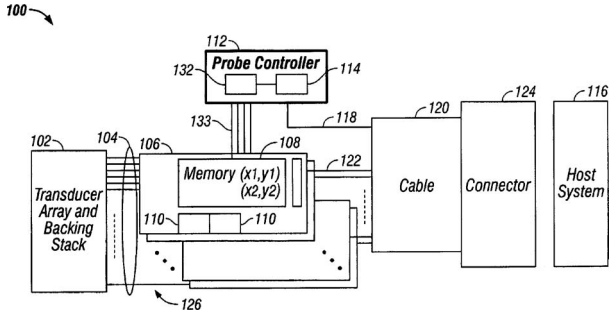
An ultrasound transducer or an optical sensor array is mounted on the tip of a catheter. The tip can be rotated around the axis of the catheter by an angle of up to 360°. The imaging array can also be rocked back and forth in a plane that includes the longitudinal axis of the catheter.—RCW

7,527,591

43.80.Vj ULTRASOUND PROBE DISTRIBUTED BEAMFORMER

Geir Ultveit Haugen *et al.*, assignors to General Electric Company
5 May 2009 (Class 600/447); filed 21 November 2003

Signal processing, memory, and control circuits are included in the ultrasound probe to perform beamforming on received signals.—RCW



7,527,592

43.80.Vj ULTRASOUND PROBE SUB-APERTURE PROCESSING

Geir Ultveit Haugen *et al.*, assignors to General Electric Company
5 May 2009 (Class 600/447); filed 21 November 2003

A signal processor that includes memory and a controller is used in the head of an ultrasound probe to process received signals in subapertures of a transducer array in the probe.—RCW

7,529,393

43.80.Vj GUIDANCE OF INVASIVE MEDICAL DEVICES BY WIDE VIEW THREE DIMENSIONAL ULTRASONIC IMAGING

Michael Peszynski *et al.*, assignors to Koninklijke Philips Electronics, N.V.
5 May 2009 (Class 382/128); filed 4 March 2004

An ultrasound visualization of a volume is used to show a medical device inserted in the body. The field of view may be shown in a cross sectional or in an enlarged format. A quantified display such as a wire-frame model derived from the three-dimensional ultrasonic image data can also be displayed.—RCW

7,530,951

43.80.Vj METHOD FOR GENERATING A THREE-DIMENSIONAL ULTRASOUND IMAGE

Jens Fehre and Bernd Granz, assignors to Siemens Aktiengesellschaft
12 May 2009 (Class 600/459); filed in Germany 11 February 2003

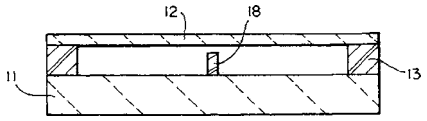
Temporally successive b-scan images in a stationary image plane containing structures that breathing causes to move are combined by a processor into a three-dimensional image.—RCW

7,530,952

43.80.Vj CAPACITIVE ULTRASONIC TRANSDUCERS WITH ISOLATION POSTS

Yongli Huang and Butrus T. Khuri-Yakub, assignors to The Board of Trustees of the Leland Stanford Junior University
12 May 2009 (Class 600/459); filed 1 April 2004

A membrane that defines the cavity of a capacitive ultrasound transducer is supported by insulating walls with a patterned isolation layer that has posts in the cavity of the transducer to prevent the transducer electrodes from coming into contact during operation of the transducer and to minimize the accumulation of charge compared to an isolation layer without a pattern.—RCW



Session 1aAAa

Architectural Acoustics and Engineering Acoustics: Spaces for Experimental Media and Spatial Sound Reproduction

Molly K. Norris, Chair
Threshold Acoustics, LLC, 53 W. Jackson Blvd., Suite 1734, Chicago, IL 60604

Chair's Introduction—8:25

Invited Papers

8:30

1aAAa1. The Experimental Media and Performing Arts Center (EMPAC) at Rensselaer Polytechnic: Defining, building, and using highest-quality spaces for hearing, seeing, and moving in space with integrated multi-modal media technology. Johannes Goebel (Experimental Media and Performing Arts Ctr. EMPAC, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180,jeg@rpi.edu)

This paper discusses studios and performance spaces built for multi-modal production, presentation, and research without compromises. Hearing, seeing, and moving in space were treated as equal in the definition of the physical properties of all spaces. Special diffusive acoustics were developed for a concert hall, a theater, and studios to be able to support any instrumental, vocal, or electronic sound from anywhere. The noise floor of the theater is as low as in the concert hall as in a studio used for video (NR 15). Only sine-wave dimmers are allowed. The integrated digital technology allows creating, recording, and projecting sound, images, and movement through thousands of audio channels, hundreds of high definition video channels, and with computer controlled rigging and flying. The world of the human senses is bridged with the digital world of technology. All noisy equipment is banned from any venue. All venues are acoustically extremely separated with one studio resting on springs, the other studio having a separate foundation, and acoustical joints everywhere. The team comprises artistic curators as well as engineers and researchers; artists and scholars are in residence; the task is to span arts, science, and technology. The center opened 2008.

9:15

1aAAa2. Small room production before large room performance. Alex U. Case (Sound Recording Technol., Univ. of Massachusetts, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

Large scale works almost always begin their acoustic life in a small room. Sound designers and composers need studio spaces that enable creativity while keeping technical matters in check. The stereo recording studio requires a paradigm shift when scaling up to four, eight, or more channels. While fundamental issues of isolation, noise control, user comfort, and ergonomics remain ever-important, new challenges arise. Despite the growth in electroacoustic complexity that accompanies multichannel systems, the sound/music creator needs core issues of level, timbre, localization, distance, envelopment, space, and so on to be presented as faithfully as possible, predictive of the future large space performance, without corruption by the small room.

9:45

1aAAa3. Utilizing reciprocal maximum length sequences within a multichannel context to generate a natural, spatial sounding reverberation. Uday S. Trivedi and Ning Xiang (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY, 12180, triveu@rpi.edu)

The development of artificial reverberation has made considerable progress in recent years. Simultaneous advances in knowledge and computational power have allowed for a rapid development and application in the design of virtual room environments. Within such situations, simulating room acoustics is critical for producing a convincing immersive experience. This research explores the application of a multi-channel loudspeaker setup that can be employed in simulating such an environment. An algorithm to shape and vary the spaciousness within the context of an artificially generated reverberation is implemented. The pseudo-random properties of reciprocal maximum-length sequences (R-MLSs) allow for a deterministic decorrelation between all channels. The use of R-MLS also provides several advantages over using traditional random noise. This paper will discuss the potential applications in creating virtual spaces and variable acoustic environments using multiple channel systems.

Contributed Paper

10:15

1aAAa4. Binaural reproduction of spherical microphone array signals. Joshua D. Atkins (Dept. of Elec. Eng., Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218, joshatkins@jhu.edu)

An efficient method for the production of binaural audio signals from spherical microphone array signals using head related impulse responses (HRIRs) is presented. The processing is done directly in the spherical Fourier transform domain which offers significant speed advantages over the conventional method of beamforming toward each HRIR measurement point. The encoding of HRIRs into the spherical Fourier domain is complicated by the fact that the available HRIR databases (CIPIC, MIT, KEMAR) use irregular sampling positions around the sphere and do not contain data for sound coming from angles below the subject. However, for spatially bandlimited processing it is possible to use an LMS technique with regularization to perform the spherical Fourier transform of these HRIRs. A real-time system with head-tracking using the eigenmike (TM) and the CIPIC HRIR database is presented which processes spherical harmonics up to third order. Since the head-rotation operation is performed digitally, this technique is especially useful for remote surveillance and virtual reality systems with complex acoustic scenes.

MONDAY MORNING, 26 OCTOBER 2009 REGENCY EAST 3, 10:40 A.M. TO 12:00 NOON

Session 1aAAb

Architectural Acoustics: Biggest Mistakes: Lessons for Practitioners, Researchers, and Young Designers

Ian Hoffman, Chair
School of Art, Design, and Architecture, Judson Univ., 1151 State St., Elgin, IL 60123

Chair's Introduction—10:40

Invited Papers

10:45

1aAAb1. Mixing studios complications. K. Anthony Hoover (McKay Conant Hoover, Inc., 5655 Lindero Canyon Rd., Westlake Village, CA 91362)

A new suite of mixing and teaching studios had been discussed for many years at Berklee College of Music in Boston. The authorization to proceed with design and construction of the studios finally came with a new Dean of the Music Technologies Department, who, without benefit of much of the background and previous discussions, contacted an old colleague for the final studio design and contracted with a reputable full-service contractor for construction services. Problems developed during construction, which apparently sensitized the client to subsequent situations, in turn requiring acoustical consulting services. This paper will review some of the project history, outline the problems, and discuss solutions and recommendations, and will address some of the concerns with the specialized construction and unusual complications.

11:05

1aAAb2. I did not see that one coming. Richard Talaske, Gregory Miller, Byron Harrison, and Evelyn Way (Talaske, 1033 South Blvd., Oak Park, IL 60302, info@talaske.com)

Occasionally we encounter something we have not before, to our detriment. Only by looking back do we gain the proper perspective to know how to avoid similar concerns in the future. We will share several examples from the Talaske body of work where things turned out differently that we had anticipated. Examples will include a theatre wall system which provided more diffusion and absorption than laboratory tests had indicated, a contractor construction error that had negative acoustic implications which were not caught until it was too late to correct, and an acoustic isolation concern between a dance studio and a theatre that passed the boundaries of our analytical assumptions. In each case, the analytical, practical, and political hurdles required to resolve each issue will be presented.

11:25

1aAAb3. Key planning steps for performance based designs. Joshua Cushner (Arup Acoust., 115 Ave. of the Americas, New York, NY 10013)

In architecture and engineering, leading firms will often value the opportunity to collaborate with unconventional clients to create bespoke building designs. However, such projects also require greater early planning and create new risks for the project team to manage. When generating "customized" designs with a specific tenant in mind rather than simply relying on industry standards for similar facilities, the rigors of a performance-based design must be understood. This case study reviews a project where key planning steps for this type of design approach were overlooked, which created conflicting project goals that were never fully resolved. While

efforts were made by team members over time to improve collaboration, the process-related problems which manifested early in the project persisted, and the design team was never able to establish cohesion amongst themselves or with the owner. This case study details the main elements which predicated the project failures and suggests alternate approaches which could have better served the goals of the project. As a more successful framework, a broadened use of the performance-based design model is reviewed.

Contributed Paper

11:45

1aAAb4. Biggest mistakes encountered on projects without acoustical consultants. Benjamin Markham, Alicia Wagner, Jeffrey Fullerton, and Jennifer Hinckley (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

As consultants, the authors have been called into dozens of projects built without the benefit of guidance from an acoustician, and the results are often disastrous (and in many cases, litigious). This presentation, replete with photographic evidence, outlines some of the biggest mistakes encountered recently by design teams that lack a critical team player: the qualified acoustical consultant. Consider a natatorium with all hard surfaces and a 6-s reverberation time to match, a fieldhouse gymnasium with a 70-dB background noise level, and a distressingly common problem in New England: residential condominiums converted from an old mill building with nothing between stacked residences except exposed, unimproved hardwood decking. Other examples relate to inadequate (or altogether absent) vibration isolation of elevator machinery and other mechanical equipment, cacophonous restaurant acoustics, blatant disregard for a nearby railway, and poor music practice facilities.

MONDAY MORNING, 26 OCTOBER 2009 RIO GRANDE EAST, 8:00 TO 11:45 A.M.

Session 1aAO

Acoustical Oceanography and Underwater Acoustics: Session in Honor of Stanley Flatté I

John A. Colosi, Cochair
Naval Postgraduate School, Dept. of Oceanography, 1833 Dyer Rd., Monterey, CA 93943

Jeffrey A. Simmen, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105

Chair's Introduction—8:00

Invited Papers

8:05

1aAO1. Sound transmission through a fluctuating ocean (1979) and ocean acoustic tomography (1995): An intertwined history. Peter F. Worcester and Walter H. Munk (Scripps Inst. of Oceanogr., Univ. of California at San Diego, La Jolla, CA 92093-0225, pworcester@ucsd.edu)

Early attempts at understanding scintillations in sound transmission through the ocean assumed that ocean fine structure is homogeneous and isotropic. It is, in fact, dominated by internal waves, which are neither homogeneous nor isotropic. *Sound Transmission Through a Fluctuating Ocean* (Flatté *et al.*, Cambridge University Press, Cambridge, England 1979) combined recently developed internal-wave models with path-integral methods to predict the fluctuations of resolved acoustic multipaths. At that time, *Ocean Acoustic Tomography* (Munk *et al.*, Cambridge University Press, Cambridge, England 1995), which uses ray travel times to determine large-scale ocean structure, had just been proposed. The wideband acoustic sources and receivers required to make travel-time measurements for tomography also provided the technology needed to quantify acoustic fluctuations. Conversely, internal-wave-induced scattering limits travel-time measurement precision. Tomography experiments at 25-km range provided some of the earliest tests of path-integral predictions. Subsequent tomography experiments provided data at ever-increasing ranges and decreasing frequencies. The Acoustic Thermometry of Ocean Climate project made measurements at megameter ranges with 1400-m vertical line array receivers. The next step is a modular distributed VLA capable of spanning the full water column that is under development to enable separation of acoustic modes using spatial filtering and to fully characterize deep-water acoustic time fronts.

8:25

1aAO2. Center for the studies of nonlinear dynamics, 1982–1985. Frank S. Henyey (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, frank@apl.washington.edu)

Stan Flatte was the third director of the Center for the Studies of Nonlinear Dynamics (CSND), from 1982 to 1985 (and for another year after he returned to Santa Cruz). Upon joining CSND, he assembled a group to work on the path integral method for predicting internal wave effects on acoustic propagation in the ocean. This talk discusses the work done by that group. Topics of this work include travel time bias, relation to moment equation methods, fourth moment calculations, and pulse spread.

8:45

1aAO3. Scaling turbulent dissipation and finestructure: The wave-propagation method. Timothy F. Duda (AOPE Dept., MS 11, Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

In the 1980s the team of Flatté *et al.* developed a theory relating the turbulent kinetic energy dissipation rate in the open ocean to finestructure feature observations (1–100 m scale), building on work of Henyey and Pomphrey. Formulas derived from this work are now prevalently used for indirectly estimating cross-isopycnal ocean mixing, a process of critical interest, using platforms and sensors unable to directly measure viscous-scale dissipation and associated turbulent transport processes. The method originally involved the internal-wave energy level. Revisions of the method introduced by follow-on investigators involve internal-wave shear and strain. The physics behind the method is that short-wavelength shear-rich internal waves propagate in a background of larger-scale internal waves, with action conserved, and can have their energy concentrated spatially. The concentrated waves are inferred to break, and in doing so provide the energy to support diapycnal mixing. The approach is one of "wave propagation in random media," a specialty of Flatté.

9:05

1aAO4. Ray methods in long-range deep ocean sound propagation. Michael G. Brown (RSMAS, Univ. of Miami, 4600 Rickenbacker Cswy., Miami FL 33149, mbrown@rsmas.miami.edu), Ilya A. Udovydchenkov, and Irina I. Rypina (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Although ray methods have limitations, they remain extremely important because they provide insight into the underlying wave propagation physics that is difficult, if not impossible, to obtain by any other means. In this talk the utility of ray methods in long-range deep ocean sound propagation will be discussed and illustrated. Topics to be discussed include ray-mode duality and action quantization; adiabatic invariance for rays and modes; beamforming and Radon transforms; caustics and catastrophes; travel time sensitivity kernels, Fresnel zones and other measures of ray widths; nonlinear resonances, KAM theory, and mode coupling; the waveguide invariant and its ray equivalent; and beam dynamics—from weakly divergent to explosive. [Work supported by ONR.]

9:25

1aAO5. Flatte fluctuation theory and its use by the US Navy. Arthur B. Baggeroer (Massachusetts Inst. of Technol., Rm. 5-204, Cambridge, MA 02139)

Fluctuations of narrowband signals and spreading of broadband ones are important concepts in the design of the sonar receiver. The first determines how long one can integrate or equivalently how narrow a bandwidth can be used to increase the gain and the second leads to the convolution of the ambiguity function and the scattering function to determine where energy from the transmission will appear in the range-Doppler plane, hence the region over which one wants to collect energy for a detection. The seminal book "Sound Transmission through a Fluctuating Ocean" by Flatte *et al.* essentially defined the concepts of saturated, partially saturated, and unsaturation transmissions which parsed the problem in the "lambda-phi" for designing sonar transmitters and receiver appropriate for the dynamics of an oceanographic region.

Contributed Papers

9:45

1aAO6. Acoustic signal horizontal coherence variability: Relationship to internal tide and storm events. Marshall H. Orr (Acoust. Div., The Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5032, marshall.orr@nrl.navy.mil), Peter C. Mignerey, and David Walsh (Naval Res. Lab., Stennis Space Ctr., MS 39529-5004)

Acoustic signal horizontal coherence has been extracted from a 31-day data set taken during the late fall/early winter time period. The coherence variability for both 300- and 500-Hz acoustic signals appears coupled to shelf/slope frontal motion, internal tides, and associate internal waves as well as changes in the surface gravity field. Acoustic signal and environmental data illustrating the coupling will be presented. The data were taken on the New Jersey Shelf with a 460-m-long bottomed array placed in ~89 m of water. The cross shelf propagation range was ~20 km. The acoustic sources were in ~60 m of water.

10:00—10:15 Break

10:15

1aAO7. Cross-mode coherences and decoupling of equations for mode intensities in two-dimensional and three-dimensional fluctuating ocean. Alexander G. Voronovich, Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80303), John A. Colosi (Naval Postgrad. School, Monterey, CA 93943), and Andrey K. Morozov (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Sound waves propagating in the ocean are scattered by internal waves (IWs) and spice. A modal approach for studies of low-frequency, long-range sound propagation in a fluctuating ocean seems to be the most adequate. This approach has been employed in a number of works, including recently published papers [A. G. Voronovich and V. E. Ostashev, J. Acoust. Soc. Am. **125**, 99–110 (2009)] and [J. A. Colosi and A. K. Morozov, J. Acoust. Soc. Am. **124**, 2598 (2008)]. Based on these two papers, in this presentation important aspects of sound scattering in a fluctuating ocean are considered for both two-dimensional (2-D) and three-dimensional (3-D) geometries. First, we compare equations for the coherence function of a sound field in 2-D and 3-D fluctuating ocean. Then, decoupling of equations for the mode intensities from equations for the cross-mode coherences is studied for the two geometries. Numerical examples of evolution of the mode intensities and cross-mode coherences with range are presented and discussed. Finally, insonification of an acoustic shadow zone by sound waves scattered by IWs is studied. [Work supported by ONR.]

10:30

1aAO8. Antipodal acoustic propagation and a half-century of ocean warming. Brian D. Dushaw (Appl. Phys. Lab., Univ. of Washington, 1013 N.E. 40th St., Seattle, WA 98105, dushaw@apl.washington.edu)

In 1960, sound signals traveling from Perth, Australia were recorded at Bermuda. Previous work focused on the path traveled by the sound [Munk *et al.*, JPO, 1876–1898 (1988)]. Calculation of the horizontal refraction of sound, across the Southern Ocean in particular, gave the perplexing result that Bermuda was in the shadow of Africa. Heaney *et al.* [J. Acoust. Soc. Am., 2586–2594 (1991)] used low-resolution atlases for global sound speed and bathymetry to obtain two viable acoustic paths between Perth and Bermuda, both influenced by bathymetry. From a modern perspective, however, the explanation of Heaney *et al.* is unconvincing [Dushaw, GRL (2008)]. High-resolution ocean models put the Perth-to-Bermuda acoustic problem into a new light. These models suggest that intense, small-scale features,

e.g., Agulhas rings near the Cape of Good Hope, would greatly influence the acoustic paths. The antipodal travel time, 13 382 s, is a measure of the ocean temperature in 1960. If the acoustic propagation issues can be fully understood, data-assimilating ocean general circulation models might be used to calculate a present-day travel time. The travel-time change over the past half-century, expected to be about 10 s based on nominal estimates of ocean warming, is a measure of ocean climate change.

10:45

1aAO9. Characterization of deep acoustic shadow zone arrivals. Lora J. Van Uffelen and Peter F. Worcester (Scripps Inst. of Oceanogr., Univ. of California at San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, lvanuffe@ucsd.edu)

Acoustic shadow-zone arrivals first observed in the late 1990s on horizontal receiving arrays in the North Pacific Ocean revealed significant acoustic energy penetrating the geometric shadow by an estimated 500–1000 m. An extensive vertical line array deployed in conjunction with 250-Hz acoustic sources at ranges of 500- and 1000-km during SPICEX, a long-range propagation experiment conducted from June to November 2004 in the North Pacific, confirmed the presence of these anomalously deep arrivals and enabled an examination of their vertical structure. Parabolic equation simulations incorporating scattering consistent with the Garrett–Munk internal-wave spectrum at full strength are able to describe both the energy contained in and vertical extent of deep shadow-zone arrivals. Several hundred acoustic receptions were recorded during the experimental deployment, enabling a statistical characterization of the energy and vertical extent of these arrivals as well as an investigation of the evolution of the arrivals throughout the nearly 6 months of acoustic receptions.

11:00

1aAO10. Seismo-acoustic modal scattering by volume heterogeneities in shallow water sediments. Darin J. Soukup and Robert I. Odom (Appl. Phys. Lab., Univ. of WA., 1013 NE 40th St., Seattle, WA 98105, odom@apl.washington.edu)

Elastic anisotropy is a nearly ubiquitous feature of marine sediments. The simplest type of sediment anisotropy is transverse isotropy, characterized by five elastic constants, and results from layered deposition. A modal scattering theory for volume perturbations of the sediment elastic moduli is presented. The scattering theory is based on the coupled mode formulation for propagation in range dependent fluid-elastic media. The Born approximation is employed to derive a modal scattering matrix. Although the perturbations of the elastic moduli are random, they may not be arbitrary in the sense that certain symmetry and energy constraints among the moduli must be respected. Mode-mode coupling matrices are computed for quasi-P-SV-SH seismo-acoustic modes, which show mode mixing and the importance of non-nearest neighbor interactions. The effects of volume scattering can be combined with rough surface scattering and also incorporated into mode coupling caused by deterministic range dependence of the material properties. This work has implications for acoustic loss estimates for low-frequency shallow water acoustic propagation. [Work supported by ONR.]

11:15

1aAO11. Deep seafloor arrivals: Scattering or multi-path from ocean thermal structure? Ralph A. Stephen (WHOI, 360 Woods Hole Rd., Woods Hole, MA 02543, rstephen@whoi.edu), Matthew A. Dzieciuch, Peter F. Worcester (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0225), Rex K. Andrew, James A. Mercer (Univ. of Washington, Seattle, WA 98105-6698.), John A. Colosi (Naval Postgrad. School, Monterey, CA 93943), and Bruce M. Howe (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

An unexplained set of arrivals has been observed on ocean bottom seismometers (OBSs) during the NPAL04 long-range ocean acoustic propagation experiment in the North Pacific. The observed intensity pattern of the OBS arrivals is significantly more complex than the waterborne arrivals seen on the deep vertical line array (DVLA). These "deep seafloor" arrivals occur later than the first PE predicted arrival; their arrival time is not predicted by acoustic PE propagation models, they do not correspond to decay from shallower turning points (as is the case for deep shadow zone arrivals), and they are not readily observed on the DVLA hydrophone just 750 m above the seafloor. The arrival structure in the observed data, in time and amplitude, varies substantially between three OBSs that are separated by less than 4 km. Could these unexplained arrivals be scattering or horizontal multi-path from persistent ocean thermal structure?

11:30

1aAO12. Internal wave strength inversion based on the shape of the axial finale. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039)

Flatte had a significant impact on my dissertation and early education in the field of ocean acoustic propagation. In particular, Flatte's work on the impact of internal wave scattering on deep ocean propagation was central to my early work in adiabatic mode scattering in the presence of deep water internal waves. In this paper, we will revisit some of Flatte's wave propagation in a random media theory and show its phenomenological impact on deep water propagation, primarily within his explanation of the Slice89 experiment. In this experiment the axial finale (latest arriving energy) is broadened in depth. Flatte correctly asserted that this was due to internal wave scattering to higher modes (with larger depth extents) and used this as a method for heuristically determining the internal wave strength. Results from Heaney [Ph.D. thesis, SIO-UCSD (1997)] will be presented using the ATOC experiment. More recent results using NPAL data as well as more recent modeling will be presented. The impact of these results on deep water anti-submarine warfare and the extension of these results to basin scale propagation (>10000-km ranges) will be made.

Session 1aPAa

Physical Acoustics: Propagation, Scattering, and Related Topics

Ralph T. Muehleisen, Chair
Illinois Inst. of Technology, Civil, Environmental and Architectural Engineering, Chicago, IL 60616

Contributed Papers

8:00

1aPAa1. Comparison of three coordinate mapping methods for sound propagation over irregular terrain. Santosh Parakkal, Kenneth E. Gilbert, and Xiao Di (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, sparakka@olemiss.edu)

In propagation calculations, a coordinate mapping to "flatten" irregular terrain is an attractive approach. However, coordinate mapping methods can be as computationally intensive as the propagation calculation itself. Hence a judicious choice of the mapping method is critical for practical calculations. We compare three coordinate mapping methods: (1) conformal mapping; (2) transformation to polar coordinates; and (3) a simple mapping introduced 30 years ago by Beilis and Tappert . The first two methods are essentially exact, while the third is, in principle, applicable only to small slopes and low propagation angles. The three approaches are outlined and a comparison is made of the accuracy and computational demands of the methods. Also, the theoretical predictions from the three methods are compared to field measurements of sound propagation over a hill. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

8:15

1aPAa2. Sound propagation in a refractive, turbulent atmosphere above a statistically rough, impedance ground surface. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80303 and Phys. Dept., NMSU, Las Cruces, NM 88003), D. Keith Wilson, and Sergey N. Vecherin (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

In recent years, considerable effort has been devoted to studies of sound propagation in a turbulent atmosphere. It was shown that temperature and wind velocity fluctuations can significantly diminish the coherence of a sound wave and, hence, degrade performance of modern acoustic sensor arrays for source detection. Outdoor sound waves are also affected by a rough, impedance surface of the ground. This paper is devoted to studies of the effects of a statistically rough, impedance ground surface on the coherence of a sound wave propagating in a refractive, turbulent atmosphere. Using the Beilis–Tappert transformation, the considered problem is reduced to sound propagation over a flat impedance surface in an atmosphere with an effective turbulence spectrum. Then, a closed equation for the coherence function of a sound field is derived using the Markov approximation similarly to that in the work of Wilson and Ostashev [J. Acoust. Soc. Am. **109**, 1909–1922 (2001)]. The previously developed moment-screen method is used to numerically solve the closed equation for the coherence function. Possible approaches are discussed for comparing the relative roles of atmospheric turbulence and surface roughness in diminishing the coherence of a sound wave.

8:30

1aPAa3. Molecular simulation of sound propagation in a gas. Takeru Yano (Dept. of Mech. Eng., Osaka Univ., Suita 565-0871, Japan, yano@mech.eng.osaka-u.ac.jp)

Molecular dynamics simulation is carried out to clarify the propagation process of high frequency and large amplitude sound in a gas. Numerically, the inter-molecular potential is assumed to be the Lennard-Jones 12-6 potential and the motions of molecules are determined by solving Newton'S equation of motion for each molecule with the leap-frog method. The gas, where the sound propagates, is sufficiently rarefied or low density so that its behavior may be governed by the Boltzmann equation. Comparison of the numerical results based on the molecular dynamics, the Boltzmann equation, and the Navier–Stokes equations, the nonequilibrium effects due to the high frequency and nonlinearity are demonstrated.

8:45

1aPAa4. Sound propagation classes for long-range assessment algorithms. Michelle E. Swearingen and Michael J. White (US Army ERDC-CERL, P.O. Box 9005, Champaign, IL 61826, michelle.e.swearingen@usace.army.mil)

Preparing noise assessments for military training activities is a significant challenge due to the short duration of the individual signals and the lack of highly detailed atmospheric conditions, due to either an absence of necessary meteorological sensors or a need to perform the assessment without prior atmospheric knowledge. To overcome these difficulties, a set of sound propagation classes has been developed. These classes narrowly define the atmospheric and ground properties and have associated mean and variance as a function of distance. This talk will provide a description of these classes and examples of how they are used.

9:00

1aPAa5. Exact image solution approach for multiple reflections. Ambika Bhatta, Miroslava Raspopvic, Max Denis, and Charles Thompson (ECE Dept., Univ. Mass. Lowell, 1 University Ave. (CACT), FA 203, Lowell, MA 01854, ambika_bhatta@student.uml.edu)

In this paper a modified method using the solution of the Laplace transform based formulation for the Sommerfeld integral to determine the strength of image sources and boundary reflections is presented. It is shown that multiple reflections can be expressed in terms of elemental branch integrals. This formulation allows for a rigorous derivation of the pressure field using numerical methods.

9:15

1aPAa6. Convergence of Born series using Padé approximants. Jing M. Tsui, Max Denis, and Charles Thompson (CACT, Umass Lowell, 1 University Ave., Lowell, MA 01854)

The solution of acoustic wave scattering can be calculated in terms of Born series in the low contrast limit. Recasting the Born series in terms of Padé approximants allows one to extend its validity at higher contrast. However, as the value of the contrast is further increased, the direct calculation of the scattering field using the Padé approximant coefficients may exhibit local divergence in the scattering volume. Origin of this problem and its possible solution will be discussed. The results will be presented.

9:30

1aPAa7. Turbulence-shear interaction pressure contributions to wind noise. Richard Raspet, Jiao Yu, and Jeremy Webster (Dept. of Phys. and Astronomy and the Natl. Ctr. for Physical Acoust., University, MS 38677)

In an earlier paper [Raspet *et al.*, J. Acoust. Soc. Am. **123**, 1260–1269 (2008)], the mean shear-turbulence interaction pressure predicted by George *et al.* [J. Fluid Mech. **148**, 155–191 (1984)] was calculated for outdoor wind noise measurements. It was found that the mean shear-turbulence interaction

pressure predictions overestimated the spectral levels at low wave numbers and were not consistent with measurements. A new calculation based on Kraichnan's method [J. Acoust. Soc. Am. **28**, 378–390 (1956)] includes two significant improvements: It uses a logarithmic wind velocity profile instead of a linear wind velocity profile, and it models the reduction in the vertical turbulence spectrum as the ground surface is approached. The effects of the change in vertical spectrum with height and the curvature of the wind profile are analyzed by comparing the two calculations. The height dependence of the predicted turbulence-shear interaction pressure spectrum is also calculated. The predictions with the improved method agree with measurements made within large (1.0 m) wind-screens. [Research supported by the U.S. Army TACOM-ARDEC at Picatinny Arsenal, NJ.]

9:45

1aPAa8. Nonuniformly spaced linear array. Jenny Y. Au (Dept. of Elec. and Comput. Eng., Ctr. for Adv. Computation and Telecommunications, Univ. of Massachusetts Lowell, 1 University Ave., Falmouth 203, Lowell, MA 01854, jenny_au@student.uml.edu) and Charles Thompson (Univ. of Mass., Lowell, MA 01854)

In this presentation, we will examine the criterion for the design of nonuniformly spaced linear arrays. The transducer location selection based on the min-max optimization will be of particular interest. The purpose is to develop low-complex array structure. The influence of transduced interaction will also be considered.

MONDAY MORNING, 26 OCTOBER 2009 REGENCY EAST 2, 10:15 TO 11:30 A.M.

Session 1aPAb

Physical Acoustics: Thermoacoustics, Cavitation, and Bubbles

E. Carr Everbach, Chair
Swarthmore College, Dept. of Engineering, Swarthmore, PA 19081

Contributed Papers

10:15

1aPAb1. Computational thermoacoustics in fibrous media. Carl Jensen and Richard Raspet (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University MS 38677, raspet@olemiss.edu)

A computational method for investigating the thermoacoustic properties of fibrous media has been developed that simulates the time-varying acoustic flow through the micro-geometry of fibrous material samples. The fibrous sample is represented in the simulation as an array of straight, cylindrical fibers, and a lattice Boltzmann based thermal fluid solver is used to calculate the response from an acoustic signal passing through the modeled geometry. It was found in development that the boundary conditions at solid surfaces must be represented to high accuracy as well as providing sufficient resolution to capture boundary layer effects near the walls and throughout the domain. These issues place limits on memory capacity that restricts the use of the method on samples of higher porosity. These issues will be discussed, and some results of the simulations will be presented.

10:30

1aPAb2. Effects of heat exchanger size on gain of thermoacoustic prime mover. Bonnie J. Andersen (Dept. of Phys., Utah Valley Univ., 800 W Univ. Pkwy., Orem, UT 84058, bonniem@uvu.edu) and Orest G. Symko (Univ. of Utah, Salt Lake City, UT 84112)

Thermoacoustic standing wave prime movers are discussed as self-sustained oscillators whose initial growth pattern of oscillation followed by amplitude limitation can be modeled by the Van der Pol equation. The best fit solutions offer quantitative comparisons of terms related to the energy supplied to the system and the losses. This study on engines near 2.6-kHz oscillations suggests that information on performance of the device is contained within the first second of oscillations. The Van der Pol equation is used to compare the performance of three hot heat exchangers. The hot heat exchangers used copper wire mesh with 40×40, 60×60, and 80×80 wires /in. The engine with the 40×40 mesh has a higher gain term and smaller loss term than either the 60×60 or the 80×80, as expected, since it has the largest wire diameter and the largest open area. The gain term for the three heat exchangers increases linearly with increasing surface area of the mesh in contact with the air in the resonator. This is evidenced in the build-up of the oscillations, in agreement with the Van der Pol equation.

10:45

1aPAb3. Generation of cavitation in mercury by means of an electromagnetic acoustic source. Qi Wang, Nicholas J. Manzi, Glynn R. Holt, Ronald A. Roy, and Robin O. Cleveland (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, qiwang@bu.edu)

An electromagnetic acoustic source (EMAS) consists of a flat electrical coil, a thin insulating membrane, and a metal plate. The EMAS is excited by discharging a capacitor through the coil, and the resulting eddy currents produce a repulsive force on the metal plate launching acoustic waves. Here an EMAS was used to excite acoustic waves in a cylindrical chamber of mercury (75-mm ID and 75-mm long) with the metal plate bonded to the bottom of the tank. The lid of the chamber had a re-entrant cylinder such that a steel plate was positioned 50 mm from the EMAS face. A laser Doppler vibrometer was used to monitor the deflection of the plate to ascertain the acoustic field. A 75-mm-diameter coil was wound with either 18, 27, or 36 turns. The EMAS plate was steel, aluminum, or just a Mylar insulating membrane with no metallic plate. Experiments were carried out for capacitances of 0.5, 2, and 4 μF, charging voltage of 0–10 kV, number of coils, and properties of the plate. For a 4-μF capacitor charged to 10 kV, cavitation was produced in the mercury with duration of 150 μs.[Supported by the ORNL Spallation Neutron Source US-DOE Contract DE-AC05-00OR22725]

11:00

1aPAb4. Boundary element simulation of fully three-dimensional bubble dynamics. Jason P. Kurtz and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

A three-dimensional boundary element method is presented for simulating bubble dynamics in incompressible, inviscid fluids (such that the fluid velocity is the gradient of a potential). The method is applied to simulate the migration and collapse of bubbles near rigid boundaries. The motivation is to model multiple cavitation bubble interactions near the hard surfaces of kidney stones that occur in shock-wave lithotripsy. In order to accurately compute the fluid velocity, the method employs a novel curvilinear representation of the bubble geometry that enforces the global continuity of the unit normal vector. Contrary to the axisymmetric setting, the global continuity of the derivative of the parametrization, known as C^1 continuity, cannot be enforced. Enforcing the continuity of the unit normal, known as G^1 continuity, guarantees that the surface gradient of the velocity potential is

orthogonal to the normal vector and enables a more accurate computation of the total gradient. The presentation concludes with several movies showing the simulated bubble migration and collapse, which are in good agreement with other known simulations (that assume axisymmetry) and experimentally obtained photographs. Results for multiple-bubble scenarios in non-axisymmetric configurations will also be presented. [Work supported by the ARL:UT Postdoctoral Fellow Program and NIH DK070618.]

11:15

1aPAb5. Acoustic streaming caused by the oscillation of an irregular surface. Katherine Aho, Elaine Vejar, and Charles Thompson (CACT, Univ. of Massachusettes Lowell, 1 University Ave., Lowell, MA 01854)

The analysis of acoustic streaming generated by oscillatory flow over a two-dimensional flexible boundary will be presented. The time-averaged slip velocity that results from the gradient of the Reynolds stress is obtained. Of particular interest is the coupling between transverse motion of boundary and longitudinal oscillation of the fluid.

MONDAY MORNING, 26 OCTOBER 2009 BOWIE (LOSOYA CENTER), 10:00 A.M. TO 12:00 NOON

Session 1aSC

Speech Communication: Talker Variation and Identification (Poster Session)

Ewa Jacewicz, Chair
Ohio State Univ., Speech and Hearing Science, 1070 Carmack Rd., Columbus, OH 43210

Contributed Papers

All posters will be on display from 10:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 10:00 a.m. to 11:00 a.m. and contributors of event-numbered papers will be at their posters from 11:00 a.m. to 12:00 noon.

1aSC1. Dialectal variations on the realization of high tonal targets for focused and non-focused syllables in Taiwan Mandarin. Yi-Hsuan Huang (Grad. Inst. of Linguist., Natl. Taiwan Univ, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan, r94142011@ntu.edu.tw) and Janice Fon (Natl. Taiwan Univ., Taipei 10617, Taiwan)

This study investigated how focus and non-focus were realized across sentential positions for different dialects of Taiwan Mandarin. Ten male subjects from the northern and central regions of Taiwan participated in this study and were asked to read Mandarin high-falling tones in focus and non-focus positions placed sentence initially, medially, and finally. Results showed that F0 maxima for focused syllables were higher than non-focused syllables across sentential positions and dialects. Furthermore, the northern dialect had higher F0 maxima for both focused and non-focused syllables than the central dialect. A significant interaction among dialects, focus status, and sentential positions was found. For northern speakers, differences in F0 maxima for focused and non-focused syllables were greatest in sentence-final positions and smallest in sentence-initial positions. However, for central speakers, the magnitude of F0 maxima differences between focused and non-focused syllables were similar across sentential positions. Reasons for the tonal variations and different declination patterns of the two dialects were discussed.

1aSC2. Multilevel modeling of speaker variation in cross-dialectal articulation rate. Ewa Jacewicz, Robert Allen Fox (Speech Percept. and Acoust. Labs., Speech and Hearing Sci., Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, jacewicz.1@osu.edu), and Lai Wei (Ohio State Univ., Columbus, OH 43210)

Articulation rate (excluding pause time) in spontaneous speech was examined for northern speakers of American English from Wisconsin and for southern speakers from western North Carolina. The corpus consisted of speech samples from 192 speakers, males and females, ranging from 8 to 90 years old. The focus of this study is to model statistically both the between-speaker and within-speaker variations. Within-speaker changes in tempo depend primarily on phrase length so that longer phrases, containing more syllables, tend to be spoken faster than shorter phrases. To capture this type of variation, mixed-effect models were used and the phrase length was included as a predictor. Fixed-effect covariates of interest included dialect, gender, age, and phrase length. The results indicate that phrase length has a highly significant fixed effect on articulation rate. When the differences in phrase length were controlled, the speech tempo of Wisconsin speakers was found to be significantly faster than of North Carolina speakers and male speech was significantly faster than female. The effects of age were also significant showing that articulation rate increases with age achieving the peak at 46 years and decreases with age thereafter. [Work supported by NIH.]

1aSC3. A linguistically-informative approach to dialect recognition using dialect-specific context-dependent phonetic models. Nancy F. Chen, Wade Shen, and Joseph P. Campbell (MIT Lincoln Lab., 244 Wood St, C266, Lexington, MA 02420)

In this work, we explore automatic approaches to learn dialect discriminating pronunciation patterns and use these patterns to automatically recognize dialects. Since linguistic literature suggests that dialect differences often occur in certain phonetic contexts [2, 7, 8, 9], we extend adapted phonetic models [Shen *et al.* (2008)] to consider phonetic contexts. We evaluate our system on classifying American and Indian English. Despite many challenges (e.g., subdialect issues and suboptimal phone recognition accuracy due to lack of word transcriptions), we discover dialect discriminating biphones compatible with the linguistic literature, while outperforming a baseline system by 7.5% (relative). Our work is an encouraging first step toward a linguistically informative dialect recognition system, with potential applications such as forensic phonetics and accent training tools. [This work is sponsored by the Command, Control and Interoperability Division (CID), which is housed within the Department of Homeland Security's Science and Technology Directorate under Air Force Contract No. FA8721-05-C-0002. N. C. F. is also supported by the NIH Ruth L. Kirschstein National Research Award and NIH/NIDCD Grant Nos. DC02978 and T32DC00038.]

1aSC4. Low-back vowel merger in Minnesotan English. Kaitlyn Arctander, Hannah Kinney, and Christina M. Esposito (Dept. of Linguist., Macalester College, 1600 Grand Ave., St. Paul, MN 55105, esposito @macalester.edu)

Current linguistic observation suggests that the low-back vowels (specifically, international phonetic alphabet [ɑ] and [ɔ]) are merging in the majority of American English dialects. There has yet to be a phonetic study of this merger which would provide empirical support for this observation in the mid-western varieties of English. The current study investigates this merger experimentally. Speakers of Minnesotan English were asked to read

a list of 100 monosyllabic words containing the low-back vowels (filler words were also produced). The first (F1) and second (F2) formants were measured automatically for each vowel. Observations of speakers suggest three possible outcomes: (1) both vowels merge to [ɑ], (2) both merge to [ɔ], or (3) that the merger results in a novel vowel which is slightly higher than [ɑ] but lower than [ɔ].

1aSC5. The influence of linguistic complexity on the acoustic and perceptual correlates of gender typicality in boys' speech. Benjamin Munson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr., SE, Minneapolis, MN 55455, munso005@umn.edu), Eden Kaiser, and Laura Crocker (Univ. of Minnesota, Minneapolis, MN 55455)

Studies have shown that adults discern differences between the speech of boys and girls as young as 4 years old (e.g., Perry, Ohde, Ashmead, JASA, 2001). Crocker and Munson (2006, NWAV) examined variation in gendered speech within a group of boys. They examined acoustic measures of vowels and consonants in the single-word productions of 30 5–12 year old boys. Fifteen of these boys were given the label gender identity disorder (GID) by a trained psychometrist based on their nonspeech behaviors. The other 15 were age matched boys without GID. Crocker and Munson found acoustic differences between the groups. These were sufficient to lead naive adult listeners to rate the boys with GID as sounding less boy-like than boys with GID. We expanded on this by examining acoustic and perceptual measures of sentences produced by the same group of boys. The group differences in perceived gender typicality were larger than for single words. These were likely the consequence of the longer speech materials that were used, as acoustic differences were comparable to those found by Crocker and Munson. Together, these findings suggest a link between the early development and gender identity of the development of gender typicality in boys' speech.

1aSC6. Acoustic changes associated with transgender speech therapy: A case study. James Dembowski (Dept. of Speech Lang. Hearing Sci., Texas Tech Univ. Health Sci. Ctr., 3601 4th St., Lubbock, TX 79430, james.dembowski@ttuhsc.edu)

This case study documents acoustic changes in speech and voice characteristics over the course of several months' therapy to feminize the voice in a male-to-female (M-F) transgendered individual. Perceptual tests show that higher speaking fundamental frequency (F0) may be the only significant difference in voices perceived as female rather than male [Petit, J. M., *MIT Encyclopedia of Communication Disorders*, edited by R. D. Kent MIT, Cambridge, MA, 2003, pp. 223–226]. Paradoxically, however, changes in F0 alone are not sufficient to produce a feminine sounding voice. Voice quality, speech rate, and inflection also contribute to the perception of femininity. This presentation documents acoustic changes in a speaker who specifically did not wish to aggressively pursue alterations of F0, but wanted to achieve a more feminine voice by other means. Therapy goals focused on speech rate, voice quality, and prosodic patterns. Acoustic measures over the course of therapy included average F0, minimum F0, F0 variability, speech rate, rms intensity, and signal to noise ratios during selected vocalic segments. Preliminary data analyses suggest that changes in vocal quality and prosodic variability also produced F0 changes, even though these were not targeted therapeutically.

1aSC7. The role of gender and ethnicity in the specification of American English vowels. Sonja A. Trent-Brown, Kyle P. Woodworth, and Joline C. Dreyer (Psych. Dept., Hope College, 35 E. 12th St., Holland, MI 49423, trentbrown@hope.edu)

This study explores acoustic variation in vowel phoneme production for adult speakers of General American English. Peterson and Barney (1952) conducted an analysis of the acoustic features of American English vowels in order to map the phonological vowel space; Hillenbrand *et al.* (1995) replicated and extended their classic study. Both studies presented acoustic measures for men and women. The findings across the two studies include target values for various acoustic cues in the speaker productions, including fundamental frequency (F0), formant frequencies, and duration measures. However, no mention is made of speaker ethnic background, and studies have since suggested that there is evidence of both perceptual and acoustic variations with respect to speaker ethnicity [Thomas & Reaser, (2004)]. The purpose of the current research is to replicate and extend the findings of the previous studies to determine whether differences exist with respect to ethnicity that do not arise from dialect features. Speakers were recorded producing words and sentences containing /hVd/ target stimuli. Speaker productions were screened for dialectal features. Temporal and spectral steady-state vowel measures of acoustic variation (F0, formant values, and vowel duration) were obtained for all speakers. Results are presented for the variations across the speaker gender/ethnic groups. [Hope College Psychology Department, The Carl Frost Center for Social Science Research, The Jacob E. Nyenhuis Student/Faculty Cooperative Grant

1aSC8. Acoustic correlates of gender and ethnicity in speaker perception. Sonja A. Trent-Brown, Joline C. Dreyer, and Kyle P. Woodworth (Psych. Dept., Hope College, 35 E. 12th St., Holland, MI 49423, trentbrown@hope.edu)

Perceptual and acoustic explorations have determined that certain cues are evident in the auditory signal that provide information about speaker characteristics such as gender and ethnicity. Perceptual studies present stimuli to listeners who make decisions based on their perceptual interpretations. Results from such studies indicate that listeners are able to identify speaker ethnicity and gender [Lass *et al.*, (1979); Boonstra *et al.*, (2006)]. Acoustic studies analyze the speech to determine which cues, spectral or temporal, help listeners make their interpretations [Walton and Orlikoff, (1994); Xue and Fucci, (2001)]. The current study presents spectral and temporal analyses to provide a more comprehensive picture of various acoustic contributions to speaker voice types that listeners may utilize in their perceptual identifications. Speech samples (single word- and sentence-length utterances) were produced by 20 male and female, African American and European American speakers of General American English. Results suggest that various spectral and temporal measures contribute to differences across voice characteristics. Differences attributed to speaker gender include fundamental frequency, intensity, pitch and amplitude perturbations, and the harmonics-to-noise ratio. Measures of stimulus intensity and pitch perturbation differed across speaker ethnicity. Differences resulting from the interaction of speaker gender and ethnicity include fundamental frequency and intensity. [Hope College Psychology Department, The Carl Frost Center for Social Science Research.]

1aSC9. Listener voice identification in foreign and accented English. Michelle Sims (Dept. of Linguist., Univ. of AB, 4-32 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada, mnsims@ualberta.ca)

This study investigates perceptual voice identification by analyzing the quantitative and qualitative acoustical characteristics listeners use to identify voices in non-native speech. Research on voice identification typically focuses on the direct quantitative comparison of voices to find unique identifiers [e.g., Morrison (2009) and Jessen (2008)]. However, in this study an experiment was run in order to analyze the acoustic cues people perceptually adhere to in identifying voices. Listeners of various native language backgrounds were asked to identify the voices of Mandarin Chinese speakers in both Mandarin and Mandarin-accented English speech. Listeners' accuracy and speaker selections in fourteen voice line-ups (Sullivan and Schlichting 1998) were recorded. Speakers' productions for all stimuli were analyzed using twelve acoustic features. The results find that though listeners were highly accurate at the voice recognition task, their errors do follow systematic trends. Quantitative and qualitative acoustic measures such as pitch, pitch variation, formant trajectories, intensity, and nasality prove to be reliable in patterning memorable and forgettable voices. That is, this study finds that certain acoustic characteristics are more salient in everyday voice identification.

1aSC10. Talker recognition using envelope modulation spectra. Susan J. LeGendre (Dept. of Speech, Lang. and Hearing Sci., Univ. of Arizona, P.O. Box 210071, 1131 E. 2nd St., Tucson, AZ 85721-0071), Julie M. Liss (Arizona State Univ., Tempe, AZ 85287-0102), Andrew J. Lotto (Univ. of Arizona, Tucson, AZ 85721-0071), and Rene Utianski (Arizona State Univ., Tempe, AZ 85287-0102)

The envelope modulation spectrum (EMS) is a spectral analysis of the low-rate (0–10 Hz) amplitude modulations of the envelope for the entire speech signal and within specific frequency bands. Using a set of predictor variables computed from these spectra, LeGendre *et al.* [J. Acoust. Soc. Am. **125**, 2530–2531 (2009)] were able to classify sentences by dysarthric sub-

types with a remarkable degree of accuracy, presumably because the EMS encodes rhythmic perturbations characteristic of each subtype. In the current study, we tested whether the EMS encodes perturbation differences within a subtype by attempting to classify sentences according to individual talkers. Stepwise discriminant analyses were performed on five sentences produced by 12 speakers with ataxic dysarthria. The 60 sentences were accurately assigned to talker 93% of the time. Thus, individual patterns of motor disruption are readily observed in the speech amplitude envelope. A similar analysis was performed with a set of normal control speakers and, again, classification of sentence by talker was remarkably high. These results suggest that talkers have idiosyncratic rhythmic patterns in their speech production that are independent of linguistic content (which was controlled in this study). The EMS may be a useful measure for talker recognition. [Work supported by NIH/NIDCD].

1aSC11. The role of source and filter characteristics in human talker identification: Experiments with laryngeal and electrolarynx speech. Tyler K. Perrachione (Dept. of Brain and Cognit. Sci., MIT, Cambridge, MA 02139, tkp@mit.edu), Cara E. Stepp (Harvard-MIT Div. of Health Sci. and Technol., Cambridge, MA 02139), Robert E. Hillman (Massachusetts General Hospital, Boston, MA 02114), and Patrick C. M. Wong (Northwestern Univ., Evanston, IL 60208)

Differences in individuals' vocal anatomy and physiology result in unique acoustic features of their vocalizations. Humans are exceptionally attuned to these variations and use them to identify familiar individuals. Although these abilities are often called "voice recognition," talker identity cues actually arise through interactions between acoustic excitation produced at the source (typically, the larynx) and both static and dynamic properties of the filter (vocal tract, articulators, and their manipulations during speech). We investigated the differential contributions of source- and filter-related information to talker identification through four experiments using laryngeal (typical) and electrolarynx speech from five talkers. Using an electrolarynx energy source removed individual differences in vocal anatomy, leaving only unique filter properties for talker identification. Listeners learned talker identity best from typical, laryngeal speech, which contained both unique source and filter cues. Listeners were also able to learn talker identity from electrolarynx speech, which homogenized talker source characteristics. Curiously, listeners did not generalize talker identity across source mechanisms: Training on laryngeal or electrolarynx speech resulted in chance performance identifying the same talkers using the other source mechanism. We consider the implications of these results for models of talker identification and articulatory compensation during electrolarynx use. [Work supported by NIH.]

1aSC12. Speaker variability when producing repeated syllables and across speech tasks. Christina Kuo and Gary Weismer (Dept. of Communicative Disord., Univ. of Wisconsin-Madison and Waisman Ctr., 1975 Willow Dr., Madison, WI 53706, kuo2@wisc.edu)

Target theory [Lindblom, J. Acoust. Soc. Am. **35**, 1773–1781 (1963)] of vowel perception and production suggests that vowels have unique and characteristic articulatory and acoustic events that constitute the canonical forms ("targets") of vowels. However, speakers do not always achieve these targets. There exists substantial variability in speech within a speaker and across speakers. Some of the challenges in understanding speech production lie in variability resulting from various factors including speaker identity, prosody, rate, and speech task. Because of the lack of data permitting good statistical estimates of target events, the following experiments were performed. First, within-speaker variations in consonant-vowel (CV) and vowel-consonant transitions in three monosyllabic words (*dock, knock*, and shock) during a sentence reading repetition task are evaluated for two speakers. Second, acoustic vowel targets in /h/-vowel-/d/ or-/t/ and selected consonant contexts (CV-/d/ or-/t/) are studied in the speech production of one individual across three speech tasks: citation, passage reading, and conversational speech. The hypothesis is a coarse-grained consistency with fine-grained variability in the acoustic signal and a continuum of changes in the acoustic vowel targets across tasks. Findings and continuing directions are discussed within the framework of target theory.

1aSC13. Variation and stability in judgments of talker similarity. Benjamin Munson and Celina C. Marnie (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, munso005@umn.edu)

Listeners judge some pairs of talkers sound more similar to one another than others. This study examined whether the parameters that listeners attend to when judging talker similarity depend on the type of experience they have listening the talkers' voices. We hypothesized that judgments of talker similarity would change over the course of experiment depending on the type of exposure that listeners have to different talkers. Listeners who identify words that talkers produce should attend less to features not relevant for word recognition (i.e., f0 and voice quality), while listeners passively exposed to talkers should not change the parameters that they attended to. Two groups of listeners participated in a three-phase experiment. In the first phase, they judged the similarity of pairs of 16 talkers. In the second, one group conducted a recognition-memory task with the same talkers' voices. The other listened passively while completing a visual-memory task. In the third, they listened to the same talkers and provided a new set of talker-similarity judgments. Multidimensional scaling was used to analyze the talker-similarity judgments. Preliminary analyses suggest that the dimensions that characterize talker similarity spaces are similar in the first and third phases for both groups of listeners.

1aSC14. Variations in intensity, fundamental frequency, and voicing for teachers in occupational versus non-occupational settings. Eric J. Hunter and Ingo R. Titze (Univ. of Utah VP Res., NCVS, 201 Presidents Circle, Rm. 210, Salt Lake City, UT 84112-9011, eric.hunter@ncvs2.org)

This study creates a more concise picture of the vocal demands placed on teachers by comparing occupational voice use with non-occupational voice use. The National Center for Voice and Speech voice dosimetry databank was used to calculate phonation time dose as well as average dB SPL and F0. Occupational voice use (9am–3pm) and non-occupational voice use (4pm–10pm, weekends) were compared from 57 subjects (2 weeks each, 8400+ h). Five key findings were uncovered: (1) previous findings of occupational (30%) and non-occupational (14%) voicing were substantiated; (2) teachers experienced a wide range of occupational voicing percentages (33%, SD $\pm$11%); (3) the occupational voice was on average only about 1-dB SPL louder than the non-occupational voice and remained constant throughout the day; (4) the occupational voice exhibited an increased pitch, trending upward throughout the day; and (5) apparent gender differences in voicing percentages, as well as changes in dB SPL and F0, were shown. Data regarding voicing percentages, F0, and dB SPL provide critical insight into teachers' vocal health. Further, because non-occupational voice use is added to an already vocally overloaded voice, it may add key insights into recovery patterns and should be the focus of future studies.

Session 1aSP

Signal Processing in Acoustics, Animal Bioacoustics, Acoustical Oceanography, and Underwater Acoustics: Categorization of Animal Acoustic Signals

Sean K. Lehman, Chair
Lawrence Livermore National Lab., 7000 East Ave., Livermore, CA 94550-9234

Chair's Introduction—8:45

Invited Papers

8:50

1aSP1. A time-frequency approach for studying propagation effects on underwater sound. Patrick J. Loughlin (745 Benedum Hall, Univ. of Pittsburgh, Pittsburgh, PA 15261, loughlin@pitt.edu)

Sound is perhaps the primary means for communication, navigation, and exploration in the ocean, for both man (e.g., sonar) and marine animals (e.g., whale and dolphin vocalizations). The propagation of underwater sound can undergo frequency-dependent effects, especially in shallow water, which can cause the sound to change from location to location. Spectrograms, or more generally time-frequency analysis, of underwater sounds shows in a natural and dramatic way the changes that can occur as the sound propagates. In this talk, we review and illustrate time-frequency analysis of sounds, with a focus on characterizing propagation effects such as dispersion and damping. A simple but accurate time-frequency approximation method is also discussed, which suggests a feature extraction approach to obtain metrics of the sound that are invariant to particular propagation effects. These features may be useful for automatic classification of different animal vocalizations. We also show how modern time-frequency methods, such as the Choi–Williams and Zhao–Atlas–Marks distributions, can reveal detail in vocalizations that is obscured in spectrograms. [Work supported by ONR 321US.]

9:10

1aSP2. Kalman filter tracking of dolphin whistle countours. Philip Top (Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94550, top1@llnl.gov)

The sounds of animals surround us daily with a cacophony of chirps, tones, and squeaks. For some creatures, underlying this symphony is a great deal of structure and potentially a complex communications channel. As one of the more intelligent creatures, the sounds of dolphins are of particular interest for further examination. Typical dolphin sounds consist of a series of broad spectrum pulses overlayed on narrowband signals that vary in frequency over time. The research in this study applies a branching Kalman filter to track the frequency varying spectral lines. The generated tracks can then be analyzed for features of interest, common characteristics, and structure as an initial step in potentially creating a decoder key.

9:30

1aSP3. Identifying individual clicking whales acoustically I. From click properties. George E. Ioup, Juliette W. Ioup, Lisa A. Pflug (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148), Christopher O. Tiemann, Alan Bernstein (Univ. of Texas at Austin, Austin, TX), and Natalia A. Sidorovskaia (Univ. of Louisiana at Lafayette, Lafayette, LA)

It has been observed that clicks within a single sperm whale coda have similar time and frequency properties, but these properties can differ from one coda to the next. Thus, it may be possible to identify individual sperm whales acoustically from the properties of their coda clicks. While sorting the codas by eye is promising, there is an obvious need to have automated identification by computer. Clustering methods present one appropriate approach to the identification of individuals. Although both K-means and self-organizing maps have been applied, the advantage of the latter, i.e., the ability to limit the number of clusters automatically, has favored its use. More recently, differences in echolocation clicks have become the basis of possible acoustic identification of individuals, also. Experiments by the Littoral Acoustic Demonstration Center in 2001, 2002, and 2007 as well as Detection, Classification, and Localization Workshop data have enabled extensive testing of these ideas. Results of the clustering for both types of clicks are encouraging. Localization and echolocation click rhythms have been used to verify the clusters. The identification has been extended to echolocation clicks of beaked whales. Illustrative results are presented. [Research supported by ONR and SPAWAR.]

9:50

1aSP4. Identifying individual clicking whales acoustically II: From click rhythms. Natalia Sidorovskaia, Philip Schexnayder (Dept. of Phys., Univ. of Louisiana at Lafayette, Lafayette, LA 70504-4210, nas@louisiana.edu), George E. Ioup, Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148), Christopher O. Tiemann, Alan Bernstein (Univ. of Texas at Austin, Austin, TX 78712), Alexander Ekimov, and James Sabatier (Univ. of Mississippi, Oxford, MS 38655)

Many species of marine mammals live in socially organized groups and exhibit cohesive behavior. Since acoustics is their primary communication tool, it has been suggested that acoustic signal characteristics and emission patterns carry attributes of individuals among the group. A dynamic approach for rhythmic analysis of echolocation and communication signals is presented. The algorithm

provides detection of a particular species in a continuous stream of broadband acoustic data and a robust (to low signal-to-noise ratio) method for association of rhythmic frequencies with individuals. The method is applied to passive acoustic recordings collected by the Littoral Acoustic Demonstration Center (LADC). The efficacy of the method for sperm whale echolocation clicks and codas, and beaked whale clicks is addressed. The algorithm is a part of an integrated tool proposed by LADC for identification of individual animals in a group. In parallel, multi-attribute similarity analysis and source localization are applied to the same volume of data. The results are integrated and compared to provide reliable identification. The proposed approach is beneficial for passive acoustic studies of marine species population and social behavior and may contribute to understanding acoustic communication among marine mammals. [Research supported by the Louisiana Optical Network Initiative, ONR, and SPAWAR.]

10:10—10:30 Break

10:30

1aSP5. What has been learned by tracking flying bats from their sonar broadcasts. James A. Simmons (Dept. of Neurosci., Brown Univ., Box G-LN, Providence, RI 02912, james_simmons@brown.edu)

Echolocating bats are active at night, often in total darkness. Although thermal-imaging infrared video is effective for documenting the unexpected diversity of bat activity, the cost and availability of cameras is a big limitation. Infrared strobe-flash photography or night-vision video both require illumination of the scene, which means the site of bat activity has to be known in advance. By far, most observations of bat behavior have been conducted acoustically, using bat detectors for listening to biosonar broadcasts or for plotting the time-frequency structure of the sounds to identify species. In the past decade, ultrasonic microphone arrays have come into use for tracking the flight of bats. Some methods involve widely spaced microphones for three-dimensional reconstruction of flight-paths in large spaces enclosed by the array. Other methods are more focused, using arrays to pinpoint the location of the sounds from off to the side, sometimes using two arrays and triangulation of the sounds sources. Important findings have emerged about the bat's sense of surrounding space and orientation of the sonar beam during flight. Acoustic-array methods coordinated with video recordings reveal new details about how bats look at scenes or interact when flying in groups. (Work supported by ONR and NIMH)

10:50

1aSP6. Categorization of animal sounds using algorithms from diverse applications. Grace A. Clark (Lawrence Livermore Natl. Lab., 7000 East Ave., L-130, Livermore, CA 94550, clark9@llnl.gov)

The proposed approaches to animal sound characterization are motivated by the philosophical theme of this special session; diverse problems, similar solutions. The literature primarily makes use of some fundamental algorithms. This paper proposes some powerful techniques that have the potential to effect significant advances in animal acoustics. Several classes of algorithms are discussed, along with examples of how they have been used to solve similar problems in diverse applications. The key algorithm areas discussed are (1) statistical feature extraction from two-dimensional or multi-dimensional hierarchical transforms, such as wavelet transforms, higher-order poly-spectral features, histogram features, texture features, etc.; (2) feature selection algorithms to choose a subset of most useful features and overcome the curse of dimensionality; (3) advanced classification algorithms; (4) performance measures that include estimates of statistical confidence for small sample sizes; and (5) sensor fusion algorithms for exploiting a variety of sensor modalities. Signal processing examples from diverse problem domains are presented. These include seismic oil exploration, ultrasonic nondestructive evaluation of materials, buried mine detection, target recognition, and others. It is demonstrated that these diverse problems have similar solutions, and it is proposed that some animal acoustics problems could be amenable to similar approaches.

MONDAY MORNING, 26 OCTOBER 2009 RIO GRANDE CENTER, 9:00 A.M. TO 12:00 NOON

Session 1aUW

Underwater Acoustics: Acoustics in Ocean Sediments

Marcia J. Isakson, Chair
Univ. of Texas, Applied Research Labs., 10000 Burnet Rd., Austin, TX 78758

Contributed Papers

9:00

1aUW1. Scattering by sinusoidal pressure-release surfaces. Darrell Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The problem of sound scattering by surfaces having sinusoidal shape has been intensively studied, yet fundamental questions remain. The present work employs this problem as a testing ground for our understanding of small-roughness perturbation theory. Extending previous work by the author, the convergence of the perturbation series is studied using numerical calculations. As has been noted previously, the Rayleigh hypothesis does not set the bounds for convergence to the correct radiated field, but is relevant to determination of the surface field. In the present work, both topics are examined in some detail, and conformal mapping is used to support some of the conclusions.

9:15

1aUW2. Three-dimensional rough surface scattering using finite elements. Sumedh Joshi and Marcia Isakson (Appl. Res. Labs., 10000 Burnet Rd., Austin, TX 78758, sumedhj@mail.utexas.edu)

In order to quantify the effects of three dimensional scattering, the scattering of a spherical incident wave from a rough, pressure release surface is

modeled using a commercially available finite element (FE) code. The surface is generated by creating random realizations from a spatial power spectrum measured as part of the experimental validation of acoustic modeling techniques sea test conducted off the coast of Isola d'Elba in 2006. Since the FE model approaches an exact solution as the discretization density increases, it can be used as a benchmark for approximate methods. Therefore, the three-dimensional (3-D) FE model will be compared with a 3-D Kirchhoff approximation solution as well as two-dimensional solutions based on the integral equation, Kirchhoff, and FE models in order to determine the range of validity of the approximate methods.

9:30

1aUW3. Scattering by unexploded ordnance with variable burial depth and seafloor roughness: A parametric study. David C. Calvo, Brian H. Houston, Joseph A. Bucaro, Larry Kraus, Harry J. Simpson, and Angie Sarkissian (Acoust. Div. Naval Res. Lab., Washington, DC 20375)

Efficient search strategies for unexploded ordnance (UXO) by sonar greatly depend on the ocean environment, particularly bottom roughness, and the mean burial depth of the target. To gain insight into how detection ranges vary with statistical properties of the ocean bottom and factors associated with object burial, we apply full-wave numerical modeling using the time-domain elastodynamic finite-integration technique (EFIT). The numerical technique—which is not restricted to small interface slopes or amplitudes—allows us to obtain a uniform picture of enhancement of acoustic bottom penetration with frequency and roughness parameters, while providing roughness backscattering levels for variable incidence angles. For the target, attention is focussed mostly on scattering near beam incidence, and comparison with experiments is made for a 5-in. rocket UXO. Good agreement with measured target strength over the 4–30-kHz band is obtained using a two-dimensional EFIT model with simple finite-aperture corrections for three-dimensional (3-D) effects. We consider both forward and backscattering and discuss the potential advantages of both. For lower frequencies, we compare with a full 3D finite-element model which compares favorably with data. [Work supported by SERDP and ONR.]

9:45

1aUW4. Measurement of sound transmission through mud at Dodge Pond, Connecticut. William M. Carey and Allan D. Pierce (Dept. of Mech. Eng., College of Eng., Boston Univ., Boston, MA 02215)

Questions important to the sonic detection of buried ordinance are whether the sound dispersion and attenuation of muddy bottoms can be predicted and verified. Wood and Weston [Acustica (1964)] measured compressional speeds in harbor mud 3% less than that of water with attenuation considerably less than those of sandy/silty sediments. A recent theoretical treatment [Pierce and Carey, POMA, 7001 (2009)], making use of the Mallock–Wood equation and of a card-house theory of the structure of mud, estimates the slow sound speed to depend on porosity as $1-(0.35)(1-\phi)$. Present measurements at frequencies between 1 and 10 kHz with a buried array in the depositional mud at the bottom of Dodge Pond, which contains considerable gas microbubbles, yield speeds of the order of 60% of the sound speed in water. The initial measurements on the disturbed sediment were found to be strongly influenced by scattering from larger bubbles whereas the results after a period of 10 months showed the effect of a smaller size distribution of bubbles. Estimates based on the Dodge Pond measurements and on the card-house theory of the propagation characteristics and of the effect of micro-bubbles are discussed. [Sponsored by SERDP-NSWC-PCD.]

10:00—10:15 Break

10:15

1aUW5. Low-frequency seismo-acoustic propagation near thin and low-shear speed ocean sediment layers. Jon M. Collis (Colorado School of Mines, Golden, CO 80401)

Accurate and efficient prediction of propagation over realistic models of elastic ocean sediments has been achieved recently using parabolic equations. A concern has been low-shear wave speed sediments that can become singular as the wave speed tends toward zero. A historic approach for treating a sediment of this type has been to assume that it is a fluid and that effects due to elasticity are negligible. This approach does not account for second order effects such as energy loss due to frequency-dependent attenuation. In addition, thin sediment layers typical of that found at the ocean bottom interface have been difficult to treat numerically. At low frequencies, layers of this type can be treated as a massive interface between the water and higher-shear speed sediment basement layers. To satisfy interface conditions across the layer, Rayleigh jump conditions are imposed [F. Gilbert, Ann. Geophys. **40**, 1211 (1997)]. A consequence of this approximation is that interface and other wave types become dispersive where they were not previously. In this presentation, the massive elastic interface is benchmarked with an elastic parabolic equation and the effects of resultant errors are quantified. In addition, the energy partition between compressional and shear energies is determined.

10:30

1aUW6. The applicability of a small towed array system to the ocean bottom sediment classification problem. William M. Carey (Dept. Mech. Eng., Boston Univ., Boston, MA 02215) and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

An autonomous array system is uniquely suited to perform near bottom ($\approx$1 m) measurements of horizontal wave number spectra that estimate water and interface wave numbers. Synthetic coherent processing provides the required gain to observe the interface wave peak 40–50 dB below the peaks of water modes. Comparable measurements were performed with an array of seismometers (Muir *et al.*, 1991). These spectra contain the information necessary to perform inversions that account for compressional and shear wave speed gradients. Spatial transforming the computed pressure field produces spectra comparable with measurements. Variations of the geoacoustic profiles show that the present gradient produces additional identifiable spectral structure corresponding to shear waves trapped within the sediments. In the experiments discussed, the sound source was at the bottom and the array was towed by a vehicle. However, it is feasible to either use a source of sound on the vehicle or on a companion vehicle to do surveys. This method is a fast, efficient, and accurate ocean acoustic measurement tool useful in rapid waveguide characterization with the appropriate inversion technology. (Work partially sponsored by ONR-OA.)

10:45

1aUW7. Relationships between intrinsic sediment attenuation, modal attenuation, and transmission loss in experiments over sandy-silty sediments. Stephen V. Kaczkowski, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY, 12180, kaczks@rpi.edu), William M. Carey, Allan D. Pierce (Boston Univ., Boston, MA 02215), and Wendell Saintval (Univ. of Miami, Miami, FL 33149)

Numerous shallow water acoustic transmission experiments over sandy-silty bottoms demonstrate that the frequency dependence of intrinsic sediment attenuation at frequencies less than 1 kHz is nonlinear. Computational analyses including modeled geoacoustic profiles have shown that good agreement with experimental data can be obtained for frequencies between 50 Hz and 1 kHz. Upper sediment attenuation values for 1 kHz are in ranges specified by Hamilton [J. Acoust. Soc. Am. **68**, 1313–1340 (1980)], and a nonlinear frequency dependent attenuation with power-law of about 1.8 is necessary. In this presentation, the relationship between the power-law exponent and modal attenuation coefficients is quantified for several nearly range-independent experiments in different locations. These include the Gulf of Mexico (1972), the Strait of Korea (ACT III, 1995), the New Jersey Shelf (1995), and Nantucket Sound (2005). The intrinsic sediment attenuation behavior with frequency and depth implies that modal attenuation coefficients may be obtained using parameters from the models of experimental environments. The behavior of range averaged transmission loss and the modal attenuations of contributing modes is examined. The former is shown to be less sensitive than the latter to parameter changes in the usual situation where the upper sediment layer is the primary source of attenuation. [Work partially supported by ONR.]

11:00

1aUW8. The effects of scattering and poro-elasticity in reflection coefficient measurements. Marcia J. Isakson and Nicholas P. Chotiros (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

Reflection coefficient measurements taken as part of the experimental validation of acoustic modeling techniques experiment in October 2006 off the coast of Isola d'Elba, Italy display frequency dependent behavior that cannot be described using elastic reflection models and spherical wave effects alone. Furthermore, analysis of the effects of scattering based on measured interface roughness does not predict the critical and supercritical angle behavior. Therefore, a poro-elastic model is developed to account for the frequency dependent behavior. A comparison of the model including poro-elasticity, spherical wave effects, and scattering with the data will be presented. [Work sponsored by Office of Naval Research, Ocean Acoustics.]

11:15

1aUW9. The role of porosity fluctuations in scattering from sand sediments and in propagation losses within the sediment. Brian T. Hefner and Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698)

Acoustic backscattering from a diver-smoothed sand sediment was measured at frequencies from 200 to 500 kHz as a function of grazing angle. The residual roughness of this smoothed surface was measured using a laser line scanning system capable of measuring sub-millimeter heights over a 4-meter track. Using the measured sediment roughness, perturbation theory underestimates the scattering strength at all frequencies for angles greater than the critical grazing angle. The absence of large, discrete scatterers in the sediment suggests that scattering from fluctuations in the sediment properties may be the dominant scattering mechanism for these angles. The sound speed and attenuation were also measured in this sediment and the attenuation was found to exhibit a linear frequency dependence similar to that observed for other sand sediments. To account for this linear attenuation, a theory that incorporated scattering losses due to porosity fluctuations into the effective density fluid model has been developed. This theory suggests that the propagation losses may be connected by the same physical mechanism to the scattering of sound from the sediment. This connection is explored in the context of these scattering and propagation measurements.

11:30

1aUW10. Measurements of sound speed, attenuation, and normal-incidence reflection coefficient in water-saturated glass beads as a function of porosity. Theodore F. Argo, IV, Matthew D. Guild, Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292), Matthias Schröter (Max Planck Inst. for Dynam. and Self-Organization, Göttingen, Germany), Charles Radin (Dept. of Mathematics, The Univ. of Texas at Austin, Austin, TX 78712), Donald J. Suntrup, and Harry L. Swinney (Ctr. for Nonlinear Dynam., The Univ. of Texas at Austin, Austin, TX 78712)

Sound propagation in water-saturated granular sediments is known to depend on the sediment porosity, but few data in the literature address both its frequency and porosity dependency. To address this deficiency, a method to control porosity [Argo *et al.* J. Acoust. Soc. Am. **124**, 2469 (2008)] was used to prepare artificial sediments composed of water-saturated glass spheres (130, 265, and 530 μm diameters) with porosities ranging from 0.37 to 0.43. Through-transmission measurements were performed for frequencies from 250 to 800 kHz to determine the sound speed and attenuation. Pulse-echo measurements for frequencies from 0.25 to 7.5 MHz were performed to determine the normal-incidence reflection coefficient. For both classes of measurements, the Fourier phase technique was used to analyze the data. Measurement results will be presented and compared to existing models. A Biot-based model was found to qualitatively describe the porosity dependence of the sound speed and attenuation, but a transition to a scattering-dominated regime was observed. [Work supported by NSF, ONR, and the Welch Foundation.]

11:45

1aUW11. Acoustic detection of mines buried in the sediment in shallow waters using a laser Doppler vibrometer. Alexander Ekimov, James Sabatier, and Vyacheslav Aranchuk (NCPA, Univ. of MS, 1 Coliseum Dr., Univ., MS 38677, aekimov@olemiss.edu)

Naval operations in littoral regions require accurate methods and sensors to detect mines, thereby preventing injury to personnel and damage to equipment. Mines may be buried in the sediment in shallow waters, on beaches, or regions adjacent to the water. Acoustic detection of a mine buried in the seabed involves two major issues: exciting the vibration of a mine and measuring the vibration response of the mine and the environment. A small laboratory-like testing facility, a loudspeaker, and a laser Doppler vibrometer (LDV) were used in experiments to detect buried mines in sand and under water. Two possible scenarios for measurement of the vibration response of a mine and the environment were used: sensing vibration of the water surface and sensing vibration of the sand surface. The on-mine and off-mine mechanical impedances were measured with the LDV. The possibility of the transmission of the mine's vibration to the surface of the water where it can be measured with an LDV was demonstrated. In these experiments, glass beads were suspended on the water surface to improve the reflected light returning to the LDV. Test results are presented and discussed. [Work supported by the Office of Naval Research Grant N00014-09-1-0016.]

Session 1pAA

Architectural Acoustics, Noise, and ASA Committee on Standards: Acoustic Test Facilities: Design, Qualification, and Testing

Douglas F. Winker, Chair
ETS-Lindgren, 1301 Arrow Point Dr., Cedar Park, TX 78613

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Qualification of anechoic chambers: 70 years on and we are still trying to get it right. Kenneth A. Cunefare (School of Mech. Eng., The Georgia Inst. of Technol., Atlanta, GA 30332)

There are now three equally "acceptable" techniques in common use for qualification of anechoic and hemi-anechoic chambers. The historical approach for modeling the inverse square law decay uses the measured sound pressure level at a given reference position from the source and applies the inverse square law therefrom to compute "theoretical" levels at other distances. A second approach, explicitly embodied in ISO 3745:2003 Annex A, seeks to estimate a source strength and source center offset location that yields a theoretical decay that matches the observed decay in a least-squares sense. This method is motivated by the concern that the true acoustic center of a sound source may not coincide with a visually identifiable point on or near the source. More recently, a third method has come into use which determines an apparent source strength that maximizes the qualified span of a traverse. Nonetheless, the 3745 standard does not constrain the analysis routine, such that one can fully conform to the standard through use of any of these three methods (or other!) and yet obtain different results from the same data. The challenges of qualification are all the more confounded by issues related to sound sources, signal type, and spatial resolution.

1:25

1pAA2. Considerations related to the standardization of methods for the qualification of free field test environments. Jeff G. Schmitt (2512 Star Grass Circle, Austin, TX 78745, jeffs@prodigy.net)

The methods for qualification of acoustic free field environments, such as anechoic and hemi-anechoic chambers, have historically been defined in Annexes to the basic sound power standards, such as ISO 3745. However, free field test environments are used for many applications other than just sound power level determination, such as basic acoustic research, audio test, and hearing sciences. There currently are no general standards for free field qualification that addresses the unique needs of these test methods. Although the basic inverse square law qualification method is very straightforward theoretically, there are a number of practical considerations that can make this qualification method technically complex, logistically difficult, and expensive to implement. Industry standards must strike the right balance between the technical needs of the test methods and the costs associated with chamber qualification, particularly in the chamber vendor/chamber customer business environment. This paper will address some of the practical issues associated with the current inverse square law qualification method and the work of ISO Technical Committee 43, Working Group 8 that is currently in the process of developing a more general and improved document for the qualification of free field environments using this method.

1:45

1pAA3. Hemi-anechoic chamber source considerations for chamber qualification in compliance with ISO 3745:2003(E). Douglas F. Winker (ETS-Lindgren-Acoust. Systems, 1301 Arrow Point Dr., Cedar Park, TX 78613, douglas.winker@ets-lindgren.com)

Source directivity specifications for qualifying the inverse square law performance are given in ISO 3745:2003(E) Annex A.2.2.2. These specifications are difficult to meet, especially for a single source. Multiple sources are required to qualify a chamber across its usable range. Sources used at low frequencies are relatively simple to design. As frequency increases, compliance becomes increasingly difficult. The design parameters will be discussed, and the advantages and disadvantages of several sources will be discussed. Both commercially available and new designs were tested in accordance with A.2.2.2 with varying results. The frequency ranges of source qualification will be evaluated as well as their impact on the chamber traversing process. Source directivity patterns will be presented and compared to A.2.2.2. Deviations from A.2.2.2 will be examined, and the reasons for these deviations will be evaluated.

2:05

1pAA4. Impulse response characterization of anechoic and hemi-anechoic chambers. David A. Nelson (Nelson Acoust., P.O. Box 879, Elgin, TX 78621)

Characterization of a measurement environment is an important component of assessing measurement uncertainty. The traditional technique in anechoic and hemi-anechoic chambers is qualification through draw-away testing: deviation from square-law spreading is estimated along various radials, for broadband noise and occasionally for tones. All locations within a volume for which the deviations are below certain criteria are considered "qualified" for testing, and measured results from these locations are typically not corrected for

the influence of the room. The draw-away test is tedious and time consuming and does not yield the information that is actually desired: precision of measurements for a particular source-receiver pair. Impulse response methods are considerably more rapid and provide a wealth of information in postprocessing including amplitude, phase, and time-delay factors for both steady-state and transient sounds, across arbitrary frequency ranges. This paper will explain how the impulse response characterization can be obtained and how it can be used to improve precision of measurements in anechoic and hemi-anechoic environments.

2:25

1pAA5. Effect of placement of test hardware on measured sound pressures in a hemi-anechoic chamber. Einar G. Ristroph, Michael C. Black, and John A. Phillips (ETS-Lindgren Acoust. Lab., 1301 Arrow Point Dr., Cedar Park, TX 78613, einar.ristroph@ets-lindgren.com)

Anechoic chamber design includes a focus on minimizing interior exposed reflective hardware such as doors, integrated test fixtures, fire suppression and safety systems, etc. Albeit, a limited amount of exposed reflective hardware is unavoidable. Oftentimes additional specimen and microphone positioning equipment is put into the chamber for actual testing after the acoustic field has been qualified by experts. The effects of reflective objects typical of specimen or microphone positioning systems on measured steady-state sound pressure levels was evaluated experimentally using collinear measurements at successive distances away from a sound source, both with and without reflective objects of specific geometries and positions relative to source and receiver placed in the acoustic field. The size and position of the reflecting objects were varied. Results are presented, and from these data, acoustic experimental design guidelines are derived.

2:45

1pAA6. Room requirements for laboratory measurement of transmission loss and impact sound insulation. Bradford N. Gover and Trevor R. T. Nightingale (Natl. Res. Council Inst. for Res. in Construction, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada, brad.gover@nrc-cnrc.gc.ca)

Laboratory measurement of transmission loss (ASTM E90) and impact sound insulation (ASTM E492) have one thing in common: measurement of the sound power radiated into the receive room by the specimen under test. In E90 the reported quantity, transmission loss (TL), is the difference between the incident and transmitted sound power, whereas for E492 the reported quantity, normalized impact sound pressure level (NISPL), is a direct measure of the radiated sound power of the floor. A derivation of the TL and NISPL equations using a statistical energy analysis (SEA) framework shows that the ASTM equations are based on the assumption of power balance and statistical sampling of the sound field. This SEA framework is used to define the acoustical conditions of the receive room (number of modes in each frequency band, the average loss factor in each band, and hence the modal overlap factor), which are then related to the physical properties such as room shape, volume, absorption, etc. The conditions necessary for meaningful measurements in SEA are compared to ASTM and ISO requirements for receive rooms. A recommendation for possible revisions to enhance test method consistency and accuracy is discussed.

3:05—3:20 Break

Contributed Papers

3:20

1pAA7. Uncertainty of sound power levels determined following Air Conditioning Heating and Refrigeration Institute Standard 220. Stephen J. Lind (Trane Commercial Systems, Ingersoll Rand, 3600 Pammel Creek Rd., La Crosse, WI 54601, slind@trane.com)

As part of the AHRI, ASA-ANSI, and ISO standard development processes, the uncertainty of sound power measurements needs to be quantified for inclusion in sound power standards. A study of three reverberation rooms with volumes of 140, 280, and 560 cm^3 was undertaken. Sound power levels for two noise sources (a horizontal shafted reference sound source and a leaf blower) were determined using ARI Standard 220. Variables investigated include source location, source orientation, room conditions, source operating characteristics, and microphone traverse length. The frequency range of interest was 25 Hz–10 kHz. Measurements were repeated and the order of tests was randomized. Of particular interest was the difference between the 140-ft^3 room and the other rooms to confirm the room qualification process required by ARI Standard 220 because the 140-ft^3 room does not meet the ARI Standard 220 qualification process but does meet the requirements in ISO 3741. This work will be helpful in understanding the sources of variation and allow for a starting point for determining if improvements are needed and if so what factors should be considered.

3:35

1pAA8. Inter-laboratory variation in sound power levels in qualified reverberant rooms. Robert Stabley (2843 Bradley Ave., Dallastown, PA 17313, stabley@cyberia.com)

Reverberant acoustic test facilities can be qualified to determine the sound power levels of broadband and tonal noise sources using the procedures defined in Air-Conditioning, Heating and Refrigeration Institute (AHRI) Std. 220, "Sound Power Testing Using Reverberant Rooms for HVAC Equipment." Member companies from AHRI's Technical Committee on Sound participated in a round robin test program in which tonal noise sources were shipped to, and tested in, a number of qualified reverberant rooms. This report summarizes the results of this effort. The mean and standard deviations of the sound power levels for multiple locations/orientations of the noise sources in each facility and for all facilities are presented. The standard deviations as a function of frequency for these sources were found to be generally less than the values established for broadband sources and therefore less than those allowed for tonal sources. Based on the comparisons of round robin test results, accurate determinations of sound power levels can be made using the substitution method in rooms qualified in accordance with AHRI Std. 220, Technical Committee on Sound, Air-Conditioning, Heating, and Refrigeration Institute

4:05

1pAA9. A proposal for systematic investigation of reproducibility deviations in Laboratory Acoustical Measurements. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

Cases have been observed in airborne sound isolation and impact noise insulation testing where apparently identical assemblies have achieved significantly different STC and/or IIC ratings, not only when tested at different test facilities but also when re-tested in the same facility [J. LoVerde and W. Dong, J. Acoust. Soc. Am. **125**, 2629 (2009); J. Acoust. Soc. Am. **123**, 3504 (2008)]. Acoustical test data will be presented demonstrating several cases and showing that the deviations are significantly larger than the expected statistical variation based on the reproducibility of the test method as reported within the applicable ASTM standards. The reason(s) for the variation in these test results are unknown, and there has been no systematic investigation of possible causes. Standards committees, test laboratories, and government agencies have not shown a satisfactory level of interest in either quantifying the deviations or exploring the reasons behind them, which may be theoretical, material, and/or procedural. A protocol is presented for a series of investigations that could be performed in an acoustical testing laboratory intended to identify and quantify causes of the variations.

4:20

1pAA10. Tape as a surface protector for field tapping machine tests. Alicia J. Wagner, Jeffrey Fullerton, and Jonah Sacks (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138, awagner@acentech.com)

Impact sound transmission measurements of floor-ceiling constructions are often performed in the field. This leads to a wide range of testing environments; some of which can be high-end, multi-family residences with expensive floor finishes. Some residents can be concerned that testing may damage or scuff their floor. The ASTM e1007-04E1 standard does not discuss or provide allowances for acceptable methods of floor protection. This study investigates various kinds of tape including storage, masking, duct, and aluminum tape to find options that act as good surface protectors while minimally influencing measurement results. Testing was conducted on hardwood and tile floors. Results will be shown comparing the impact sound transmission frequency spectra of the floors using each tape with the frequency spectra measured using the bare floor. Suggestions for modifications to future standards are proposed to address this concern.

MONDAY AFTERNOON, 26 OCTOBER 2009 RIO GRANDE EAST, 1:00 TO 5:15 P.M.

Session 1pAO

Acoustical Oceanography and Underwater Acoustics: Session in Honor of Stanley Flatté II

Timothy F. Duda, Cochair
Woods Hole Oceanographic Inst., 98 Water St., Woods Hole, MA 02543

Lisa M. Zurk, Cochair
Portland State Univ., Electrical and Computer Engineering Dept., 1900 SW Fourth Ave., Portland, OR 97207

Chair's Introduction—1:00

Invited Papers

1:05

1pAO1. Stanley Flatte: Atmospheric optics contributions. Rod Frehlich (CIRES, Univ. of Colorado, UCB 216, Boulder, CO 80309, rodney.frehlich@colorado.edu)

Stanley Flatte and his research group produced many key contributions to the field of atmospheric optics: both theoretical and based on numerical experiments. The results of numerical simulation of wave propagation through the random atmosphere (scintillation), the new theoretical predictions of scintillation statistics, and the implications of this work for understanding experimental results will be reviewed. Much of this work is based on techniques applied to acoustical problems of sound propagation in the ocean. The unique properties of the atmosphere that are responsible for the large magnitude of optical scintillation were revealed by these careful analysis. Theoretical predictions of basic scintillation statistics for many conditions of interest were found to be lacking. New applications, unresolved issues, and future directions of this research will also be discussed.

1:25

1pAO2. Amplitude and phase fluctuations of seismic waves and characterization of small-scale heterogeneities in the earth. Ru-Shan Wu, Xiao-Bi Xie, and Yingcai Zheng (Earth and Planetary Sci., Univ. of California, Santa Cruz, CA 95064)

Due to the limitation of sparse seismic observations on the earth's surface, the deterministic picture of small-scale heterogeneities in the earth is difficult to obtain in general. Statistical characterization is a viable alternative to study the distribution and properties of the small-scale heterogeneities. The pioneering work of Aki in 1973 modeled the heterogeneities with a single-layer uniform random medium having only three parameters: scale, perturbation strength, and layer thickness. This simple model is consistent with the observations which are limited to only the transversal coherence functions of seismic wave (phase and amplitude) across an array and it has no depth resolution. The following work in 1970s and early 1980s were based on similar models. Flatte introduced the key idea of angular coherence functions using different earthquakes at large epicentral distances in the late 1980s (Flatte and Wu, 1988), which opened a new area of earth heterogeneity characterization by extending the single, uniform layer model to a depth-dependence model.

Later, the angular coherence function was extended to the joint transverse-angular coherence function, which significantly increases the depth resolution of the heterogeneity spectra. In this talk, we summarize the research results along this direction conducted by a group of scientists in the earth science at University of California, Santa Cruz, collaborated with Flatte in different stages.

1:45

1pAO3. Statistics of the forward field propagated through three-dimensional random internal waves in an ocean waveguide. Purnima Ratilal (Dept. of Elec. and Comput. Eng., Notheastern Univ., 360 Huntington Ave., Boston, MA 02115), Tianrun Chen, and Nicholas Makris (MIT, Cambridge, MA 02139)

Internal waves randomize the forward propagated acoustic field in both continental shelf and deep water environments. This has implications in underwater communications and remote sensing using wide area sonar systems. Here, we derive compact analytic expressions for the mean, variance, mutual intensity, spatial covariance, and temporal coherence of the acoustic field propagated through three-dimensional (3-D) random internal waves. These take into account multiple forward scattering effects and are derived in terms of the statistics of the internal waves. For typical continental shelf environments, we show that when internal wave height exceeds the acoustic wavelength, the acoustic field becomes so randomized that the expected total intensity is dominated by the field variance beyond moderate ranges. This leads to an effectively saturated field that decays monotonically and no longer exhibits the periodic range-dependent modal interference structure present in nonrandom waveguides. 3-D scattering effects become important when the Fresnel width exceeds the cross-range coherence length of the internal wave field. Our derivation makes it possible to predict the coherence time scale of field fluctuations in ocean-acoustic measurements from knowledge of the oceanography. It is used to explain the time scale of acoustic field fluctuations observed at megameter ranges in various deep ocean-acoustic transmission experiments.

2:05

1pAO4. Single-path acoustic scintillation results from the Shallow Water 2006 Experiment. Dajun Tang, Daniel Rouseff, Frank Henyey, and Jie Yang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th Str., Seattle, WA 98105, djtang@apl.washington.edu)

In "sound transmission through a fluctuating ocean," Flattxe *et al.* described saturation of a single acoustic path as that path becoming a number of interfering uncorrelated micropaths due to refraction by internal waves. The probability density function of intensity becomes exponential with a scintillation index of 1.0. In deep water, however, full saturation is not achieved due to weak scattering and absorption. Mid-frequency (1–10 kHz) data from the Shallow Water 2006 Experiment are used to determine single-path intensity statistics. At a range of 1 km in water 80 m deep, an acoustic path is isolated that went through two upper turning points separated by a single bottom reflection. The data were collected during a period when large nonlinear internal waves were absent. The scintillation index calculated from the data increases with frequency until reaching a maximum of 1.2 around 6 kHz. It then decreases to 1.0, suggesting that single-path saturation can be achieved at mid-frequencies in shallow water. The probability density functions of intensity at various frequencies show a trend toward exponential. Because shallow water internal waves are dominated by the first mode, uncorrelated micropaths are an unlikely mechanism for producing the observed saturation. [Work supported by ONR.]

2:25

1pAO5. Sound transmission through the internal-wave field: The inverse problem. John Colosi (Dept. of Oceanogr., Naval Postgrad. School, Monterey CA 93943, jacolosi@nps.edu)

Buoyed by the early success of ocean acoustic tomography for ocean mapping, and of path integral theory combined with the Garrett–Munk internal-wave model for explaining acoustic fluctuations, Flatte and others could not resist the temptation of using statistical moments of the acoustic field to learn something about the spatial and temporal distributions of ocean internal waves. Indeed, internal waves are of intense oceanographic interest because of their relation to tides, eddies (which account for 99% of the ocean kinetic energy), and the mixing processes that maintain the thermocline. Because of the integrating property of acoustics and the sensitivity of certain acoustic observables to different parts of the internal-wave spectrum, acoustical techniques like those described by Flatte offer a unique view into internal-wave behavior. Recent acoustic and oceanographic measurements made in the Philippine Sea using a large vertical aperture array and high temporal sampling may provide the first field demonstration of internal-wave tomography, progress of which would be quite pleasing to Flatte.

Contributed Papers

2:45

1pAO6. Robust observables for mode tomography. Tarun K. Chandrayadula, Kathleen E. Wage (Dept. of Elec. and Comput. Eng., Mailstop 1G5, George Mason Univ., Fairfax, VA 22030), James A. Mercer, Bruce M. Howe, Rex K. Andrew (Univ. of Washington, Seattle, WA 98195), Peter F. Worcester, and Matthew A. Dzieciuch (UCSD, La Jolla, CA 92093)

Deep water tomographic inversions should include the low modes to improve estimates of the axial sound speed. Unfortunately, the low-mode signals are strongly affected by internal wave fluctuations, making it very challenging to estimate their arrival times. This talk describes the use of matched subspace detectors (MSDs) to measure mode travel times for ranges up to 400 km. The MSD approach was first proposed by Scharf and Friedlander [IEEE Trans. Signal Process. **42**, 2146–2157 (1994)] as a way to detect signals that lie within a particular subspace. In this work the subspace for each mode is defined by analyzing simulations of mode propagation through independent realizations of the internal wave field. Simulations show that the MSD-based travel time estimation approach yields higher accuracy and less variance than other methods such as peak-picking. The end result is improved inversions for sound speed near the sound channel axis. In addition to the simulation study, mode signals from the 2004 Long Range Ocean Acoustic Propagation EXperiment (LOAPEX) are used to invert for the sound speed profile (SSP) across the LOAPEX path. The inverted SSP is compared with environmental measurements made during LOAPEX. [Work sponsored by ONR.]

3:00—3:15 Break

3:15

1pAO7. Comparison of fluctuations of a broad-band pulsed acoustic signal in the Philippine Sea to simulation. Andrew W. White, Frank S. Henyey, Rex K. Andrew, James A. Mercer (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105), Peter F. Worcester, and Matthew A. Dz-

ieciuch (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238)

The North Pacific Acoustic Laboratory (NPAL) has conducted several long-range (>500-km) experiments in the North Eastern and North Central Pacific Ocean. In April 2009, NPAL conducted an engineering test/pilot study in the more oceanographically energetic Philippine Sea in preparation for a longer-duration experiment in 2010. Transmissions from a ship-suspended source were to be recorded on a vertical line array at a comparatively shorter range of ≈100 km. Statistics of fluctuations of the acoustic field are found by simulated broad-band signal propagation through a model of the Philippine Sea environment. Model parameters are chosen following analysis of environmental data taken during the engineering test. Simulation-based predictions are compared to experimental data and to predictions of acoustic fluctuations in the less-energetic North Eastern and North Central Pacific Ocean. [Support for this work was provided by the Office of Naval Research.]

3:30

1pAO8. Effect of random hydrodynamic inhomogeneities on low-frequency sound propagation loss in shallow water. Andrey A. Lunkov and Valeriy G. Petnikov (A. M. Prokhorov General Phys. Inst., RAS, 38 Vavilov Str., Moscow 119991, Russia, landr2004@mail.ru)

Low-frequency (100–500 Hz) sound propagation loss on the US Atlantic continental shelf and in the Barents Sea in the presence of stochastic surface waves, and for the US Atlantic shelf also in the presence of internal waves, is studied for the range of up to 150 km by means of numerical simulations. Surface waves simulations are carried out with Pierson–Neumann spectrum. Internal wave field realizations are modeled according to frequency spectrum of thermocline displacements that were recorded in Shallow Water'06 experiment. Qualitative difference between sound propagation loss behavior on the US Atlantic shelf and in the Barents Sea is demonstrated for summertime conditions even without random inhomogeneities. It is shown that whereas internal waves, on average, have a weak effect on propagation loss, surface waves result in its considerable increase in both areas under wintertime conditions with a wind speed of more than 9 m/s. [The authors wish to acknowledge and thank Dr. J. F. Lynch for discussing this work. The work was supported by the Russian Foundation for Basic Research Project 08-02-00283 and by the Civilian Research and Development Foundation Project No. REC 010.]

3:45

1pAO9. Amplitude and phase front variations for individual modes in Shallow Water 2006 Experiment. Mohsen Badiey (College of Earth, Ocean, and Environment, Univ. of Delaware, Newark, DE 19716, badiey@udel.edu), Boris Katsnelson, Andrey Malikhin (Voronezh Univ., Voronezh 394006, Russia), and James Lynch (WHOI, Woods Hole, MA 02543)

Using modal selection techniques, developed for an L-shaped array, the amplitude and phase front variations for individual modes are considered for shallow water waveguides in the presence of internal waves (IWs). Experimental data collected during the Shallow Water 2006 Experiment include acoustic transmissions during several IW events. In this paper, we focus on one of these IW episodes (i.e., event 50) where a train of nonlinear IW was moving approximately across an acoustic track. Low-frequency LFM signals with band 270–330 Hz were transmitted from a source moving with the advancing IW front and were received by an L-shaped array for few hours. During this time, the angle between the acoustic track and the IW front changed while transmissions continued. The waveguide in the presence and the absence of a train of IW was observed providing excellent environmental input for modeling. Fluctuations of the intensity as well as the modal phase are studied using experimental data and modeling of the waveguide during this period. [Work supported by ONR321OA and CRDF].

4:00

1pAO10. Arrival time variations of pulses in shallow water and low-frequency acoustical underwater positioning. Boris Katsnelson (Dept. of Phys., Voronezh Univ., 1, Universitetskaya Sq., Voronezh 394006, Russia, katz@phys.vsu.ru) and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

Variations in acoustic arrival time in shallow water are presented to establish a method for long range underwater acoustical positioning systems (UWPSs) with possible applications for underwater vehicles. At the present time, there are different methods used for UWPS for only short distances (about a few hundred meters), relating the sound signals to distance based on ocean physical properties. It is possible to use arrival times of high-frequency signals (up to 100 kHz) radiated by a set of reference sound sources (such as 3–4), placed at different points in an area to determine the position. However, the main problem for long range positioning (i.e., operation area of several hundred to thousand kilometers) is significant spatial and temporal variabilities of waveguide environmental properties affecting the signal travel time. To establish limits for low-frequency acoustical UWPS, here we use recent field observations in the SW06 experiment to estimate signal arrival time variations caused by different hydrodynamic conditions observed simultaneously during acoustic transmissions. Based on preliminary mapping of the operation area, and the subsequent optimization, the accuracy related to the waveguide dynamics is presented. [Work supported by ONR321OA and CRDF.

4:15

1pAO11. Observation of shallow water turbulences and model comparisons. Tokuo Yamamoto (Div. of Appl. Marine Phys., Univ. of Miami, Miami, FL 33149) and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

Observations of acoustic wave propagation through moving media reveal that the tidally generated turbulent currents occur around the clock. Only turbulent amplitude oscillates with geo-time. The Kolmogorov model of turbulent energy spectra and measured turbulent energy spectra is in good agreement. Turbulence is a poor conductor of gas because of thick bubble packed layer near the sea surface. Analytical solution shows strong vertical and horizontal circulations. It is anticipated that the bubble packed from the surface layer would be submerged into the turbulent currents. The turbulent energy data are compared with the internal wave energy. For most cases compared the turbulent energy is greater than that of the internal wave energy. From measured turbulent energy, the turbulence field and Reynolds stress field are obtained. The vertical and horizontal circulations in the turbulence are extracted using inverse methods.

4:30

1pAO12. Measurement and modeling of high-frequency acoustic channels for a moving source in shallow water. Joseph M. Senne, Aijun Song, and Mohsen Badiey (Ocean Acoust. Lab, Univ. of Delaware, Newark, DE 19711, sennejm@udel.edu)

Acoustic systems used in underwater communications are sensitive to both scattering and Doppler shifts. In order to assess these sensitivities, an experiment (KAM08) was conducted during summer 2008 under variable environmental conditions. During the experiment, high-frequency (<10 kHz) acoustic signals were emitted from a towed source and measured by multiple hydrophone arrays moored in 100-m depth. Environmental parameters including wind, surface waves, and water-column temperature were collected concurrently with acoustic transmissions, which took place under variable sea states. Linear frequency modulated (LFM) signals and coherent *m*-sequence transmissions in the frequency band of 13–19 kHz were used to determine detailed channel characteristics for multiple source-receiver geometries. Both parabolic equation and ray-based models were run with moving source and simulated sea surface configurations in order to accurately reproduce experimental conditions. [Work supported by ONR 321OA].

4:45

1pAO13. Efficient high-frequency underwater acoustic propagation through random media with wavefront predistortion by singular value decomposition: a communication perspective. Lisa J. Burton (Dept. of Mech. Eng., MIT, 77 Mass. Ave., Cambridge, MA 02139, lisab@mit.edu), Andrew Puryear, and Pierre F. J. Lermusiaux (MIT, Cambridge, MA 02139)

Acoustic propagation through the fluctuating ocean environment severely limits the capacity of existing underwater communication systems. Specifically, surface waves, small scale turbulence, and fluctuations of physical properties generate random signal variation that creates deep fades and limits reliable communication to low data rates. This paper presents new theoretical signal processing techniques that use channel state information to provide high-rate reliable communication. Specifically, a general framework for efficient propagation through the ocean random media is presented, where efficiency is defined in terms of minimizing bit error rate. Based on results from this propagation framework, a communication architecture that optimally predistorts the acoustic wave via spatial modulation by singular value decomposition and detects the acoustic wave with optimal spatial recombination to maximize reliable information throughput is presented. This effectively allows the system to allocate its power to the most efficient propagation modes while mitigating intersymbol interference. The system consists of a plane of transmit apertures (acoustic emitters) and a plane of receive apertures (hydrophones). This method is applicable to link ranges of 100–1000 m using frequencies between 10 and 30 kHz. Many emerging applications, including real-time ocean observation systems, have the potential to benefit from the proposed architecture. [Support from NSF-GRFP and ONR.]

5:00

1pAO14. Virtual time series experiment (VIRTEX) simulation tool for underwater acoustic communications. J. S. Kim, H. C. Song, W. S. Hodgkiss (Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238, jeasoo@ucsd.edu), and M. Siderius (Portland State Univ., Portland, OR 97207-0751)

Recently a virtual time series simulation tool for underwater acoustic communications, referred to as virtual time series experiment (VIRTEX), has been developed [Siderius and Porter, J. Acoust. Soc. Am. **124**, 137–150 (2008)]. The code is based on time-varying impulse responses, capable of simulating the time series incorporating relative motion between a transmitter and a receiver as well as time-evolving moving boundaries. In this paper, VIRTEX is applied to KAM08 experimental environment to simulate the characteristics of acoustic transmissions under varying ocean conditions. Specifically, a comparison is made between KAM08 data and the simulated time series in terms of scattering functions that characterize Doppler and delay spreads along with representative channel impulse responses. It is found that VIRTEX can be useful for designing and planning communication experiments and evaluating the performance of communication systems.

MONDAY AFTERNOON, 26 OCTOBER 2009 REGENCY EAST 2, 1:30 TO 5:00 P.M.

Session 1pBB

Biomedical Ultrasound/Bioresponse to Vibration: Bubbles, Cavitation, and Heating

Parag V. Chitnis, Chair

Riverside Research Inst., Lizzi Ctr. for Biomedical Engineering, 156 William St., New York, NY 10038

Contributed Papers

1:30

1pBB1. Inertial cavitation thresholds during standing wave disruption using random phase modulation. Caleb H. Farny, Sai Chun Tang, and Greg T. Clement (Dept. of Radiology, Brigham and Women"s Hospital, Harvard Med. School, 221 Longwood Ave., Boston, MA 02115, cfarny@bwh.harvard.edu)

In low-frequency transcranial ultrasound therapeutic applications, such as sonothrombolysis, standing waves have been identified as a potential source of undesirable collateral damage due to cavitation. Recently, a method for disrupting the standing wave pattern was developed by modulating the phase of the driving signal, producing a beam pattern with significantly lower ripple in pressure along the acoustic axis. Here we present a study conducted to determine the inertial cavitation threshold in tissue-mimicking phantoms placed in a standing-wave-inducing chamber when the 272-kHz focused therapy transducer was excited separately by a monochromatic or phase-modulated continuous-wave signal. A 10-MHz receiver was used as a passive cavitation detector whose signal was digitally sampled to examine changes in the broadband signal amplitude. Results indicate that the pressure threshold for inertial cavitation is higher during random phase modulation, suggesting that such an approach would be suitable for transcranial ultrasound therapy. These results agree with a simulation of similar excitation signals on the bubble dynamics, which suggest that disruption of timing on bubble forcing reduces the explosive growth required to induce inertial cavitation. [Work supported by NIH Awards R25 CA089017-06A2 and U41 RR19703.]

1:45

1pBB2. Closed-loop adaptive feedback control of cavitation during high-intensity focussed ultrasound therapy. Natalie Hockham, Manish Arora, and Constantin-C Coussios (Inst. of Biomedical Eng., Univ. of Oxford, Old Rd., Campus Res. Bldg., Oxford OX3 7DQ, United Kingdom)

Prior experimental evidence indicates that the occurrence of inertial cavitation at the HIFU focus enhances the local rate of heat deposition, due to increased absorption of the broadband emissions associated with collapsing bubbles. However, initiating and sustaining cavitation activity throughout HIFU exposure remain a major challenge. Previous open-loop attempts at maintaining cavitation activity by varying the HIFU pulse duration have necessitated the use of duty cycles on the order of 20%, which are too low to achieve ablative temperatures. The present work presents the design, validation, and testing of a real-time closed-loop adaptive controller for sustaining thermally relevant cavitation within the HIFU focal region at duty cycles in excess of 95%. During 1.067 MHz HIFU exposure of an agar tissue-mimicking material, cavitation activity is continuously monitored and quantified using a single-element broadband passive cavitation detector (PCD) positioned coaxially with the HIFU transducer. The HIFU amplitude is adaptively modulated in direct response to the PCD signal, making it possible to maintain cavitation activity and a temperature rise in excess of 30 °C for more than 30 s. Most importantly, the acoustic energy needed to achieve such focal heating is only a fraction of that required in the absence of cavitation control.

2:00

1pBB3. Determination of ultrasound contrast agent shell rupture thresholds using double passive cavitation detection. Daniel A. King, Alayna C. Roberts, Michael J. Malloy, Alexandar Haak, Christian C. Yoder, and William D. O'Brien, Jr. (Bioacoustics Res. Lab., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews Ave., Urbana, IL 61801, daking3@illinois.edu)

Double passive cavitation detection involves the confocal alignment of one transmit and two passive receive transducers. Highly dilute mixtures of ultrasound contrast agents (UCAs), including Definity, Optison, and non-commercial microbubbles, were prepared such that on average only a single UCA was present at the confocal region. A 3-cycle incident pulse repeating at 10 Hz was varied in frequency from 0.9 to 7.1 MHz and in peak rarefactional pressure (PRP) from 200 kPa to 6 MPa. The acquired signals were classified according to the presence or absence of a postexcitation signal (PES), a quantifiable aspect of the acquired temporal signal which is hypothesized to occur due to the rupture of the UCA shell and therefore indicate the collapse of the microbubble. The resulting percentages of UCAs with PES were fitted with modified logistic regression curves to determine percentage collapse thresholds. The percentage of signals exhibiting PES was found to increase as frequency decreased or as PRP increased while holding other acoustic parameters constant, in agreement with alternative methods of measuring collapse. Definity was found to have a lower threshold of collapse than Optison, while the thresholds of the non-commercial UCAs depended on their physical properties. (Work supported by NIH Grant No. R37EB002641.)

2:15

1pBB4. Response of ultrasound contrast agents to static overpressure. Parag V. Chitnis, Paul Lee, Jonathan Mamou, and Jeffrey A. Ketterling (F. L. Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., 9th Fl., New York, NY 10038)

This study investigated the static pressure threshold for contrast agent (CA) rupture. Polymer-shelled CAs (mean diameter of 3 μm) with a constant shell-thickness-to-radius ratio (STRR) were subjected to overpressure in a cylindrical test chamber containing an optical window on its top and bottom faces. A video-microscope (50× magnification, 128 μm FOV, 1 Mpixel CCD) was used to image the CAs while static pressure was increased from 0.5-22 psi by pumping water into the chamber at a constant flow rate. An image processing algorithm was used to determine the change in diameter as a function of static pressure for each CA (total of 130). The majority of CAs did not respond to static overpressure until the threshold for shell rupture was exceeded, resulting in abrupt destruction. The probability density function of rupture pressures exhibited four lobes and four subsets that were obtained by a *k*-means algorithm. Measurements indicated that mean rupture pressures in each group were independent of CA diameter. This observation is in agreement with the classical theory for a perfectly spherical shell, which predicts that the critical compression load is a function of STRR and not of the diameter alone. [Work supported by NIH EB006372.]

2:30

1pBB5. An *in vitro* study of the relationship between the size of lipid-coated microbubbles and frequency-dependent attenuation coefficient. Yanjun Gong (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, ygong@bu.edu), Mario Cabodi, and Tyrone M. Porter (Boston Univ., Boston, MA 02215)

Monodisperse populations of lipid-coated microbubbles were produced using a flow-focusing microfluidic device. The microbubbles were formed by forcing a lipid solution and perfluorocarbon gas through a narrow orifice simultaneously. The average bubble size depended on the lipid solution flow rate, the gas pressure, and the size of the inlet microchannels and narrow orifice. We have successfully produced monodispersions of microbubbles with average diameters ranging from 4–11 μm. Monodisperse bubble populations were suspended in an exposure vessel, and the attenuation coefficient between 1–7.5 MHz was measured using an acoustic spectroscopy method. The frequency at which the peak attenuation coefficient was measured was dependent on the mean bubble diameter. For example, microbubbles with diameters of 4 and 8 μm had peak attenuation coefficients at approximately 3.0 and 1.5 MHz, respectively. Additionally, it was noted that the bandwidth of the frequency-dependent attenuation coefficients was wider for polydisperse populations compared to monodisperse populations of lipid-coated microbubbles. Implications of these results on the acoustic characterization and use of monodispersions of lipid-coated microbubbles for molecular imaging will be discussed.

2:45

1pBB6. Quantitative measurements of microbubble interactions with surrounding vessels and tissues. Hong Chen, Wayne Kreider, John C. Kucewicz, Michael R. Bailey, Andrew A. Brayman, and Thomas J. Matula (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Understanding the coupled interactions between acoustically activated microbubbles and surrounding tissues is important for both diagnostic and therapeutic applications. However, little is known about ultrasound-induced coupled interactions between microbubbles and blood vessels. In this work, high-speed microscopy was used to investigate the coupled interactions inside vessels of *ex vivo* rat mesentery. A 1-MHz focused ultrasound transducer was used to transmit a single pulse to tissue samples having blood vessel diameters of 10–100 μm. Bubble-tissue interactions were captured with a high-speed camera and an inverted microscope. For pressure amplitudes ranging from 1 to 7 MPa, several phenomena were observed: (a) microbubble oscillations caused localized dilation and invagination of the vessels, (b) microbubble jetting occurred where most jets were directed away from the adjacent vessel wall, and (c) connective tissue surrounding the vessels moved with the vessels. The talk will focus on quantitative measurements of vessel motion, strain, and jet velocities. [Work supported by NIH Grants 5R01EB000350, 5R01AR053652, 5P01DK043881, and DK070618.]

3:00

1pBB7. A model for the dynamics of ultrasound contrast agents in soft tissue. Shengping Qin and Katherine Ferrara (Dept. of Biomedical Eng., Univ. of California, Davis, One Shields Ave., Davis, CA 95616, spqin@ucdavis.edu)

For a bubble suspended in an incompressible and infinite liquid, the dynamics of a clean gas bubble can be described by the Rayleigh–Plesset (RP) equation. The RP equation has been applied to approximately simulate the behavior of ultrasound contrast agents (UCAs) in blood, and extended RP equations have been proposed to account for the effects of the UCA shell and surrounding soft tissue. These models produce consistent results with experimental measurements for low acoustic pressure scenarios. For applications of UCAs in therapeutic medicine, the transmitted acoustic pulse can have a peak negative pressure up to a few megapascals, resulting in violent oscillations of UCAs and discrepancies between measurements and predictions using these extended RP equations. Based on the Keller equation, we developed a new model to describe the dynamics of UCAs in the body while taking account of liquid compressibility and shell and soft tissue effects. Liquid compressibility is approximated to the first order, and the shell is treated either as a layer of visco-elastic solid or a viscous liquid. Finite deformation of the shell is included. Typical UCAs with a lipid or albumin shell are investigated, and the effects of soft tissue and blood are examined.

3:15

1pBB8. Ultrasound and ultrasound contrast agent induced angiogenesis: A survival study. Chenara A. Johnson (Dept. of Bioengineering, Univ. of Illinois, 405 N Mathews, Urbana, IL 61801, cjohns42@illinois.edu) and William D. O'Brien, Jr. (Univ. of Illinois, Urbana, IL)

Ultrasound (US) and ultrasound contrast agents' (UCAs), ability to induce a particular bioeffect, angiogenesis, has been explored as a means of restoring blood flow to ischemic muscle. Determining the progression of the angiogenic response and how the acoustic pressure affects that response are integral to the evaluation of bioeffects, US therapy development, and understanding the physical process. Because UCAs demonstrate an increasing percentage of collapse with increasing US pressure, this study explored the effects with a high percentage of UCA collapse (i.e., high US pressure). Angiogenesis was examined in terms of capillary density in rat gracilis muscle (N=21) using H&E, and two immunohistochemical stains: CD-31 and CD-34. Several survival groups (0, 3, 6, 13, 20, and 27 days) were used to de-

termine when the capillary density returned to normal levels. Inflammation cells were also quantified via a color threshold analysis as a percentage of the total section. When compared to the control, high US pressure and UCAs caused capillary density to decrease by 39% and inflammation cells to increase by 30–175% of the inflammation cells at day 0. These values returned to normal levels within 3–6 days. (Work supported by NIH Grant No. R37EB002641.)

3:30—3:45 Break

3:45

1pBB9. The relationship between acoustic backscattering and heating during bubble mediated high intensity focused ultrasound. Peng Zhang and Tyrone Porter (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

It has been demonstrated that cavitating bubbles can increase the heating rate during HIFU thermal ablation. We are developing a pressure-sensitive phase-shift nanoemulsion (PSNE) to nucleate acoustic cavitation in a controlled manner. In this study, we demonstrate the ability to nucleate cavitation with PSNE dispersed throughout acrylamide gels. This allowed us to investigate the acoustic power scattered and radiated by cavitating bubbles in a controlled manner. PSNE was vaporized with short pulses from a 1.133–MHz-HIFU transducer. An spherically focused broadband transducer (10 kHz–20 MHz) positioned coaxially with the HIFU transducer was used to receive emissions from bubbles located at the overlapping foci. The rf signal from broadband transducer was split into two channels. One was filtered with a 10-MHz high-pass filter while the other was not, to record emissions from inertial cavitation (IC) and total backscattered signal, respectively. The temperature change was measured with a thermocouple embedded in gels. A strong correlation was found between temperature change, IC emission, and total backscattered acoustic energy. The relationship between temperature change and bubble density was studied by changing the amount of PSNE suspension added to the gel. Finally, the roles of scattering and IC in the heating effect were discussed.

4:00

1pBB10. Tissue heating in vibro-acoustography. Shigao Chen, Azra Alizad (Dept. of Physio. and Biomedical Eng., Mayo Clinic College of Medicine, Rochester, MN 55905), Wilkins Aquino (Cornell Univ., Ithaca, NY 14853), Matthew Urban, James Greenleaf, and Mostafa Fatemi (Mayo Clinic College of Medicine, Rochester, MN 55905 fatemi@mayo.edu)

Vibro-acoustography (VA) is an imaging method based on ultrasound radiation force. In this method, a two dimensional image is formed by scanning two co-focused ultrasound beams with slightly different frequencies over the object (biological tissue) in a C-scan format and recording resulting sound that is emitted from the focal region at the difference frequency. As in other ultrasound-based imaging modalities, it is important to study the thermal safety of the technique. This paper addresses tissue heating due to a VA scan using a concentric confocal transducer. The three dimensional ultrasound intensity field calculated by Field II is used with the bio-heat equation to estimate tissue heating. Results calculated with thermal conduction and with blood perfusion, with conduction and without perfusion, and without conduction and without perfusion are compared. Maximum temperature rise occurs in the transducer near field during VA scans. Heating during one VA scan is below 0.05 °C for soft tissue with attenuation coefficient of 0.3, 0.5, or 0.7 dB/cm MHz. It is concluded that temperature rise in tissue is not significant. [Work supported by NIH Grants CA 91956, CA 127235, and CA121579. Some of the techniques are patented by Mayo Clinic and the some of the authors (M.F. and J.F.G.).]

4:15

1pBB11. Acoustic droplet vaporization for enhancement of thermal therapies. Paul L. Carson, Man Zhang, Mario L. Fabiilli, Oliver D. Kripfgans, Kevin J. Haworth (Dept. of Radiology, Univ. of Michigan, 3216 Med. Sci. I, Ann Arbor, MI 48109-5667), William W. Roberts, and J. Brian Fowlkes (Univ. of Michigan, Ann Arbor, MI 48109-5667)

Vaporization of superheated microdroplets of perfluorocarbon can be triggered by transcutaneous ultrasound for the local production of gas bubbles. Acoustic droplet vaporization (ADV) was performed in the canine renal artery or segmental artery for blood flow control by capillary occlusion to accelerate rf ablation exclusively in the target vascular bed and for future trapping of released drugs. Occlusions of 82% were obtained for >20 min in the two best exteriorized canine kidneys with intracardiac droplet injections. Intravenous injection studies avoided induced cardiac arrhythmias. In four of the six IV studies, mean achieved renal artery flow reduction was 75% for 3 min. Perfusion by colored microspheres showed larger regional reductions in some cases. Droplets that administered IV or intra-arterially were vaporized by transcutaneous ultrasound in five studies with 3.5–5.5-cm tissue paths to produce increased kidney cortex echogenicity, but not substantial occlusion. Improved occlusion is expected by enhanced bubble coalescence and creation of a narrower droplet size distribution at diameters close to the transpulmonary limit. Achieved ADV bubble concentrations also appear adequate for enhancement of HIFU therapy by reflection of acoustic energy exiting the focus, increasing absorption exclusively therein, and revealing targeted tissues. [Work supported in part by NIH Grants R01-EB000281 and S10-RR022425.]

4:30

1pBB12. Optimal control point distributions for waveform diversity. Robert J. McGough, Matthew R. Jennings, and Xiaozheng Zeng (Dept. of Elec. and Comput. Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824)

Waveform diversity is a method for tumor heating with ultrasound phased arrays that extend the concept of multiple focusing to achieve an optimal sequence of power depositions. Waveform diversity minimizes the power deposited at control points in normal tissue while maximizing the power at control points in the tumor through an approach that employs semidefinite programming. As with other spot scanning and multiple focusing techniques, the performance of waveform diversity depends on the location of the control points. For example, single spot scanning is limited by intervening tissue heating, so the ideal spot scanning approach in large tumors focuses on the back half of the tumor and allows intervening tissue heating to fill in the remaining tumor region. In contrast, waveform diversity combined with mode scanning tends to preferentially deliver heat beyond control points placed in the tumor. In an effort to exploit this feature, the energy delivered to control points on the front face of the tumor is maximized, and the energy delivered to control points along the tumor axis and directly behind the tumor is minimized. Results show that this control point distribution applied to waveform diversity combined with mode scanning achieves conformal heating in a 3-cm-diameter spherical tumor model that is heated by a 1444 element spherical section array.

4:45

1pBB13. Estimation of mechanical properties of tissue using laser-induced microbubble interrogated by acoustic radiation force. Sangpil Yoon, Salavat R. Aglyamov, Andrei B. Karpiouk, and Stanislav Y. Emelianov (Dept. of Biomedical Eng., Univ. of Texas at Austin, 107 W. Dean Keeton St., Austin, TX 78712, s.yoon@mail.utexas.edu)

Measurements of mechanical properties of tissue can be used in many applications ranging from disease diagnosis to treatment planning and monitoring. In this paper, we present a technique to measure mechanical properties of the tissue undergoing laser surgery during which the cavitation bubble is introduced. Utilizing the laser-induced microbubble, the mechanical properties of the tissue were assessed by measuring displacement and oscillations of the bubble under applied acoustic radiation force. The bubbles were produced using a nanosecond pulsed laser operating at 532 nm. Using the lens (0.92 numerical aperture, 11.5-mm focal length), the laser beam was focused inside the gelatin to create microbubbles ranging from 80- to 250-cm diameter. The bubbles were then interrogated using acoustic pulses of different durations and amplitudes. First, the size of each laser-induced microbubble was estimated by measuring the laser-produced shock wave and pulse-echo ultrasound signals. Then, a 1.5-MHz transducer generated an acoustic radiation force to displace the microbubble while a 25-MHz transducer probed the microbubble continuously to measure the resulting displacement. Finally, using theoretical model, viscoelastic properties of the medium were assessed from the measurements of the bubble dynamic behavior. The results were verified using optical methods.

Session 1pED

Education in Acoustics and ASA Student Council: Project Listen Up

Kent L. Gee, Cochair
Brigham Young Univ., Dept. of Physics and Astronomy, Provo, UT 84602

James M. Sabatier, Cochair
National Ctr. for Physical Acoustics, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—3:00

Invited Papers

3:05

1pED1. Bending waves and pressure waves in solid rods. Uwe J. Hansen (Dept. of Chemistry & Phys., Ind. State Univ., Terre Haute, IN 47809)

Dimensionally identical aluminum rods are excited by a hammer impulse. One rod is suspended by rubber bands at a distance of nearly 1/4 of the length of the rod near the rod's ends. The rod is excited near the end in a direction perpendicular to the rod axis. The other rod is supported by a clamp at the middle of the rod. This rod is excited by an impulse on the rod's end face in the direction along the rod's axis. In the first case the standing bending wave has a wavelength of nearly the length of the rod. In the second case the length of the standing wave is nearly twice the length of the rod. The resulting pitches observed clearly indicate that the propagation speed of the longitudinal wave by far exceeds the propagation speed of the transverse wave.

3:25

1pED2. Drumhead resonances in circular elastic membranes. Joseph R. Gladden, III (Dept. of Phys., Univ. of Mississippi, University, MS 38677)

A classic problem in continuum mechanics is determining the normal modes of vibration of a thin elastic membrane with a circular geometry that is clamped at the edges. The analytical solution to the displacement field of such a problem are Bessel functions of the first kind. In this talk, I will describe the construction of a drumhead apparatus consisting of a latex sheet with a circular geometry in which the tension can be varied. Normal modes can be excited using a simple speaker and a strobe light can be used to slow the motion allowing close inspection of the displacement field. The various modes can then be compared to analytical solutions. The shift in natural frequencies of the system due to changing boundary conditions can be demonstrated by placing a hand close to the vibrating membrane.

3:45

1pED3. Why we have two ears-take two: A revised experiment on sound localization. Ralph T. Muehleisen (Civil and Architectural Eng., Ill. Inst. of Tech., Chicago, IL 60616)

Whether they fly, swim, crawl, or walk, vertebrates have binaural auditory systems. Two ears provide redundancy in case of injury to the auditory system, improve the ability to detect quiet sounds, and perhaps most importantly, improve the ability to localize sound. A simple experiment in sound localization using headphone based auralization was presented at a previous meeting. In response to some very useful suggestions a new experiment has been developed. In this experiment, a subject sits in a swivel chair and a sound source is placed a distance away. The subject is rotated a random amount in the swivel chair. The tester then plays the test signal and the subject indicates the perceived direction of the sound. The experiment is repeated with one ear plugged using a high isolation insert earplug. The tests are performed with two types of sounds; a transient noise source with high-frequency tones and/or broadband components such as bird call and source that is primarily low-frequency tones such as a fog horn. The experiments are best performed in quiet rooms with few reflections.

4:05

1pED4. A demonstration of acoustic damping using bubbly liquid for Project Listen Up. Preston S. Wilson, Theodore F. Argo, IV, and Chad A. Greene (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

Acoustic resonances are exploited in many natural and human-made systems, and demonstrations of the phenomena have been presented previously [Greene *et al.*, J. Acoust. Soc. Am. **124**, 2568 (2008); Argo IV *et al.*, J. Acoust. Soc. Am. **125**, 2625 (2009)]. In practice, resonance is always accompanied by energy loss, which limits the maximum excursion of a system. When sufficiently large, this loss can alter the resonance frequency of the system. An air bubble in water is an example of a lossy acoustic resonator and is well-suited for demonstrating acoustic damping in the Listen Up paradigm. For the proposed demonstration, students will observe the resonance of a water-filled drinking glass by striking it with a metal spoon to produce a strong ringing sound. They will then introduce damping into the system by using a stream of water to entrain bubbles which will deaden the ringing response of the system. The

primary mechanism for the acoustic energy dissipation is conversion into heat during bubble oscillation. An explanation of this process is presented with a detailed description of the proposed experiment.

4:25

1pED5. Alvin Lucier's method for determining the acoustic modes of a room. Andrew Piacsek (Dept. of Phys., Central Wash. Univ., 400 E. University Way, Ellensburg, WA 98926-7422)

Students can learn about standing waves in rooms as they recreate the classic 1970 recording by Lucier, "I Am Sitting in a Room". In this work, the composer records himself describing the process in which he is engaged. The approximately 1-min recording of his speech is then played back and re-recorded in the same acoustic environment (an unspecified room). The recording and playback process is repeated several times; the complete work consists of a sequential playback of each iteration. From an acoustically unremarkable, if slightly reverberant, initial recording, the speech is altered with each iteration as the resonance frequencies of the room are amplified relative to other frequencies. Eventually, intelligibility is lost and all that is heard are the modulated tones comprising the harmonic series. Using only a microphone, a digital recording device, a playback system, and a suitable room, students can reproduce this effect while investigating how the room dimensions, shape, and wall materials affect the final outcome.

MONDAY AFTERNOON, 26 OCTOBER 2009 REGENCY EAST 1, 1:00 TO 3:55 P.M.

Session 1pNS

Noise and Architectural Acoustics: Sustainable Building Systems and Acoustic Performance

Erica E. Ryherd, Cochair
Georgia Institute of Technology, Mechanical Engineering, 771 Ferst Dr., 124 Love Bldg., Atlanta, GA 30332

Kelly A. Aston, Cochair
Artec Acoustics, 114 W. 26th St., New York, NY 10001

Invited Papers

1:00

1pNS1. Updated American Society of Heating, Refrigerating and Air-Conditioning Engineers Guidelines and Standards on Sustainable Systems Design. Erica E. Ryherd (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, erica.ryherd@me.gatech.edu) and Tom Lawrence (Faculty of Eng., Univ. of Georgia, Athens, GA 30602)

Many designers look for guidance with the various standards targeting sustainable, green buildings. The American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) publishes a number of useful design resources related to green buildings. The resources are widely used among mechanical systems engineers and facilities designers. This talk will present an overview of several ASHRAE resources which include acoustics aspects. "ASHRAE Green Guide: The Design, Construction and Operation of Sustainable Buildings" is a book to teach designers how to effectively participate on green building design teams and fundamental green design concepts. The book takes an integrated building systems perspective by addressing how engineered systems interact with and are influenced by architectural design. It also includes guidance on the U.S. Green Building Council's LEED® rating system. ASHRAE is also in the final stages of preparing a new Standard for High-Performance Green Buildings (Standard 189.1). This Standard is co-sponsored by the U.S. Green Building Council. Standard 189.1 is being written as an ANSI-level Standard and intended for adoption into state and local building codes.

1:20

1pNS2. Washington State's Sustainable Schools Protocol. Daniel C. Bruck and Jeanette Hesedahl (BRC Acoust. & Technol. Consulting, 1741 1st Ave. S., Ste. 401, Seattle, WA 98134)

With the High-Performance Public Buildings Act signed into law in 2005, Washington State developed the Washington Sustainable Schools Protocol as a goal-setting and planning tool for school districts to clearly communicate design goals and meet requirements for state-funded building projects. Based on the Collaborative for High Performance Schools Criteria and similar to the US Green Building Council's Leadership in Energy and Environmental Design rating system, the Washington Sustainable Schools Protocol system requires school facility projects to achieve a minimum level of points in the categories related to site, water, materials, energy, indoor environmental quality, and innovation. A minimum acoustic performance level for classrooms and other learning spaces is the prerequisite to achieving up to three additional points with improved acoustic performance and audio enhancement. Although clearly a compromise, the acoustics related requirements and optional credits within the Washington Sustainable Schools Protocol has set the starting point for supporting acoustically effective speech communication and learning environments for publicly funded schools in the State of Washington. This paper summarizes the development of the Washington Sustainable Schools Protocol, current acoustical related credits, and lessons learned that may be useful for other jurisdictions considering the development of their own standards.

1:40

1pNS3. Integrating acoustical issues in the design of "green" schools. Gary W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611), Chris Jones, Robert M. Lilkendey, and Hyun Paek (Siebien Assoc., Inc., Gainesville, FL 32607)

ANSI 12.60 and LEED for schools are relatively explicit about the acoustical requirements needed to meet each standard. In practice, many schools are built that develop in an uneasy way trying to meet LEED criteria because a method to integrate acoustical design principles with the basic architectural design scheme of a school has not been assimilated into the professions. This paper presents a core set of acoustical planning principles that can be implemented early in the design process so that LEED and ANSI criteria for room finishes, sound isolation, and building equipment noise can be met as the design progresses.

2:00

1pNS4. Acoustic performance measurement protocol for evaluating "green" buildings. Kenneth P. Roy (Bldg. Products Technol. Lab., Armstrong World Industries, 2500 Columbia Ave, Lancaster, PA 17604, kproy@armstrong.com)

Since a majority of buildings have been shown to be "acoustically challenged," and since "Green" design has only recently begun to address acoustic needs, when the USGBC approached ASHRAE to develop performance measurement protocols (PMPs), acoustics was included as a building IEQ factor. Accordingly, the PMP document provides guidance for the measurement and reporting of building performance, and includes protocols for performance measurement of energy use, indoor environmental quality, and water use. The IEQ protocols apply to thermal comfort, indoor air quality, lighting, and acoustics. The purpose of the PMP is to provide a consistent set of protocols, over a range of cost/accuracy, to enable an appropriate and accurate performance comparison of buildings, especially those claiming to be green, sustainable, and/or high performance. The protocols identify what is to be measured, how it is to be measured, and how often it is to be measured. Additionally, the protocols where developed for three levels of cost/accuracy (low, medium, and high), to provide realistic choices for consistent performance benchmarking. Design of buildings to meet high levels of acoustic performance will require a higher "level of proof" that those design goals have, in fact, been met. This document will be reviewed.

2:20—2:35 Break

2:35

1pNS5. Sustainable design for audio, video, and broadcast facilities. Russ Berger and Richard Schrag (Russ Berger Design Group, Inc., 4006 Belt Line Rd., Ste. 160, Addison, TX 75001, russ@rbdg.com)

Meeting sustainable design goals in facilities that include technical production spaces for audio, video, or broadcast presents special challenges to the design and construction team. Materials selections and sources affect not only the acoustical finishes, but may also influence larger building systems. Daylighting goals must be balanced with appropriate acoustical isolation. The integration of alternative chilled air delivery systems and lighting sources must be weighed against critical technical function. As a result, the acoustician's role should expand to include an understanding of the sustainability goals, the impact of acoustical design recommendations, changes to the specification process, and the necessary coordination efforts with the rest of the participants. Examples from radio and television broadcast facilities, music recording studios, and performance spaces are given.

2:55

1pNS6. Mechanical system noise control in green laboratory design. Benjamin Markham, Doug Sturz, and Jeffrey Fullerton (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

The goals of energy efficiency and low noise are often mutually supportive in the design of mechanical systems for laboratories. This presentation outlines some basic design principles for indoor and outdoor noise controls of laboratory mechanical systems designed to be sustainable and energy efficient. Two themes emerge: the inherent connection between improving energy efficiency and minimizing noise (and minimizing the need for sound attenuation devices that increase system pressure and hence decrease efficiency) and the occasional conflicts between energy efficiency and low noise. The practical pros and cons of specific design guidelines and noise control strategies will be presented, focusing on systems with high-energy efficiency and relatively low system pressures and airflow velocities. Some examples of covered topics include equipment selection guidelines, laboratory exhaust discharge attenuation design or avoidance strategies, variable speed drive design for cooling tower applications, duct layout guidelines, duct leakage requirements, and sound attenuator usage and selection strategies.

3:15

1pNS7. Noise and energy consumption in heating, ventilation and air conditioning systems. Philip Sanders (Charles M. Salter Assoc., Inc., 325 South First St., San Jose, CA 95113, philip.sanders@cmsalter.com) and Peter Rumsey (Rumsey Engineers, Oakland, CA 94607)

Reduced noise and reduced energy consumption are often inter-related in HVAC systems. This presentation reviews some simple engineering principals supporting this assertion and uses case studies to illustrate it. Acoustical consultants collaborating with architects and mechanical engineers may be able to use this information to show the acoustical benefits of energy efficient heating and cooling systems, thus influencing owner decisions in favor of HVAC systems that consume less energy.

1pNS8. Thinking holisitcally about sustainability, acoustics, and architecture. Carl P. Giegold and Scott D. Pfeiffer (53 W. Jackson Blvd., Ste. 1734, Chicago, IL 60604, cgiegold@thresholdacoustics.com)

Highly rated sustainable buildings invariably incorporate energy conservation concepts that come into conflict with tried and true approaches to interior acoustics, isolation, and noise control. The use of concrete slabs for energy storage exposes hard surfaces where we often want soft ones, natural ventilation opens rooms to each other and to often noisy, urban environments, and even seemingly benign concepts like chilled beams can end up replacing one kind of noise with another. Even so, sustainable approaches often have characteristics that are beneficial in acoustic terms, not least in its potential for reducing or eliminating sources of equipment noise. Current turmoil in the design professions as they search for better ways to embrace and explore sustainability presents an opportunity to install greater awareness of the aural world into the built environment. Through architectural case studies both built and in-progress, we look at examples of how interior acoustics and noise control have been approached in ways that solve aural, visual, and environmental challenges holistically.

MONDAY AFTERNOON, 26 OCTOBER 2009 BOWIE (LOSOYA CENTER), 1:30 TO 4:30 P.M.

Session 1pSC

Speech Communication: Issues in Speech Production (Poster Session)

Benjamin R. Munson, Chair
Univ. of Minnesota, Speech, Language, Hearing Science 115 Shevlin Hall, 164 Pillsbury Ave., Minneapolis, MN 55455

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of event-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pSC1. Effect of phonological neighborhood density on word duration in spontaneous speech. Yao Yao (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94709, yaoyao@berkeley.edu)

It has been widely shown that words with many phonological neighbors are harder to recognize than words with few neighbors (L. & Pisoni 1998, among others). However, the effect of phonological neighborhood density on word production is less clear. On one hand, high-density words are easier to produce (Vitevitch 1997, 2002), which suggests that they should be reduced in speech. Meanwhile, these words are often hyperarticulated, probably to make it easier for the listener (Wright 1997, Munson & Solomon 2004, Scarborough 2002). The present work aims to distinguish the two accounts by examining the effect of neighborhood density on word duration. We use data from the Buckeye corpus, which has 40 American English speakers, each recorded for about an hour. A mixed-effect model is built to predict the durations of all CVC monomorphemic content words in the corpus, with speaker and word as random effects. The results show that, after controlling for other factors, neighborhood density has a small but robust negative effect on word duration, i.e., the more neighbors a word has, the shorter it becomes. The effect of average neighbor frequency is in the same direction. Present results provide unambiguous evidence for the talker-oriented account concerning word duration.

1pSC2. Effects of prosodic structure on the relative timing of articulators in English lateral production. Susan S. Lin (Dept. of Linguist., Univ. of Michigan, 440 Lorch Hall 611 Tappan St., Ann Arbor, MI 48109-1220)

Previous research has established that American English speakers tend to produce syllable-final /l/ with movement of the tongue dorsum preceding movement of the tongue tip. However, the results of these studies differ with respect to the articulator timing in syllable-initial /l/, with some claiming synchrony (Browman and Goldstein, 1995) and others claiming asynchrony in the direction opposite that of syllable-final /l/ (Gick, 2003). This study uses ultrasound imaging to investigate the relative timing of the tongue tip and dorsum during production of syllable-initial and syllable-final /l/ in multiple prosodic contexts. Prosody has a significant effect on both duration and extent of articulator movement in speech production—onsets of larger prosodic units involve larger and longer movement than onsets of smaller prosodic (Keating, 2006). The explanation that these effects result from speakers' attempts to render perceptually more clear the segments that initiate phrases and utterances suggests that examining these segments at varying prosodic positions may provide insight into speakers' knowledge of speech perception. Current preliminary results show that American English speakers may utilize at least two distinct timing relations in initial laterals, supporting a position that speaker knowledge may be variable between speakers.

1pSC3. Assessing acoustic measures of the spontaneous phonetic imitation of vowels. Molly Babel (Dept. of Linguist., Univ. of BC, Totem Field Studios, 2612 West Mall, Vancouver, BC V6T 1Z4, Canada, mbabel@berkeley.edu) and Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455)

It is well established that people imitate fine phonetic detail of another talker in shadowing tasks [Goldinger, Psychol. Rev. **105**, 251–279 (1998)] and in interactive conversation [Pardo, J. Acoust. Soc. Am. **119**, 2382–2393 (2006)]. That is, the acoustic characteristics of a model talker's production (MTP) are more similar to a participant's shadowed production (PSP) than they are to that participant's baseline production (PBP), typically elicited in a reading task. This presentation compares two methods of assessing the acoustic distance between the vowels in PSP and PBP monosyllabic words taken from two previous studies [Babel, thesis, University of California (2009); Kaiser & Munson (unpublished)]. One of these measures is the F1/F2 Euclidean distance between PSP and PBP vowels. This measure does not take into account the direction of the difference. The other [adapted from Titze, *Principles of Voice Production* (1994)] compares the distance and direction of the F1 and F2 values in the PSP and PBP vowels relative to those of the MTP. The merits of each of these methods are assessed by comparing them to measures of listener judgments of the similarity of PSPs and PBPs to the MTPs in AXB perception tasks.

1pSC4. The effects of "Smooth Speech" on the phonetic features of stuttered speech: A case study. Caroline Menezes and Lee Ellis (Dept. of Health and Rehabilitative Services, Univ. of Toledo, 2801 W. Bancroft St., Toledo, OH 43623)

This is a preliminary study focused on analyzing the acoustic characteristics of stuttered speech before, during, and after a short treatment using the Smooth Speech protocol. Smooth Speech technique, among other strategies, teaches stutterers to consciously control air-flow during speech through continuous airflow, gentle onsets for vowels, and soft contacts with consonants and to synchronize speech phrasing structure with the breath stream. In our case study Smooth Speech was shown to have a noticeable impact on the intelligibility and fluency of the speaker's speech immediately following training. This study is a preliminary investigation to understand the phonetic features, both segmental and supra-segmental, that were specifically effective in increasing speech intelligibility. Speech samples included a monolog a read passage for both pre-test and post-test sessions, and a monolog for the training period. Acoustic analysis included evaluation of phoneme duration, syllable duration, pause durations, and pitch range. Phrasing structure was also analyzed. We conclude with a discussion on the relationship of speech rate and breath control and its effect on intelligibility.

1pSC5. Classifying place of articulation and palatalization of fricatives based on cepstral coefficients and spectral moments. Laura Spinu, Jason Lilley, Irene Vogel (Dept. of Linguist. and Cognit. Sci., Univ. of Delaware, 46 E Delaware Ave., Newark, DE 19716), and H. Timothy Bunnell (Alfred I. duPont Hospital for Children, Wilmington, DE 19803)

Hidden Markov models (HMMs) with two types of acoustic features were used to classify fricatives by place of articulation and palatalization status. The data were recordings of 31 native speakers of Romanian who produced a total of 3674 fricatives. Segments from four places (labial, alveolar, postalveolar, and dorsal) were examined, each of which appeared as plain and palatalized (with a palatal secondary articulation). Both the fricatives and the preceding vowels were divided by HMM training into three acoustically uniform regions, corresponding to the three states of the HMM models. Separate sets of monophone HMMs were trained using (a) the first four spectral moments plus rms amplitude and (b) the first five Bark-cepstral coefficients. Generally, the first and second regions/states of the fricatives were more important in classifying the segments by place, while the third state contributed more to the classification by palatalization status. The success of the classification depended on the specific combination of predictor variables (acoustic features and regions) used. Thus, the overall accuracy in classifying segments according to place or palatalization ranged from 62% to 95%. These analyses shed light on the differential distribution over time of acoustic features related to place and palatalization.

1pSC6. Acoustic cues to the voicing contrast in developing coda stop consonants: Toward comparison with adult caregivers. Helen M. Hanson (ECE Dept., Union College, 807 Union St., Schenectady, NY 12308), Stefanie Shattuck-Hufnagel (MIT RLE, Cambridge, MA 02138), and Margaret Capotosto (Union College, Schenectady, NY 12308)

Hanson and Shattuck-Hufnagel (Proceedings of Acoustics '08) analyzed coda stop consonants for three children from the Imbrie corpus (A. Imbrie, Ph.D. thesis, MIT, 2005). Following the closure, voice bars were frequently observed in [+voice] stops, while [−voice] codas often had a period of noise overlapping the preceding vowel and extending into the closure. Possible explanations for this noise include motoric immaturity, imitation of something in their caregivers' productions, or the child's decision to provide additional cues to the voicing contrast (enhancement). In this study, the second possibility was investigated in recordings of each child's primary caregiver. Coda stop consonants were analyzed in the words "bug," "tub," "cup," and "duck," specifically the durations of voice bars and vowel-final noise. As expected, voice bars mainly occur in voiced codas; in contrast, vowel-final noise occurs mainly with voiceless codas, and its duration and degree of overlap with the vowel/closure vary across subjects. Similar measures for the children will be compared with those of their primary caregivers. [Work supported by NIH Grants Nos. DC00075 and DC 008780, and a NSF CPATH grant.]

1pSC7. Transitional cues in fricative noise in Greek /s/-stop and stop-/s/ sequences: Children versus adults. Asimina Syrika, Jan R. Edwards (Dept. of Communicative Disord., Univ. of Wisconsin-Madison, 1975 Willow Dr., Madison, WI 53706, syrika@wisc.edu), Fangfang Li (Univ. of Lethbridge, Lethbridge, AB T1K 3M4, Canada), Mary E. Beckman (Ohio State Univ., Columbus, OH 43210-1298), and Benjamin Munson (Univ. of Minnesota, Minneapolis, MN 55455)

Greek is one of the few languages to allow /s/-stop and stop-/s/ sequences at three places of articulation (bilabial, dental, and velar) in word-initial position. This study measured the coarticulation of /s/ with the following or preceding stop in native Greek-speaking children's and adults' productions of real words beginning with /sp/, /st/, /sk/, /ps/, /ts/, and /ks/, in a variety of vowel contexts. The aim was to determine whether stop place cues are signaled effectively in the spectrum of adjacent /s/ in both types of consonant sequences attested in Greek. Fast Fourier transform spectra were calculated for overlapping 10-ms windows from the beginning to the end of fricative noise and spectral moments were computed in each window. Systematic differences in the fricative spectra as a function of the adjacent stop consonant were observed in the /s/+stop sequence (as in English) and also in the stop+/s/ sequence (where the differences in the portion after the stop burst proper mirrored the patterns at the end of the fricative in the clusters in the other order). Coarticulation was comparable across age groups, although there was greater variability in children's productions. [Work supported by NIDCD Grant No. R01DC02932 and NSF Grant No. BCS-0729140.]

1pSC8. Phonetic realization of glottal stop in Shugni. Mingzhen Bao (Dept. of Modern and Classical Lang., Univ. of Kentucky, 1055 Patterson Office Tower, Lexington, KY 40506, mingzhenbao@uky.edu)

Shugni is an underdocumented Indo-European language, mainly spoken in the Pamir mountains. Orthographically, Shugni used the apostrophe to symbolize the glottal stop phoneme, and the symbol only occurred word initially. In this paper, phonetic manifestation of glottal stop was examined, and the adjacent vowel [a] was compared under five conditions: (1) word-initial glottal stop followed by [a] and a consonant #[ʔaC1], (2) word-medial [C2aC1] sequence where [a] was also syllable-medial, (3) word-medial [C2aC1] sequence where [a] was syllable-final, (4) word-final [a] preceded by a consonant [C2a]#, and (5) a sequence of [C2a]# [ʔaC1] where glottal stop was word-initial and intervocalic. Four acoustic parameters were measured, and glottal stop was defined as (1) an absence of formant transitions entering or leaving [a], (2) an abrupt start or end of [a]'s magnitude, (3) identifiable stop closure, and (4) irregular pulses produced by glottal tension. The results showed that irregular pulses were noticed in some but not all [a]s of words which began with the apostrophe. No other significant acoustic differences were noted among conditions. The results suggested that glottal stop in Shugni be described as non-phonemic vowel laryngealization or creaky phonation. The word-initial apostrophe had to be treated cautiously to interpret the variation.

1pSC9. Investigating the incomplete neutralization of flaps in North American English. Michael Colley (Dept. of Linguist., MS 23, Rice Univ., 6100 Main St., Houston, TX 77005, michaeltcolley@yahoo.com)

The flapping of intervocalic apical stops in American English results in the neutralization of voicing in words like writing and riding, as compared to write and ride. Some evidence has suggested that speakers maintain a distinction in the length of the vowel preceding the flaps, thus making the neutralization incomplete [e.g., Patterson and Connine (2001)]. Proponents of the idea of incomplete neutralizations often use it to challenge the idea of discreteness of phonetic categories [Port (1996)]. However, claims of incomplete neutralizations in English and other languages have not been universally accepted [Manaster Ramer (1996)]. This study tests whether listeners are able to use differences in vowel durations before flaps to determine the underlying voicing of flaps in pseudo-words. To avoid any influence from orthography, the pseudo-words were presented as auditory stimuli to the participants, and the responses were recorded orally. Participants who heard stimuli in which one set of vowels were 20 ms longer than the other set were more likely to interpret the flaps as /t/ than participants who heard stimuli in which all pre-flap vowels had identical durations. However, participants were not able to use the differences in vowel durations to determine the underlying voicing of the flaps.

1pSC10. Nasal coarticulation in clear speech. Anthony Brasher (Dept. of Linguist., Univ. of Michigan, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48109-1220)

This study tests whether speakers, when trying to speak clearly, employ variable enhancement strategies as a function of phonetic environment. Using aerodynamic and acoustical methods, this study examines the effects of

phonemic context and speaking modality and on the spatial and temporal extent of anticipatory nasal coarticulation in English. Target words are English (C)VNCvoiced (e.g., *bend*) and (C)VNCvoiceless (e.g., *bent*) words spoken in either clear or citation speech modes. In order to enhance the percept of /n/ in clear speech, speakers increase the duration of the nasal consonant in CVNCvoiced words but marginally increase, or even decrease, /n/ duration in CVNCvoiceless words. While highly variable, airflow results suggest little difference on anticipatory nasalization as a function of speech mode. These results argue against models predicting a global reduction in coarticulation in clear speech.

1pSC11. Tone-glottal coarticulation in a complex tone language. Christian T. DiCanio (CNRS-Dynamique du Langage, UMR 5596, Université Lyon-2, 14 Ave. Berthelot, Lyon 69363, France, cdicanio@gmail.com)

A majority of the research on coarticulation has focused on consonant and vowel timing, but work on the coarticulation of tone or phonation-type contrasts has received relatively less attention until recently [Brunelle (2003); Miller (2007); Peng (1997); Silverman (1997); Xu (1994); (1997); (1998); (1999); (2004); (2005); Xu *et al.* (2006); Xu and Liu (2006); Xu and Sun (2002); Xu and Wang (2001)]. The present work investigates the coarticulatory timing relationships between the nine tonal contrasts and three laryngeal contrasts (creak, glottal closure, and breathiness) of Itunyoso Trique [DiCanio (2008)]. Acoustic and electroglottographic (EGG) data from Trique were collected, comparing the realization of the tones in the different laryngeal conditions. While creaky phonation induces an equivalent pitch lowering for a variety of tones, breathy phonation induces pitch lowering in an asymmetrical fashion where higher tones have a greater pitch decrease than lower tones. These particular effects are not predictable from a perceptually-driven model of coarticulation, but accord well with the body-cover model of vocal-fold vibration [Titze (1994); Story & Titze (2003)]. The timing of non-modal phonation in Itunyoso Trique reflects a language-specific coarticulation strategy, but the observed pitch perturbation effects are best-explained with a speaker-oriented model of coarticulation.

1pSC12. Lexical tone and segmental coarticulation. Augustine Agwuele (Dept. of Anthropology, Texas State Univ. San Marcos, San Marcos, TX 78666)

In studying anticipatory coarticulation in emphatic CV sequences [Lindblom *et al.*, J. Acoust. Soc. Am. **121**, 3802–3813 (2007)] and in fast speech [Agwuele *et al.*, Phonetica **68**, 194–209 (2008)], it was shown that prosodically induced modulations of speech introduce additional sources of phonetic variation separate from those generated by vowel context. Both studies were successful in dissociating these additional prosodic effects from the underlying contextual vowel variation and quantifying them. The focus of this current investigation was to study CV coarticulation, using a similar set of regression metrics, but now applied to a tonal language (Yoruba). It is to be determined whether lexical tone impacts the coarticulation of consonant and vowel. Specifically, words, do tones, produce additional effects on anticipatory CV coarticulation separate from those generated by vowel-context similar to emphatic stress and increased tempo as documented by the aforementioned studies? Finally, of interest also is to examine whether different lexical tones impact C+V coarticulation differently.

1pSC13. Vowel normalization using stop consonant loci. Christian Koops (Dept. of Linguist., Rice Univ., 6100 S. Main St., Houston, TX 77005, ckoops@rice.edu)

Vowel normalization techniques used in sociophonetics aim to eliminate the influence of vocal tract length differences on resonant frequencies so that speaker-specific vowel configurations can be compared more precisely. In vowel-extrinsic methods, a speaker's F1-F2 range is scaled up or down, and vowel spaces are aligned using the geometrical F1-F2 mean. However, a speaker's F1-F2 mean may be systematically biased if his or her vowel configuration is affected by a set of parallel vowel shifts, e.g., fronting or raising of several vowels in a chain shift. To address this problem, an alternative method of deriving speaker-specific scaling factors was tested based on a vowel-independent acoustic measure: the locus (virtual origin) of the second formant transition following the /d/-release. Forty adult speakers of American English (20 male, 20 female) were recorded reading ten repetitions of 16 /dVd/-words. From these data, individual F1-F2 means and locus equations were calculated, including each speaker's F2 locus frequency for /d/. Preliminary results based on the corner vowels /i, a, u/ show that F2 locus and F1-F2 mean are strongly correlated (r-squared=0.84). Statistical modeling is underway to determine whether scaling factors derived from locus analysis are superior to scaling factors derived from F1-F2 means.

1pSC14. Fine-grained control of voice-onset time production, lexical usage-frequency, and phrasal context. Mark VanDam (Boys Town Natl Res. Hospital, 555 N 30 St., Omaha, NE 68131, vanmarj@boystown.org), Noah H. Silbert, and Robert F. Port (Indiana Univ., Bloomington, IN 47405)

Voice-onset time (VOT) in English maps onto the phonological categories voiced or voiceless, but fine-grained control of VOT production may vary as a function of additional linguistic features (e.g., lexical frequency, phrasal context, indexical features). This study investigated the patterns of VOT production in words that varied (a) by lexical usage-frequency (high versus low) and (b) in selected linguistic contexts (isolation, carrier phrase, unfamiliar phrase, familiar phrase). Results show that talkers produced longer VOT for voiceless stops in low-frequency words in unfamiliar and familiar phrases but not in isolation or carrier sentences. Neither lexical frequency nor phrasal context seemed to induce changes in VOT in voiced stops. Thus, it appears that talkers make use of fine temporal distinctions within the voiceless category but not the voiced. This suggests a complex asymmetry in how continuously varying VOT maps onto more discretely varying phonological categories.

1pSC15. Gestural drift in Serbian-English speakers. Stephen Tobin (Dept. of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs, CT 06269-1020, stephen.tobin@uconn.edu)

Following Tobin [J. Acoust. Soc. Am. **125**, 2757 (2009)], perceptually guided changes in speech production were examined in a group of Serbian-English speakers. Participants were recorded producing Serbian and English sentences after periods of time in the US and in Serbia. Voice onset times (VOTs) and vowel formants were examined with the aim of determining whether gestural targets as well as patterns of gestural timing are affected by exposure to a language. It is predicted that in both Serbian and English, VOTs will be longer after time in the US and shorter after time in Serbia. Based on comparisons of available American English and Serbian vowel formant data, it is expected that F1's will increase and F2's will decrease after time in Serbia while F1's will decrease and F2's will increase after time in the US. The results will be compared with those of Tobin's Spanish-English speakers and discussed with relation to recent research on phonetic convergence, attunement, and articulatory setting.

1pSC16. Acoustic correlates of devoiced Japanese vowels. J. Kevin Varden (Ctr. for Liberal Arts, Meiji Gakuin Univ., 1518 Kamikurata-cho, Totsuka, Yokohama 244-8539, Japan, varden@gen.meijigakuin.ac.jp)

The analysis of acoustic correlates of devoiced vowels produced by young native Japanese speakers is presented. In addition to gradient amounts of voicing seen in vowels undergoing devoicing, gradient amounts of fricativization can be observed, supporting an analysis of devoicing in Japanese as gestural overlap of both glottal gesture and oral stricture. Specifically, the surfacing of voiceless consonant-devoiced vowel sites as voiceless fricatives, evidenced by high-frequency frication in spectrograms and a significantly higher number of zero crossings, will be discussed, as will the interplay between devoicing and fricativization in the data set. Analysis of the formant frequency information at the vowel sites, preserved in the frication and allowing the perception of the original vowels even in the face of heavy loss of vocality, is also presented.

1pSC17. Preliminary report of devoiced vowels in Chinese. Setsuko Shirai, Li-Chin Yang, and Yuda Lai (Univ. of MingDao, 369 Wen-Hua Rd., Peetow, ChangHua, Taiwan, 52345 Republic of China)

This is the preliminary report of the devoiced vowels in Chinese spoken in Taiwan. We are currently investigating the environments of devoiced vowels. We recorded the sentences read by ten Taiwanese who were born and grew up in the center of Taiwan. The sentences ended with high vowels ([i] or [u]) following aspirated consonants. The tones of these final vowels varied. We observed no voice marks after the final consonants in some sentences. In addition, some vowels consist of a couple of waves. The observed devoiced vowels are as follows: [i] in [xuechi] "wiping out of dis-

grace," [cuopi] "turf," [dajuchi] "big molar tooth," and [ruci] "in this way," and [u] in [dongdoufu] "dried tofu." Although these devoiced vowels were observed, not all speakers produced these devoiced vowels. The highly devoiced vowel was [i] in [xuechi] "wiping out of disgrace," whose final vowel [i] four out of ten speakers devoiced. In addition, the vowels produced by other two speakers consist of only two waves. Furthermore, we observed a very short vowel in the middle of the sentence (the first [u] in [tuputaopi] "spit out the skins of grapes").

1pSC18. Articulatory variability within experimental sessions. Anders Lofqvist (Dept. Logopedics, Phoniatrics and Audiol., Lund Univ., SE 221 85 Lund, Sweden, lofquist@haskins.yale.edu)

This paper examines articulatory and acoustic variability within recording sessions lasting about 25 min, with particular emphasis on overall drift during the session. Articulatory movements were recorded using a magnetometer system. The subjects were five speakers of Japanese, producing words with long and short consonants. Fifty tokens of several utterance types were obtained. For this analysis, ten tokens were binned in five different bins according to where they occurred during the recording session. Statistical analysis showed that for the five speakers, differences between the bins occurred for between 30% and 60% of the utterances and measurements. Most differences occurred for acoustic measures of segment duration and for a kinematic measure of tongue movement duration. During the course of the recording session, segment durations mostly decreased, but the patterns differed both within and across speakers. These results are useful for planning experiments, and for Power analysis of the number of tokens required for statistical analysis. [Work supported by NIH and the Swedish Research Council.]

1pSC19. Acoustic and ultrasound investigation of English vowel-to-vowel coarticulation. Jonathan C. Yip and Patrice S. Beddor (Dept. of Linguist., Univ. of Michigan, 440 Lorch Hall, 611 Tappan St., Ann Arbor, MI 48104, jonyip@umich.edu)

This study examined the hypothesis of whether vowel-to-vowel coarticulation effects are stronger when a particular vowel segment is under focus in clarified speech than when it is not under a condition of focus. In one condition, native speakers of American English produced disyllabic word pairs that varied in the quality of their first vowel (stressed V1), while in the second condition, the same group of speakers produced the word pairs that varied in the quality of their second vowel (V2). The degree of coarticulation was determined by the amount of spectral and/or articulatory displacement that occurred during the production of V2-position [schwa] due to the influence of V1. Due to the need to increase the likelihood of coarticulatory effects during speakers' productions, recording trials were sped up to a conversational speaking rate. Spectral measurements (first and second formants) of ə in both conditions were taken, as well as tongue-surface position measurements pertaining to the position of tongue body during the production of ə under the two conditions. Preliminary acoustic analysis indicates that speakers produced V2-position ə with greater coarticulatory effects in the V1-clarifying condition than in the V2-clarifying condition.

MONDAY AFTERNOON, 26 OCTOBER 2009 RIO GRANDE WEST, 1:00 TO 3:00 P.M.

Session 1pSPa

Signal Processing in Acoustics: Underwater Channel and Source Characterization

Natalia A. Sidorovskaia, Chair
Univ. of Louisiana at Lafayette, Dept. of Physics, P.O. Box 44210, Lafayette, LA 70504-4210

Contributed Papers

1:00

1pSPa1. Physically constrained maximum-likelihood mode filtering and its application as a pre-processing method for underwater acoustic communication. Joseph C. Papp, James C. Preisig, and Andrey K. Morozov (Woods Hole Oceanograph. Inst., MS 44, Woods Hole, MA 02543, jpapp @whoi.edu)

Mode filtering is most commonly implemented using the sampled mode shapes or pseudo-inverse algorithms. [Buck *et al.* (1998)] placed these techniques in the context of a broader maximum *a posteriori* (MAP) framework. However, the MAP algorithm requires that the signal and noise statistics be known *a priori*. Adaptive array processing algorithms are candidates for improving performance without the need for *a priori* signal and noise statistics. A variant of the physically constrained, maximum-likelihood (PCML) algorithm [Kraay and Baggeroer (2007)] is developed for mode filtering that achieves the same performance as the MAP mode filter yet does not need *a priori* knowledge of the signal and noise statistics. The central innovation of this adaptive mode filter is that the received signal's sample covariance matrix, as estimated by the algorithm, is constrained to be that which can be physically realized given a modal propagation model and an appropriate noise model. Simulation and data processing results from the Shallow Water 2006 experiment are presented along with the application of the mode filter as a broadband pre-processor for multichannel underwater acoustic communication systems. [This work was supported by ONR through ONR Grant Nos. N00014-05-10085 and N00014-06-10788.]

1:15

1pSPa2. Minimum Hellinger distance classification of passive underwater acoustic signals. Brett E. Bissinger (Elec. Eng. Dept., PA Box 30, State College, PA 16804, beb194@psu.edu), R. Lee Culver, and N.K. Bose (Penn State Univ., State College, PA 16804)

Passive source classification in the underwater environment is a challenging problem in part because propagation through the space- and time-varying media introduces variability and uncertainty in the signal. Acoustic propagation codes can predict received fields accurately but they are sensitive to input environmental parameters which cannot be known exactly. This uncertainty in environmental knowledge used in signal predictions results in imperfect statistical class models. Classifiers that rely on simulations of the environment must therefore be robust to imperfect statistical models. Maximum likelihood methods provide ideal performance when the class models are correct but their performance quickly deteriorates when class models are imperfect. Minimum distance methods generally offer robustness to mismatches at the expense of performance, with that trade-off governed by the distance metric used. Several well-studied distance metrics are discussed and it is explained how the Hellinger distance offers robustness to outliers while retaining the performance of a maximum likelihood method, properties that make it well-suited for classification of passive underwater acoustic signals. Examples are provided to quantify the robustness properties of the Hellinger distance when applied to classification of passive underwater acoustic signals and compare its performance to the maximum likelihood method. [Work supported by ONR Undersea Signal Processing.]

1:30

1pSPa3. Underwater channel estimation and tracking using mixed norm optimization techniques. Ananya Sen Gupta and James Preisig (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Insitution, 98 Water St., Woods Hole, MA 02543)

A mixed norm optimization technique to track the impulse response of the underwater acoustic channel is presented. The channel is assumed to follow a delay-Doppler spread function model, which is known to be sparsely distributed over the complex channel coefficients. The ill-conditioned nature of this estimation problem, along with the need to optimize over a complex field, poses challenges to employing popular sparse estimation techniques that currently exist in the literature. The time-varying nature of the underwater channel also makes efficient tracking difficult. The motivation behind our proposed algorithm is to robustly estimate and efficiently track the sparse delay-Doppler coefficients directly over the complex field. We have formulated the channel estimation problem as that of minimizing over a convex combination of the L1 norm of the sparse coefficients to be estimated and the L2 norm of the estimation error. We analyze the difference between our proposed algorithm to existing L2-constrained L1 optimization techniques and present results showing performance on simulation generated and field data. [Work supported in part by ONR Grant Nos. N00014-05-10085 and N00014-07-10184 and in part by Postdoctoral Scholar Program at the Woods Hole Oceanographic Institution, with funding provided by the Doherty Foundation.]

1:45

1pSPa4. High-resolution time delay estimation using a combined correlator-based and phase measurement approaches. Said Assous, Clare Hopper, Mike Lovell (Dept. of Geology, Univ. of Leicester, Leicester LE1 7RH, United Kingdom, sa251@le.ac.uk), David Gunn, Peter Jackson, and John Rees (Br. Geological Survey, Nottingham NG12 5GG, United Kingdom)

Time delay estimation is a fundamental step in source localization or beamforming applications. It has attracted considerable research attention over the past few decades in many fields including radar, sonar, seismology, geophysics, and ultrasonics. Various techniques are reported in the literature, mainly the correlator-based methods. However, with these approaches, when the time delay is not an integral multiple of the sampling rate, it is necessary to either increase the sampling rate or use interpolation to get better resolution. In this paper, we present a new approach for time delay estimation based on combining the correlator-based technique and the phase information obtained from the received signal. A cross-correlation based approach is applied first to get an initial estimate then the DFT is applied to a window of the signal starting at this initial estimate to calculate a related phase. This phase will define an additional fractional inter-sample time delay, which is corrected to provide more accurate estimation. Simulated examples are given with synthetic signals. An underwater application for distance and speed of sound measurements is then demonstrated. Finally, the proposed method is shown to outperform the correlator-based approach and a high resolution is achievable without the need of oversampling or interpolation.

2:00

1pSPa5. Underwater speed of sound and ranging measurements using the sliding discrete Fourier transform. Said Assous, Clare Hopper, Mike Lovell (Dept. of Geology, Univ. of Leicester, Leicester LE1 7RH, United Kingdom, sa251@le.ac.uk), Laurie Linnett (Fortkey Ltd., Edinburgh EH1 2JL, United Kingdom), David Gunn, Peter Jackson, and John Rees (Br. Geological Survey, Nottingham NG12 5GG, United Kingdom)

It is often desirable to extract parametric information from a signal in order to determine its properties with high accuracy. Such parameters might be the time delay, the amplitude, the frequency, and/or the phase. These parameters are usually highly correlated with the material through which the signal has passed or from which it has been reflected. Time delay estimation is a fundamental step in source localization or beamforming applications, and has attracted considerable research attention over the past few decades in many fields. In this paper, we propose a novel time delay estimation approach based on the sliding discrete Fourier transform (DFT). The analysis window is slid sample by sample over the received single frequency tone burst signal with the DFT evaluated successively. The optimal peak spectrum and its corresponding phase are used to estimate the time delay of the signal. We use the corresponding time as the first estimate, which is improved based on the related phase. Simulated examples are given with synthetic signals. An underwater application for speed of sound measurements using this approach compared to those obtained by a commercial device is then demonstrated. Finally, the proposed method is shown to outperform standard correlator-based approaches.

2:15

1pSPa6. Noise performance of a novel total least squares approach to source location estimation. Wm. Garth Frazier, Kenneth E. Gilbert, and Xiao Di (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677, frazier@olemiss.edu)

One of the advantages of the time-of-arrival (TOA) representation for the source location problem is that it can be formulated such that the unknown variables appear linearly in the describing equations. It is known that to obtain a unique three-dimensional solution four measurements are required in the TOA case. Typically, more measurements than the minimum are available and this naturally leads to an over-determined problem for which an optimal estimate is sought. One method of solution is weighted least-squares (LS). The noisy measurements include the TOAs of the impulse at each acoustic receiver as well as the locations of the receivers. In the linear formulation these measurement values appear on both sides of the equation $Ax=b$, and not just on the right side. Application of a linear total least-squared (TLS) error method permits representation of the uncertainty on both sides of the equation. Moreover, the additional computations required in the TLS approach are not significantly greater than the LS approach. In this paper, we examine the noise performance of TLS solutions and demonstrate the significant reduction in bias obtained over the LS approach at the expensive increased sensitivity. Results from simulation and field experiments are provided.

2:30

1pSPa7. Active clutter reduction through fusion with passive sonar data. Jason M. Aughenbaugh, Bryan Yocom, Brian R. La Cour, and Tom Yudichak (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029 Austin, TX 78713-8029, jason@arlut.utexas.edu)

Active and passive sonar systems provide different types and qualities of data. For example, passive sonar typically provides bearing information and classification clues. Active systems can provide the complementary range and Doppler shift information. Similarly, the types of false alarms—passive interferes and active clutter—may be from different objects. The combination of information from both systems should yield a better estimate of the target state, but traditionally, such systems are used separately. A method for fusing active and passive sonar data is presented with example results. The fusion engine is a Bayesian tracker. The key step for fusing active and passive data is the formation of the likelihood function for each type of measurement. This process is described at a high level, but the emphasis on the results of fusing the data. The example problem involves simulated active and passive data, including the presence of passive interferers and active clutter objects. The results show the value of using active systems to refine passive estimates and more importantly suggest the potential power of using passive data to mitigate active clutter. [This work was supported by the Office of Naval Research Contract No. N00014-06-G-0218-01.]

2:45

1pSPa8. Statistical characterization of bistatic acoustic clutter from a shipwreck in the 0.8–3.5-kHz region. John, R. Preston (ARL, Penn State Univ., P.O. Box 30, State College, PA 16804, preston@ciao.arl.psu.edu)

Sonar clutter is defined as target like echoes that are detected and tracked by automated signal processors. Sonar clutter is one of the primary limitations to active ASW. Some statistical analysis of clutter from bistatic measurements will be presented. The received data are taken from the Five Octave Research Array that has been used to collect extensive monostatic and bistatic data in two recent sea trials on the Malta Plateau off Sicily called Clutter 07 and BASE 07. Statistical characterizations of bistatic clutter from

and near a shipwreck are presented. Related work by D. Abraham focused on monostatic characterizations of clutter. This work uses data from the above mentioned sea trials to characterize non-Rayleigh behavior of bistatic clutter from pulsed sources and to compare bistatic clutter with monostatic clutter. K-distributions with their shape and scale parameters are used to describe non-Rayleigh behavior together with the models of Abraham and Lyons to infer scatterer densities. The ability to geo-reference key statistical measures of clutter near the shipwreck that can allow CFAR processors to properly set thresholds and reduce false alarms is demonstrated. Also included are presentations of the shape parameter versus bistatic aspect angle. [Work supported by ONR code 321US.]

Session 1pSPb

Signal Processing in Acoustics: Sensor Optimization Techniques

R. Lee Culver, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Contributed Papers

3:30

1pSPb1. Cost efficient sensor placement and configuration to fulfill spatially inhomogeneous coverage preferences. Sergey N. Vecherin, D. Keith Wilson (Signature Phys. Branch, U. S. Army ERDC/CRREL, 72 Lyme Rd., Hanover, NH 03755, sergey.n.vecherin@usace.army.mil), and Chris L. Pettit (U. S. Naval Acad., Annapolis, MD 21402)

Finding a most cost efficient sensor network to fulfill specified coverage requirements is an important practical problem. Current research presents a strict formulation of this problem as the binary linear programming problem. The solution addresses the questions of how many, what type, and where to place sensors to fulfill arbitrarily complicated coverage preferences with minimal cost. Cost can reflect different sensor attributes, e.g., actual sensor values, their number, or accessibility of the terrain in which they are placed. Sensor performance is evaluated in terms of a detection probability and is applicable to any type of sensor, including acoustic. Geometrical spreading and attenuation of a signal from its source to a sensor are taken into account when calculating detection probability. For large spatial grids of candidate sensor locations, it is infeasible to determine the strict, optimal solution. To overcome this problem, a fast algorithm is presented for finding a suboptimal, but highly satisfactory, solution. The algorithm accommodates many practical needs, such as sensor availability, relocation of existing sensors to improve their coverage, and sensor-to-sensor communication. The efficiency of the algorithm is demonstrated through a comparison with another heuristic sensor placement strategy that places sensor one-by-one in the previously worst covered spots.

3:45

1pSPb2. Using wireless networks for real-world acoustic measurement applications. Kurt Veggeberg (Sound and Vib. Segment, Natl. Instruments, 11500 N. Mopac C, Austin, TX 78759, kurt.veggeberg@ni.com)

Wireless technology extends the concept of personal computer based data acquisition beyond the limits of cables and wired infrastructure for new remote or distributed measurement applications. This is an overview of wireless networking basics and how to deploy reliable wireless measurement in a variety of outdoor or harsh environments for reliable and secure data acquisition systems. Examples of distributed outdoor noise monitoring systems will illustrate networking layers and topologies for specific applications using IEEE 802.11 networks for reliable and secure wireless data acquisition systems. These include a distributed, synchronized data acquisition system for the extended measurement of sonic booms and a portable noise and vibration management system with multiple front ends connected wirelessly throughout a construction site to be compliant with local regulations and contractual obligations for noise and vibration levels.

4:00

1pSPb3. Research on the applications of time reversal method for noise control by use of micro-electro-mechanical-system array microphones for cell phones. Yu-Hao Hsieh, Sheng-Che Lin, Bo-Hsien Wu, Gee-Pinn Too (Dept. of Systems and Naval Mechatronic Eng., Natl. Cheng Kung Univ., No. 1 University Rd., Tainan 70101, Taiwan), and Shyang-Jye Chang (ITRIS, Tainan 70955, Taiwan)

The purpose of this research is to construct a noise reduction system in 3G cell phones that combines time reversal method (TRM) and the micro-electro-mechanical-system (MEMS) microphone array. Time reversal method has advantages of compensate distortion due to path effect in propagation and signals focus at the original location. This research contains theoretical reasoning, TRM simulations, and experiments. The main point of this study is to estimate the impulse response function accurately to form a virtual path model, because desired signal needs to be obtained at listener's location, not source location. The simulations present that the procedure is effective under different environmental conditions and parameters. Both normalized cross-correlation coefficient (NCCC) and signal-to-noise ratio (SNR) are used to estimate the simulation result. In the experiments, a MEMS microphone array is constructed and simulates the configuration as it is on the cell phone. The results demonstrate that the procedure can separate a specific source from a combination signal of multiple sources, and increase the NCCC and SNR of the specific source signal. Finally, the enhancement level of the result is discussed. [This study is supported by Center of Nanotechnology and Microsystems Engineering, National Cheng Kung University, with the Project no. 98C066.]

4:15

1pSPb4. Improving mp3 capability of mobile phones by linking acoustic information with vibrations. Cheol Hong Kim (Chonnam Natl. Univ., Dept. of Electron. and Comp. Eng. 300 Yongbong-Dong, Buk-Gu Gwangju 500-757 KR, chkim22@chonnam.ac.kr), Young Jin Park, and Jong-Myon Kim (Univ. of Ulsan, Korea)

Consumers around the world have strong demand for mobile devices such as mobile phones and mp3 players. For this reason, the mp3-playing mobile phone has become one of the most popular mobile devices, especially for the younger generation. In this study, we propose an efficient method of utilizing small vibration motors embedded in the mobile phones to improve mp3 capability. The small vibration motors are generally embedded in the mobile phones to implement tangible user interface. The proposed method, which links acoustic information with areas on the mobile phone surface, utilizes the existing embedded small vibration motors to enhance mp3 capability. The use of additional small vibration motors can maximize mp3 capability but can cause additional power consumption and heat dissipation in the mobile phones. Therefore, the proposed method utilizes the pre-embedded vibration motors to generate vibrations according to the

acoustic information from mp3. Experimental results, performed by using up-to-date mobile phone platform over 50 persons, demonstrate that the proposed method is suitable for mp3-playing mobile phones in case that the users want to feel vibrations with mp3. [Work supported by the MKE, Korea, under the ITRC supervised by the IITA [IITA-2009-(C1090-0903-0008)].]

MONDAY AFTERNOON, 26 OCTOBER 2009 RIO GRANDE CENTER, 1:00 TO 4:15 P.M.

Session 1pUW

Underwater Acoustics: Scattering From Targets in the Ocean

Philip L. Marston, Cochair
Washington State Univ., Physics and Astronomy Dept., Pullman, WA 99164

Jon La Follet, Cochair
Washington State Univ., Physics and Astronomy Dept., Pullman, WA 99164

Contributed Papers

1:00

1pUW1. Inversion for material, geometrical, positional, and radial speed parameters of a spherical target in a waveguide. John A. Fawcett (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, john.fawcett@drdc-rddc.gc.ca)

In this paper, a moving spherical target in a Pekeris waveguide is considered. The classification of a target from its free space echo has been previously studied by many authors. In a waveguide, the received echo results from the coherent interference of the various multipaths incident upon and scattered from the target. When the position of the target in the waveguide is imprecisely known, the echo classification problem is more difficult . In addition, if the target is moving with an unknown radial speed, this further complicates the problem. In this paper, we consider a single omnidirectional transmitter and a single receiver and investigate the estimation of the material, geometrical, positional, and dynamical parameters of the sphere from the received echo. We use a simple simulated annealing approach to investigate the parameter estimation. We also consider various two-dimensional cost-function surfaces to illustrate the uncertainties in the parameter values.

1:15

1pUW2. Backscattering by a solid aluminum cylinder near a boundary: Experiment and finite element model comparisons. Jon La Follett and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, lafollej@mail.wsu.edu)

Modeling can be a useful tool for understanding the mechanisms that affect target backscattering. Comparing model predictions with experimental data serves to build credibility for the modeling approach and can give an estimation of accuracy. Measurements were made of backscattering by a solid aluminum cylinder, having a length to diameter ratio of 2. The target was suspended just below the surface of a water tank to simulate some aspects of backscattering when resting on the seabed. The target was rotated in a plane parallel to the free surface in both the free field and proud target configurations to obtain the spectral response (color) as a function of tilt. These measurements were compared with predictions made by a previously developed approximate finite element method [Williams *et al.*, J. Acoust. Soc. Am. **125**, 2701 (2009)] designed to model axisymmetric targets. Model predictions of acoustic color were able to accurately produce many of the bright features present in the tank data. However, for the proud configuration at broadside and larger values of wavenumber radius product, *ka* greater than 6, there are some discrepancies. [Research supported by ONR.]

1:30

1pUW3. Bistatic line-scan holography of scattering by metallic objects near a free surface. Jon La Follett and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, lafollej@mail.wsu.edu)

Bistatic scattering measurements were performed with metallic targets in water located near a free surface to simulate aspects of scattering by a proud target resting on ocean sediment. A water filled cylindrical aluminum shell and a solid aluminum cylinder were studied, both having a length to diameter ratio of 2. In addition a solid 45 deg aluminum cone was studied. Previously, an acoustic holographic method [Baik *et al.*, J. Acoust. Soc. Am. **121**, 3203 (2007)] was applied to bistatic scattering measurements where the source, free field target, and receiver scan line were all contained in a plane. That method was shown to be helpful for identifying contributions to the scattering. In the present work, this method is modified to analyze bistatic scattering by proud targets where both the source and receiver are at a grazing angles with respect to the free surface. The holographic images are compared with bistatic synthetic aperture sonar images for certain targets. Some of the image features were caused by the proximity of the target to the surface. [Work supported by the Office of Naval Research.]

1:45

1pUW4. Evolution of navy diver exposure standards for deterministic underwater sound in the 100–500 Hz band. Frederick M. Pestorius (Appl. Res. Labs., Univ. of Texas at Austin, 10000 Burent Rd., Austin, TX 78758, pestoriu@arlut.utexas.edu), David Fothergill, and Edward Cudahy (Naval Submarine Medical Res. Lab., Naval Submarine Base, Groton, CT 06349)

The US Navy developed low-frequency active (LFA) sonar that had the potential for adverse exposure of divers to high levels of LFA sound. A number of studies into the effects on divers of deterministic waveforms in the frequency band 100–500 Hz were conducted. Animal, human diver, and modeling underwater sound exposure studies were completed across a broad array of organ systems. Human diver studies have investigated underwater hearing thresholds, aversion, wet suit attenuation, and psychophysical response. Tests have included both laboratory and field studies at sound pressure levels up to 160 dB *re* 1 μPa. Ultimately, over 500 exposures using over 100 different test subjects were conducted. These tests yielded a baseline set of guidelines for the exposure of Navy and recreational divers to deterministic sonar transmissions in the subject test band. A summary of the initial and subsequent testing and a tracing of the evolution of the current guidelines will be presented.

2:00

1pUW5. Detection and localization of a target in shallow water between two source-receive ultrasonic arrays: A small-scale experiment demonstration. Christian Marandet, Philippe Roux (LGIT, BP 53, Cedex 9, 1381 rue de la Piscine, 38400 St Martin d'Heres France, christian.marandet@obs.ujf-grenoble.fr), and Barbara Nicolas (Gipsa Lab., BP 46, Cedex 9, 38402 Grenoble, France)

We experimentally demonstrate the detection and localization of a wavelength-size target in a shallow ultrasonic waveguide between two source-receive arrays. The waveguide represents a 1.5-km-long, 50-m-deep ocean acoustic channel at the 1/1000 scale, in which two coplanar arrays would record the broadband transfer matrix around 3 kHz. Invoking reciprocity principle, we perform a time-domain double beamforming algorithm simultaneously on the source and receive array. This array processing projects the multi-reverberated acoustic signals recorded between each source and each receiver into an equivalent set of eigenrays defined by their launch and receive angles. Comparison is made between the intensity of each eigenray with and without the target in the ultrasonic waveguide. When detection is achieved, localization is performed through the back projection of each eigenray into the acoustic waveguide weighted by the intensity fluctuation induced by the target. The use of the diffraction-based sensitivity kernel for each eigenray improves the localization of the target. Detection and localization experimental results will be shown in the presence of surface waves for ultrasonic arrays at 1 and 3 MHz.

2:15

1pUW6. Backscattering and synthetic aperature sonar imaging of partially exposed acoustically penetrable polymer cylinders. Timothy M. Marston and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu)

The scattering by a partially exposed circular cylinder parallel to a flat surface can have a complicated dependence on the amount of exposure h and the grazing angle. Even in the case of high-frequency scattering by a rigid cylinder, geometric analysis must be extended to eliminate abrupt jumps caused by transitions in the number of rays [K. Baik and P. L. Marston, IEEE J. Ocean. Eng. **33**, 386–396 (2008)]. The present investigation concerns transitions in the ray contributions as a function of h for the more complicated case of a penetrable plastic cylinder lowered through a free surface. In this case it is helpful to observe the impulse response as a function of h and the time t. The slopes of trajectories in the (t, h) plane are useful for classifying the scattering mechanisms and the number of reflections from the adjacent flat interface. Some mechanisms (including transmission and refraction through the cylinder) fade with a decreasing h due to a loss of aperture. The evolution of SAS images can also be explained in this way. The investigation is also important for understanding backscattering by cylinders partially covered by sediment. [Work supported by ONR.]

2:30

1pUW7. Burial depth spatial decay rate reduction using bistatic detection: Laboratory demonstration. Timothy M. Marston, Aubrey L. Espana, David B. Thiessen, and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu)

For some situations evanescent waves are important in the detection of objects below the sea floor. The scattering by the object is reduced for increased burial depth and the spatial decay rate of the scattering depends on the grazing angle of the incident wave as well as the grazing angle of the receiver (for the case of bistatic reception). The decay rates are related by extended reciprocity [Cerruti *et al.*, J. Acoust. Soc. Am. **125**, 2732 (2009)]. The present research gives laboratory measurements of bistatic scattering by two types of small targets (a liquid filled shell and an aluminum cylinder) in an experiment using immiscible liquids [Osterhoudt *et al.*, IEEE J. Ocean. Eng. **33**, 397–404 (2008)]. With a receiver position analogous to one in the water column nearly directly above a buried target, as predicted, the measured spatial decay rate of the scattering becomes simply that of the evanescent wave. Design of the experiment and placement of the hydrophone required computational simulation to reduce complications from tank wall reverberations. [Work supported by ONR.]

2:45

1pUW8. Enhanced coupling to target modes from vertical acoustic pressure gradients caused by proximity to an interface and grazing incidence. Aubrey L. Espana, Jon R. La Follett, Timothy M. Marston, and Philip L. Marston (Dept. of Phys. and Astronomy, Washington State Univ., Pullman, WA 99164-2814)

In prior research, it was demonstrated that for end-on illumination of a small horizontal aluminum cylinder, an elastic mode was preferentially excited with evanescent wave illumination in comparison to the case of a propagating wave [Espana *et al.*, J. Acoust. Soc. Am. **125**, 2732 (2009)]. A free-field finite element method simulation revealed that associated with this mode there was a rocking motion at the end of the cylinder. The evanescent wave coupled strongly because of its transverse dependence (along the vertical coordinate) on the excitation. Another way to generate transversely dependent excitation is to place the cylinder parallel to a flat reflecting surface and illuminate it with a propagating wave at grazing incidence. This enhancement of the coupling was demonstrated by placing the cylinder adjacent to a free surface and monitoring the end-on backscattering as a function of the position relative to the interface. The coupling to this mode depends on how the horizontal cylinder is positioned vertically. The response of this mode is also enhanced with the cylinder's axis below the surface near a pressure minimum of the approximately horizontal Lloyd mirror pattern. [Research supported by ONR.]

3:00—3:15 Break

3:15

1pUW9. Impact of target scattering properties on active waveguide invariance. Richard L. Campbell, Jr. and Lisa M. Zurk (NEAR-Lab., Portland State Univ., 1900 SW 4th Ave., Portland, OR 97201, rlcamp@pdx.edu)

The waveguide invariant is a parameter which quantifies time-frequency structure observed in broadband, time-varying sonar geometries. This structure is only weakly dependent on the specific properties of the ocean waveguide, and thus processing schemes utilizing this invariant relationship between range and frequency have the potential to be robust in unknown or poorly known environments. Recent work has extended the concept of waveguide invariance to active sonar, and experimental demonstrations in near-monostatic active geometries using a vertical plate scatterer have been presented (Quijano, 2008). In this study, scaled tank experiments are extended to true bistatic geometries using a variety of canonical scatterers (i.e., spheres, plates, and cylinders), in order to explore effects that are unique to the bistatic scenario. The resulting spectrograms indicate that the target scattering matrix can be the dominant influence on the time-frequency structure. Results from normal mode and raytracing simulations are presented to interpret the tank data and predict the striation structure in realistic ocean environments.

3:30

1pUW10. Acoustic beams with modulated helicity and a simple helicity-selective acoustic reciever. Timothy M. Marston and Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164 2814, marston@wsu.edu)

Acoustic helicoidal beams have an azimuthal phase gradient and an axial null in amplitude. The sign and magnitude of the azimuthal phase gradient determine the helicity and topological charge of the beam. Beams with a unit-magnitude topological charge have been generated with a four-quadrant sectored array, adjacent quadrants being driven with a 90 deg phase offset [B. T. Hefner and P. L. Marston, J. Acoust. Soc. Am. **106**, 3313–3316 (1999)]. The scattering by symmetric targets placed on the axis of such beams preserves the helicity and the axial null [P. L. Marston, J. Acoust. Soc. Am. **124**, 2905–2910 (2008); F. G. Mitri, Ann. Phys. **323**, 2840–2850 (2008)]. The present work shows how to rapidly alter (or modulate) the helicity with an appropriately timed reversal of the excitation of a single pair of opposite quadrants. A simple four-element helicity-selective detector (using appropriately offset elements) was also demonstrated. These beams may be useful for digital acoustic communication and they have a modulated axial angular momentum flux. They should also generate a modulated acoustic torque on targets. [Work partially supported by ONR.]

3:45

1pUW11. Propagation of low-frequency guided waves in three-dimensional submerged heterogeneous pipes. Jill P. Bingham and David C. Calvo (Acoust. Div., Code 7145, Naval Res. Lab., Washington, DC 20375, jill.bingham.ctr@nrl.navy.mil)

We numerically investigate the propagation of low-frequency guided waves in a fluid-filled elastic pipe submerged in a liquid. In a vacuum, axisymmetric modes travel through the pipe wall. When filled with a fluid of low sound speed, the pipe system can develop modes that are mainly supported by the interior fluid. As mode cut-offs are approached, an increasing amount of energy penetrates the surrounding fluid. In order to determine these cut-off frequencies, we have numerically obtained dispersion curves for the guided pipe modes from the characteristic equations for thick- or thin-walled layered systems in cylindrical coordinates. As a complement to the separation of variable method, we also compute group speed curves from the propagation of interior source pulses through the medium using the time-domain elastodynamic finite-integration technique (EFIT). A time-frequency analysis of the received signals at select probe locations reveals the group speed structure. We also compute the stability of the modes in the presence of heterogeneous contents such as gas or solid particulates using direct simulation with the EFIT code. Finally, we consider propagation through three-dimensional structures such as valves.

4:00

1pUW12. A new mixed optimization algorithm for array pattern synthesis. Wang Yingmin, Qi Hua, and Yu Yanli (College of Marine Eng., Northwestern Polytechnical Univ., Xi'an 710072, China, ywang@nwpu.edu.cn)

A numerical pattern synthesis algorithm based on adaptive notched noise and genetic optimization procedure is presented. For a given array, this technique allows one to attain a set of weight coefficients that shape the array with designed main beam in given directions and lower sidelobes. The algorithm may search for optimization attainable patterns for the specific elements if the desired pattern cannot be found. The new algorithm combines adaptive array principle with genetic technique and a mixed two-step optimization approach is formulated. The first step is based on adaptive theory and the elements of the given array are supposed as the elements of an adaptive array. The desired direction of the main beam is chosen to be the steering direction and sidelobes are controlled by assigning a number of interfering noise throughout the so-called sidelobe area. An initial weight coefficient or pattern is obtained after several iterations. Then an energy function based on the difference between desired pattern and the initial pattern is introduced. A simple genetic algorithm is used to minimize the energy function. The effectiveness of the algorithm is demonstrated on an example of a double-circular array.

Separate registration fee required

MONDAY AFTERNOON, 26 OCTOBER 2009 RIO GRANDE CENTER, 7:00 P.M. TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture: Podcasting Demystified: From Concept to Production For Every Budget

Juan I. Arvelo, Chair
Johns Hopkins Univ., Applied Physics Lab., Laurel, MD 20723-6099

Chair's Introduction—7:00

Invited Paper

7:05

1eID1. Podcasting demystified: From concept to production for every budget. Kathleen P. King (Graduate School of Education, Fordham Univ., 113 W. 60th St., New York, NY 10023)

Podcasting offers an inexpensive and yet powerful way to reach people around the globe with your message in audio or video format. Leveraging our global societies' access of 24/7 web-based platforms to meet their information and learning needs provides a vibrant distribution engine waiting for people of all disciplines to create and post their work. This tutorial will provide an overview of the basic "how-to's" of planning, producing, and hosting your podcast episodes and series. Podcasting possibilities may include redistributing content, and events, supplementing learning experiences, extending the traditional classroom, and integrating new opportunities. Intellectual property, copyright issues, sound production, equipment options for every budget, free software for recording, editing, and production, and resources will be discussed as well. The author hosted and produced over eight series, reached over 6 million people through her podcasting work, written two books, and presented many seminars on this, other digital media, and distance learning topics. Utilizing this experience, insight into the many different formats, audiences, hosting features, and applications of podcasting will be presented. As examples of applications, consider that podcasting can be thought of as teacher-created, student-created, or already existing content. We will discuss how you might use these different forms for different valuable purposes in teaching and learning settings.

Session 2aAAa

Architectural Acoustics: ETS-Lindgren Acoustic Test Laboratory and Factory Tour

Douglas Winker, Chair
ETS-Lindgren, 1301 Arrow Point Dr. Cedar Park, TX 78613

A tour of the new ETS-Lindgren Acoustic Research Laboratory in Cedar Park, Texas will be conducted. The schedule for the day includes a technical discussion on acoustic test facilities followed by a lunch break and factory tour. As a bonus, the tour will feature demonstrations of the material presented in the technical discussion.

Tour participants will see several state-of-the-art chambers for acoustic test services, including a hemi-anechoic chamber and two reverberation chambers, impedance tubes and supporting acoustic test equipment and software. The laboratory offers product noise emission testing and structural/architectural acoustic testing.

Product noise emission testing is commonly performed in the double-walled hemi-anechoic chamber that is designed to measure very low noise emissions from products and devices at 80 Hz and above Outside chamber dimensions are 8.5 m long $\times$ 8.5 m wide $\times$ 7 m high. This chamber is ideal for testing sound power and pressure levels as well as small fan noise. Products tested include Information Technology Equipment (ITE) such as laptop computers and associated printers, home appliances, garden equipment — essentially any noise emitting device may be tested in this chamber. Commonly referenced standards for testing in this chamber include ISO 3744, SO 3745, ISO 7779, ISO 11201, and ECMA 74. Structural/architectural acoustic testing is performed in the reverberation chambers. With transmission loss testing of wall samples, windows, doors, automobile panels and the like, design engineers can determine how much sound energy is transmitted through a product in the chambers. The source chamber measures 7.4 m long $\times$ 5.9 m wide $\times$ 4.8 m high; the receive chamber measures 7.4 m long $\times$ 9.2 m wide $\times$ 6 m high. ASTM E90, ASTM C423, ASTM E596, and ISO 3741 are the most commonly referenced standards for testing in these chambers.

To enhance chamber performance, the hemi-anechoic inner chamber sits on a 50 ton isolated concrete slab while the reverberation chambers sit on individual floating concrete slabs. The laboratory is ISO 17025 accredited under the US Department of Commerce NIST National Voluntary Laboratory Accreditation Program (NVLAP) Lab Code 100286–0. The tour will also feature a stop in ETS-Lindgren's ISO 9001 certified factory. Tour participants will see how acoustic chambers are constructed.

Bus loading will begin promptly at 8:00 am outside the main entrance of the Hyatt Regency Hotel, on Losoya Street. Tour participants will travel in a luxury air-conditioned motor coach for approximately 90 minute ride to Cedar Park (near Austin). The bus is equipped with bathroom facilities. Refreshments will be provided upon arrival at ETS-Lindgren. A traditional Texas Style BBQ lunch buffet will be served at noon. Snacks in the afternoon and a treat for the return bus ride to San Antonio will also be provided. The bus will depart Cedar Park by 2:30 pm for an arrival at the Hyatt Hotel before 5:00 pm, traffic permitting. Please note tour attendance is limited to 50 people and reservations will be confirmed in the order received until space is filled. There is no fee to attend, but you must have a prior reservation to board the bus. To make your reservation, please visit www.ets-lindgrenregistration.com/ASAtour. For more information, please contact Janet O'Neil, janet.oneil@ets-lindgren.com or phone +1.425.868.2558.

Session 2aAAb

Architectural Acoustics, ASA Committee on Standards, Noise, and Speech Communication: Classroom Acoustics: New Design Approaches—Both Successes and Failures

Kenneth W. Good, Jr., Cochair
Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17603

Pamela J. Harght, Cochair
BAi, LLC, 4006 Speedway, Austin, TX 78751

Chair's Introduction—7:55

Invited Papers

8:00

2aAAb1. Predicting acoustic performance in reconfigurable classrooms. Kenneth P. Roy and Kenneth W. Good, Jr. (Bldg. Products Technol. Lab., Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604, kproy@armstrong.com)

Both laboratory component tests and field system evaluations were conducted to screen architectural designs relative to sound quality performance within a classroom, and noise intrusion between classrooms. Recognized standards such as ANSI S12.60 for Acoustics in Schools and other rating systems such as LEED for Schools and CHPS provide guidance on performance/design needs for K-12 schools. But what about postsecondary schools and especially those designed with reconfigurable architectural elements such as relocatable walls and raised floor system—can these be made to work? Initial results from this ongoing research will be discussed, including wall performance, ceiling/plenum effects, and sound system effects.

8:20

2aAAb2. Case study in classroom acoustics measurements. Kenneth Good and Kenneth Roy (Armstrong World Ind. 2500 Columbia Ave., Lancaster, PA 17603, kwgoodjr@armstrong.com)

Recent work was done related to exploring variations and interpretations of measurement techniques in a classroom environment. The primary focus of the work was related to quantifying room to room isolation; however, minimal background noise and reverberation time measurements were included. This presentation will discuss the differences in the measurements methods, results, and when each may be appropriate.

8:40

2aAAb3. Modular classroom acoustics: Where we are and where we should be going. Norman H. Philipp and Lily M. Wang (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, philipp.norman@gmail.com)

This paper presents a review of an 18-month study into the acoustical characteristics of modular (or relocatable) classrooms within the Omaha Public School District in Omaha, NE. The study included measurements of reverberation times, interior background noise levels (both occupied and unoccupied), and exterior facade sound insulation properties. From analysis of the gathered data, recommendations are made regarding sound insulation classification systems and background noise levels for the renovation of existing modular classrooms. These results may be used to inform the pending addendum to ANSI S12.60-2002, regarding acoustic guidelines for modular classrooms. Suggestions for future research are also discussed.

9:00

2aAAb4. Acoustics in a high school gymnasium. Stephanie Hoeman, Jon Birney, Hannah Schultheis, Shane Kanter, and Bob Coffeen (The Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66046, stephanie.hoeman@gmail.com)

A high school gymnasium constructed with unintentionally interesting acoustics was recently scheduled for a new sound reinforcement system. It was determined that the sound reinforcement system would be of little consequence until the architectural acoustics of the space were addressed. The gymnasium was constructed of entirely concave shapes which resulted in severe sound focusing. The space was extremely reverberant and exhibited many distinct, distracting reflections. Impulse responses were made in the space and computer models were made. The goal was to find a solution to reduce unwanted reflections and shorten the reverberation time. Since this project was for a public school with limited funds, utmost importance was placed on a solution that would be economical and uncomplicated to install. Acoustical analysis and auralizations of the space in its current condition and the projected performance of recommendations were compared to find an effective, cost-efficient solution the school board could implement to improve the architectural acoustics before the sound reinforcement system was added.

9:20

2aAAb5. Current trends in K-12 classroom design. Jessica Molter and Jonathan Hodge (Pfluger Assoc., LP 209 E Riverside Dr., Austin, TX 78704)

Presentation of current trends in K-12 Classroom Design with an emphasis on items that contribute to acoustic performance of the space.

9:40

2aAAb6. Classroom reinvented, a quantum leap in classroom design. Kenneth Good (Acoust. Privacy Enterprises, LLC, Mount Joy, PA 17552, kwgoodjr@acousticprivacy.com)

How do you design an environment to maximize your impact on a child's development when you are limited to just an hour or two a week? Traditional academic education or biblical instruction, the environment is critical to success in reaching children and youth. This case study will explore the acoustical performance and design strategies that went into LCBC's new and nontraditional building for children and students.

Contributed Paper

10:00

2aAAb7. Optimizing the signal-to-noise ratio in speech rooms using passsive acoustics. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

Adults with normal hearing require a roughly 0-dB signal-to-noise ratio for good speech intelligibility in classrooms and lecture halls. However, significantly higher values may be needed to compensate for neurological immaturity, sensorineural and conductive hearing losses, language proficiency, and excessive reverberation. ANSI 12.60 addresses ways to lower the noise interference due to background levels and reverberation time. However, it is also possible to increase the signal, by reflecting or diffusing early reflection . While speech power is delivered in the vowels which are predominately in the 250–500-Hz frequency range, speech intelligibility is delivered in the consonants, which occur in the 2–4-kHz frequency range. Therefore, effective core learning designs can incorporate scattering surfaces, rather than surfaces that absorb in the 2–4-kHz region, on the front wall, lower side walls, and central ceiling areas, to increase the speech signal. The decay time can be controlled with broadband absorption on the perimeter of the ceiling and upper wall surfaces. A computer model analysis of various speech environments will be presented.

TUESDAY MORNING, 27 OCTOBER 2009 REGENCY EAST 2, 10:20 A.M. TO 12:00 NOON

Session 2aAAc

Architectural Acoustics, Noise, ASA Committee on Standards, and Speech Communication: Classroom Acoustics: An Update

David Lubman, Cochair
DL Acoustics, 14301 Middletown Ln., Westminster, CA 92683-4514

Kenneth W. Good, Jr., Cochair
Armstrong World Industries Inc., 2500 Columbia Ave., Lancaster, PA 17603

Invited Papers

10:20

2aAAc1. Working towards an enforceable standard for classroom acoustics. Lois Thibault (US Access Board, 1331 F St. NW, 1000, Washington, DC 20004)

The US Access Board has proposed to reference ANSI/ASA S12.60-2002 (R-2009) in the 2012 International Building Code (IBC), which will provide for local enforcement. If this proposal is adopted, the Board will then update its 2004 ADA-ABA Accessibility Guidelines to harmonize with the IBC. This paper will review the history of classroom acoustics initiatives, including the standard, and update attendees on the current process. Other classroom acoustics activity will be highlighted, particularly sustainability and green design proposals, and the 2010 re-authorization activities under the Individuals with Disabilities Education Act.

10:40

2aAAc2. The acoustic treatment in classroom refurbishment: A double blind experimental study examining the acoustic and auditory environment of the cellular classroom. David C. Canning (Div. of Psych. and Lang. Studies, Univ. College London, 26 Bedford Way, London, WC1H 0AP, United Kingdom and Hear2Learn, United Kingdom, canningd@gmail.com)

This paper will present data from the Essex Study of mainstream schools designed for inclusion of children with special hearing difficulties. It will present data from the experimental study which was set up to guide the specification of classroom acoustic performance standards. The UK Special Educational Needs Legislation requires education authorities to meet the special educational needs of all children. The major need considered in the provision of children with special hearing requirements, including deaf and hard of hearing children, is their ability to function when there is competing sound. Control of competing sound is in part the job of the teacher, as the dominant source of sound is created by the occupants, but the acoustician has a significant role to play. The acoustic performance of a classroom has considerable impact on every aspect of teaching and learning and ultimately on whether a child's special educational needs can be met within an inclusive educational setting. The findings of the study have implications for the acoustical performance of all schools, given that every school might reasonably be expected to provide for children with special hearing requirements. [The support of Sweyne Park School, Essex County Council, The National Deaf Children's Society, and the Federation of Property Services is acknowledged.]

11:00

2aAAc3. Auralizing adult-child listening differences. Peggy B. Nelson (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, peggynelson@umn.edu), Jonah Sacks, and Jennifer Hinckley (Acentech, 33 Moulton St., Cambridge, MA 02138)

Substantial data from previous research show that children and adults require different acoustical conditions for good understanding. For example, adults can understand the majority of speech when the audibility of the speech is reduced to 40% or 20%, but young children need 80% or 60% audibility for the same level of understanding. Also, while adults need 4 to 6 bands of vocoded speech to reach good performance levels, children need 8–12 bands. While adults experience a release from masking when signals and background noise arise from different angles, children do not gain the same benefit, and in fact may experience masking from background noises coming from any direction. J.H. and J.S. from Acentech prepared auralizations that demonstrate these differences between children and adults. Those auralizations will be presented and discussed as possible educational tools.

11:20

2aAAc4. Integrating acoustical issues in the design of high-performance schools. Gary W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611), Chris Jones, Robert M. Lilkendey, Hyun Paek, and Reece Skelton (Siebein Assoc., Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

ANSI 12.60 presents acoustical performance criteria for interior finishes to control reverberation inside classrooms, provide sound isolating wall and floor/ceiling assemblies between rooms, and limit noise from building equipment to allow perception and understanding of human voices within educational occupancies by young listeners with normal hearing and some degrees of hearing impairments. In practice, many schools are built that develop in an uneasy way trying to meet or sometimes to avoid meeting ANSI criteria because of perceived difficulties and expense in achieving the required results. A method to integrate acoustical design principles with the basic architectural design scheme of a school that can be implemented early in the design process was developed so acoustical performance can be simply implemented in school projects. This paper presents a core set of acoustical planning principles that can be implemented early in the design process so that ANSI criteria for room finishes, sound isolation, and building equipment noise can be met as the design progresses.

11:40

2aAAc5. High-performance acoustic ceilings make quiet classrooms quieter. David Lubman (DL Acoust., 14301 Middletown Ln., Westminster, CA 92683, dlubman@dlacoustics.com) and Louis C. Sutherland (Consultant in Acoust., 27803 Longhill Dr., Rancho Palos Verdes, CA 90275)

Highly beneficial noise level reductions of 5–10 dB are reported in occupied classrooms equipped with highly sound absorbing ceilings. The incremental cost for such ceilings is nominal. These "Lombard effect" benefits apply to small classrooms with very low reverberation times (0.5 s or less) which is less than the 0.6-s maximum specified in ANSI S12.60-2002. The amount of noise reduction depends on teaching/learning style. In the typical noise reduction lecture classrooms (single talker), the typical noise reduction benefit is 5 dB in lecture classrooms (single talker) and group learning classrooms (multiple talkers). Unfortunately, these benefits require unoccupied background noise levels (BNLs) of about 35 dBA. They are not expected in typical noisy American classrooms (BNL $\sim>$45 dBA). This underscores the importance of compliance with ANSI's 35-dBA BNL limit rather than the lenient and unsupported 45-dBA limit permitted in recent LEED and California CHPS guidelines. Details are reported in the outstanding study "Acoustic Ergonomics of School" by Oberdorster and Tiesler [University of Bremen, Germany (2006)]. Similar benefits were reported earlier in a case study by Sutherland ["The role of soundscape in children's learning," J. Acoust. Soc. Am. **112**, 2412–2413 (2002)].

Session 2aAOa

Acoustical Oceanography: Acoustic Inversion

Juan I. Arvelo, Chair
Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Contributed Papers

8:00

2aAOa1. Seabed sound speed and attenuation from broadband acoustic measurements in the Shallow Water 2006 experiment. Lin Wan, Ji-Xun Zhou, Peter H. Rogers (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, lin.wan@gatech.edu), and David P. Knobles (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713)

In the Shallow Water 2006 experiment, a set of broadband combustive sound source (CSS) signals was measured by two L-shaped arrays separated by 20 km on the New Jersey continental shelf with a water depth of around 72 m. The measured CSS data, which exhibit modal dispersion, are utilized to infer values for the seabed sound speed and attenuation. The seabed sound speed is estimated by matching the theoretical and measured modal dispersion curves, which are extracted using an adaptive time-frequency analysis technique. The frequency dependence of the seabed attenuation is inferred by minimizing the difference between the theoretical and measured modal amplitude ratios in the 50–400-Hz band. The resultant seabed attenuation is similar to the low-frequency seabed attenuation data obtained at 20 locations in different coastal zones around the world by Zhou *et al.* [J. Acoust. Soc. Am. **125**, 2847–2866 (2009)]. The depth dependence of the seabed attenuation will be discussed. [Work supported by the Office of Naval Research.]

8:15

2aAOa2. Time domain geoacoustic inversion using back-propagation on an L-shaped array. Cheolsoo Park, Peter Gerstoft, Woojae Seong, and William S. Hodgkiss (Marine Phys. Lab., Univ. of California, San Diego, La Jolla, CA 92093-0238)

This paper presents inversion of Shallow Water 06 experimental data measured on an L-shaped array (SWAMI 32). The array was deployed in 70-m water depth and consists of an equally spaced 10-element vertical line array (VLA) and a 256-m-long bottom moored 20-element horizontal line array (HLA). A mid-frequency (1100–2900-Hz) chirp source was towed at 35-m depth along a circular track around the VLA with radius 190 m. Time domain inversions using VLA, HLA, and both arrays were carried out and results compared. For the inversions, a multi-step optimization scheme is applied to the data using very fast simulated reannealing. The objective function is defined by the power of the back-propagated signal from the array to the source. At each step, water column sound speed profile, experimental geometry, and geoacoustic parameters are inverted successively. Accurate HLA positions were essential for the HLA and the HLA+VLA inversions. Finally, the inversion results were compared with other results near the site. [Work supported by ONR.]

8:30

2aAOa3. The estimation of geoacoustic parameters via low frequencies (50–100 Hz) for selected Shallow Water 06 test data. A. Tolstoy (ATolstoy Sci., Inc., 1538 Hampton Hill Circle, McLean, VA 22101, atolstoy@ieee.org) and Yong-Min Jiang (Univ. of Victoria, Victoria, BC, V8W 3P6, Canada)

This work will demonstrate the geoacoustic inversion "success" on data of using only one or two low frequencies, multiple ranges, and multiple realizations for geoacoustic inversion of actual SW06 data. The data used are the same as those processed by Jiang and Chapman and involves three ranges (1, 3, and 5 km) and multi-tonal continuous wave data collected on a 16 phone vertical array. Multiple realizations of the data were used where each reduced the non-uniqueness a bit. The multiple ranges and frequencies also reduced the non-uniqueness of the suggested solutions. However, there still remains a significant number (hundreds) of possible "solutions," (values of ctop, cbot, hsed, and chsp) for which MFP$<$0.9 (including those suggested by Jiang and Chapman). The use of higher frequencies requires refinement of more parameters (such as the ocean sound-speed profile, source depth and range, water depth, and phone locations) but would not necessarily improve estimates of such bottom parameters as chsp and hsed. Thus, non-uniqueness of bottom parameters is an issue which may well exist for all inversion approaches.

8:45

2aAOa4. Ocean tomography with acoustic daylight: A case study. Oleg A. Godin, Nikolay A. Zabotin (NOAA/Earth System Res. Lab., CIRES, Univ. of Colorado, Boulder, CO 80305, oleg.godin@noaa.gov), and Valery V. Goncharov (P. P. Shirshov Oceanology Inst., Russian Acad. of Sci., Moscow 117997, Russia)

Ambient and shipping noise in the ocean provides acoustic illumination, which can be used, akin to daylight in the atmosphere, to visualize objects and characterize the environment [Buckingham *et al.*, Nature (London) **356**, 327–329 (1992)]. It has been shown theoretically [O. A. Godin, Phys. Rev. Lett. **97**, 054301 (2006)] that, under rather general conditions, deterministic travel times between any two points in an inhomogeneous, moving or motionless, time-independent medium can be retrieved from the cross-correlation function of non-diffused acoustic noise recorded at the two points, without a detailed knowledge of the noise field's sources or properties. Using the data obtained during the 1998–1999 Billboard Array Experiment [Worcester *et al.*, J. Acoust. Soc. Am. **117**, 1499–1510 (2005)], this paper demonstrates the feasibility of a tomographic reconstruction of the sound speed field from cross-correlation of acoustic noise recorded on a pair of vertical line arrays (VLAs) in deep water. Limitations of the noise data inversion associated with the ocean temporal variability, the VLA horizontal separation, recording bandwidth, and the noise directionality are analyzed. Prospects of long-range water column tomography with acoustic daylight are discussed. [Work supported by ONR.]

9:00

2aAOa5. Two-point coherence of acoustic noise recorded by the North Pacific Acoustic Laboratory billboard array. Oleg A. Godin and Nikolay A. Zabotin (NOAA/Earth System Res. Lab., CIRES, Univ. of Colorado, Boulder, CO 80305, oleg.godin@noaa.gov)

Acoustic noise in the ocean contains extensive information about the noise sources and the propagation environment. In particular, two-point correlation function of noise (NCCF) is known to have peaks which correspond to acoustic travel times between the two points provided the noise field is sufficiently diffuse. In this paper, measurements of acoustic noise performed during the 1998–1999 Billboard Array Experiment [Baggeroer *et al.*, J. Acoust. Soc. Am. **117**, 1643–1665 (2005)] are re-examined with the goal of extracting environmental information. NCCF is evaluated by averaging time series of noise recorded on various vertical line arrays that comprise the Billboard Array. For any two hydrophones, NCCF is found to have a number of robust peaks. Possible generation and propagation mechanism responsible for various features of the NCCF are discussed. Statistical distributions

of noise are utilized to differentiate between the NCCF peaks of different origins and to identify those peaks, which can be used to retrieve information about the sound speed without detailed knowledge of the noise sources. [Work supported by ONR.]

9:15

2aAOa6. Measurements of sound speed in bubbly liquids under high-pressure conditions. Chad A. Greene and Preston S. Wilson (Mech. Eng. Dept. and Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78712-0292)

Methane hydrates occur naturally on the ocean bottom and in the upper layers of sediment on continental shelves. Seismic surveying could be used to locate methane hydrates; however, their low-frequency acoustic properties are not well-known. In addition, these properties can vary dramatically depending on whether the methane is in a gas or solid phase. As a step toward better understanding the three-phase case of gassy sediments in water, the two-phase case of methane gas bubbles in water was investigated. Wood's equation is often used to model sound propagation in bubbly liquids and has been widely verified by experiments at atmospheric pressure. However, there is little information in the literature verifying the validity of Wood's equation at high pressures. Low-frequency (0.5–10-kHz), resonator-based sound speed measurements were obtained for air bubbles in water and methane bubbles in water under pressures ranging from 1 to 10 atm at room temperature. The results are presented and compared to the predictions of Wood's equation. [Work sponsored by ONR.]

9:30

2aAOa7. Estimating bubble density from attenuation measurements through an underwater explosion. Fred D. Holt, IV and R. Lee Culver (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

Underwater explosions have been studied intensively in the United States since 1941 [e.g., Cole (1945)]. Research to date primarily focuses on the initial shock and subsequent pressure waves caused by the oscillations of a "gas globe" that is the result of a charge detonation. These phenomena have relatively short timescales (typically less than 2 s). However, as the gas globe rises in the water column and breaks the surface, it leaves behind a residual bubble cloud which has been markedly less studied. A recent experiment measured the spatial and time-dependent acoustic response of the bubble cloud resulting from a charge detonated at 50-ft depth. A directional projector was used to propagate a linear FM (5–65-kHz) pulse through the bubble cloud to an array of hydrophones placed on the opposite side of the charge in order to measure attenuation. This talk will focus on the methods used to estimate bubble density size spectra from the attenuation measurements, those of Commander and McDonald (1991), Caruthers (1999), and Czerski (2009). [Work sponsored by the Office of Naval Research, Code 333.]

9:45

2aAOa8. Sediment shear as a perturbation in geoacoustic inversions and an explanation of the anomalous frequency dependence of the attenuation. Allan D. Pierce and William M. Carey (Dept. of Mech. Engr., Boston Univ., Boston, MA 02215, adp@bu.edu)

The depth dependent shear wave speed in marine sediments is much less than both the compressional and water column sound speeds. The neglect of shear in geoacoustic inversions is usually justifiable, but at frequencies less than 300 Hz, the loss of acoustic energy from the water column because of nonreturning radiation of shear waves into the bottom dominates the loss due to the intrinsic attenuation of the sediment. To account for this in a simple manner, a perturbation theory that takes advantage of the small shear speed has been devised. The canonical problem addressed takes the disturbance as having a horizontal trace velocity ω/k in the x direction, this being somewhat less than the compressional speed in the bottom but substantially larger than the depth-dependent shear wave speed. To lowest order the stress components σ_{zz} and σ_{xx} satisfy the same reduced Helmholtz–Bergmann equation as if the bottom were an inhomogeneous fluid. The shear stress σ_{xz} satisfies an approximately uncoupled equation governing a shear wave propagating vertically downward. Coupling occurs at the interface, and the power carried off by the shear wave can be approximately determined. Results succinctly and quantitatively explain why geoacoustic inversions yield attenuation frequency exponents less than 2.

TUESDAY MORNING, 27 OCTOBER 2009 RIO GRANDE EAST, 10:15 A.M. TO 12:00 NOON

Session 2aAOb

Acoustical Oceanography: Acoustics and Ocean Acidity I

Timothy F. Duda, Cochair
Woods Hole Oceanographic Inst., 98 Water St., Woods Hole, MA 02543-1053

Peter F. Worcester, Cochair
Univ. of California San Diego, Scripps Inst. of Oceanography, 9500 Gilman Dr., La Jolla, CA 92093

Chair's Introduction—10:15

Invited Paper

10:20

2aAOb1. A brief history of the discovery of the low frequency sound absorption mechanism in seawater and its *p*H dependence. David G. Browning (139 Old North Rd., Kingston, RI 02881), Vernon P. Simmons (335 Burgundy Rd., Healdsburg, CA 95448), and William H. Thorp (2 Brook St., Noank, CT 06340)

In the 1960s tests with a new surface ship sonar indicated that low-frequency (less than 10 000 Hz) propagation loss was greater than predicted by the existing $MgSO_4$ absorption model. Propagation measurements in the SOFAR channel showed that the anomalous absorption was up to ten times greater than predicted and could be fitted by a 1-kHz relaxation. This was not without controversy at the time, but interest in this fundamental parameter spurred a major measurement effort by the U.S. Navy Underwater Sound Laboratory and other laboratories in the US, Canada, New Zealand, Europe, and Australia, resulting in at-sea measurements from Hudson Bay to

the Tasman Sea. Yeager and Fisher conducted laboratory t-jump measurements that identified a boron based low-frequency relaxation mechanism in seawater. Simmons and Fisher followed with resonant sphere measurements to quantify the resulting acoustic absorption. A compilation of the at-sea results showed, somewhat surprisingly, a correlation with *p*H: the lower the *p*H, the lower the absorption. Mellen quantified the *p*H dependence by extensive laboratory resonator measurements. These results were confirmed by measurements conducted in China. The possible application of acoustically monitoring ocean *p*H was first suggested by Browning and Mellen in 1990.

Acoustical Oceanography Mini Tutorial

10:45

2aAOb2. Rapidly changing ocean *p*H and the increasing transparency of the ocean to sound. Peter G. Brewer (MBARI, 7700 Sandholdt Rd., Moss Landing, CA 95039, brpe@mbari.org)

The intrinsic sound absorption coefficient (á, dB/km) of seawater is *p*H dependent with significant effects at 10 kHz and below. Ocean *p*H is declining from fossil fuel CO_2 invasion, from excess nutrient input, and from climate change reducing ocean ventilation. These effects produce reductions in ocean borate and carbonate species such that an ~18% decrease in á in the upper ocean has occurred today. Reasonable projections based on IPCC scenarios predict changes of 40% or more by mid-century [Hester *et al.*, Geophys. Res. Lett. **35**, L19601 (2008)]; larger changes in the sound channel are very likely. The projected increased transparency of the ocean to sound has attracted strong international environmental attention, and also provides an opportunity for acoustic detection and monitoring of such changes over large ocean regions. How strong is the basis for such assertions, can they be tested, and how well can environmental effects be predicted? This tutorial shows that the rate of change in *p*H in the sound channel has been underestimated, and how the unusual 1970s era reliance on the Soviet Gorshkov atlas as a *p*H data resource came about. This now presents a challenge for convergence of modern acoustic, environmental, and global change needs.

Invited Paper

11:35

2aAOb3. Long-term trends in ambient noise levels. George V. Frisk (Dept. of Ocean Eng., Florida Atlantic Univ., 101 N. Beach Rd., Dania Beach, FL 33004, gfrisk@seatech.fau.edu)

This paper addresses the subject of long-term trends in ambient noise levels, a topic of great interest to both the scientific community and the general public. This attention stems primarily from concerns over the effects of apparently increasing sound levels on marine mammals. A growing, though limited, body of literature suggests that low-frequency noise levels increased approximately 15 dB during the period 1950–2000, an amount that corresponds to about 3 dB/decade. One hypothesis states that this increase is predominantly anthropogenic in nature and can be attributed to increased commercial shipping activity, which, in turn, can be linked to global economic growth. As a result, a direct correlation may be drawn between ambient noise levels and the behavior of the global economy. This special session addresses an additional consequence of global economic activity, namely, increased ocean acidification leading to decreased absorption and therefore to increases in ambient noise levels associated with distant shipping. This paper also suggests topics of considerable interest for future research including (1) the relative contributions of commercial shipping activity versus ocean acidification to ambient noise levels and (2) the effect of the current economic downturn on noise levels. [Work supported by ONR.]

TUESDAY MORNING, 27 OCTOBER 2009 PECAN, 8:00 A.M. TO 12:00 NOON

Session 2aEA

Engineering Acoustics and Signal Processing in Acoustics: Acoustic Measurement and Models for Sensors and Arrays

Dehua Huang, Chair
Naval Undersea Research Center, 1176 Howell St., Newport, RI 02841-1708

Chair's Introduction—8:00

Invited Papers

8:05

2aEA1. Multidomain modeling of a variable reluctance transducer. Stephen C. Thompson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

The variable reluctance magnetic transducer consists simply of a magnetic air gap and a magnetic return path that are supplied with static and dynamic sources of magnetic flux. The dynamic flux is generated by a coil current. The static flux may come either from a permanent magnet or a coil. The attractive magnetic force across the gap is modulated by the dynamic flux to provide the mechanical excitation for the device. The variable reluctance device is fundamentally nonlinear for at least three reasons: (1) the mechanical force across the gap is a quadratic function of the total magnetic flux, (2) the changing gap dimension changes the reluctance in the magnetic circuit so that the flux does not change linearly with coil current, and (3) saturation of the magnetic circuit may be an important aspect

of the design that must be included in the modeling. In practical cases, the dynamic variation in gap dimension is large enough that a linearized approximation is insufficient to predict the performance. An approximate model using a set of nonlinear differential and algebraic equations will be discussed that can predict the stability and performance of variable reluctance transducers.

8:25

2aEA2. Lead magnesium niobate-lead titanate solution based giant-piezoelectric crystals for next generation of acoustic transduction devices. Pengdi Han (H. C. Mater. Corp., 479 Quadrangle Dr., Ste.-E, Bolingbrook, IL 60440)

The PMN-PT (binary solid solution of lead magnesium niobate and lead titanate) based piezoelectric crystals have been commercialized now. The giant-piezoelectric crystal can be broadly utilized for the next generation of acoustic transduction devices. In this paper, the major concepts and current status of the crystals and products will be presented and discussed in terms of applications. Following a brief review on the history of development of crystal growth and commercialization in the past decade, the detailed physical properties of the PMN-PT and newly developed PIN-PMN-PT crystals (high depoling temperature and high driving field) are presented and discussed in terms of applications in ultrasound transducers. At present (001)-seeded single crystals of 3 and 4 in. diameters are commercially available. In order to sufficiently utilize the advantages of the crystals, discussions focus on the following: (1) The differences between single crystal and PZT ceramics. (2) The concept of so called "domain-engineering" (artificially domaining) in context of elasto-piezo-dielectric matrices. (3) The relationships between the property and ferroelectric domain structure. (4) How to select the crystal products. In summary, the achievement of the development and fabrication of the giant-piezoelectric crystals has led to a new era that there are new opportunities readily to be explored for the next generation of acoustic transduction devices.

8:45

2aEA3. Analysis models for the Underwater Sound Reference Division low-frequency acoustic calibration systems. Dehua Huang and Anthony Paolero (NUWC, Newport, RI 02841)

The state of an art for calibrating acoustic transducers at very low frequencies is by way of a confined and well understood environment. The Underwater Sound Reference Division (USRD) has three such calibration systems, called systems K, J, and L, respectively. Each system has a cylindrical tube, of certain length and diameter, that determine a cutoff calibration frequency. System K operates in a standing wave mode condition. Systems J and L both operate in traveling wave mode conditions, where plane waves propagate from one end of the tube to the other. Optimally locating a calibrated transducer and an unknown within the tube provides the proper configuration for calibration. This paper demonstrates the simulation tools to predict performance of aforementioned low-frequency calibration systems. The mathematical model, the GUI coding, simulation, predicted, and test results will be presented.

9:05

2aEA4. Electroacoustic transducer and array modeling tools. Ender Kuntsal (Int. Transducer Corp. (ITC), 869 Ward Dr., Santa Barbara, CA 93111, ekuntsal@itc-transducers.com)

The availability and the accuracy of transducer and array design tools are becoming even more important with the changing economical climate. The engineers would like to evaluate their designs more quickly in place of the prototyping, which requires additional manufacturing and testing time and expense. The software models must be user friendly, provide results that can be relied on, and also be affordable. The strength of a transducer and array design engineering team is still extensively based on experience and background, but these tools make the process more efficient. There are many transducer and design software programs used in industry and academia. Some are based on analytical solutions and some use more complicated numerical methods such as finite element models, which have the capability of modeling piezoelectric materials and acoustical parameters. Among these are ATILA, PAFEC, MAVART, NASTRAN, PZFlex, and other powerful programs such as CHIEF and TRN. Some of these programs can be used to design both resonators and arrays while taking baffling and/or acoustic interaction effects into account. In this presentation, a summary of commonly available software packages and their features will be described.

9:25

2aEA5. High-power single crystal based projectors. Richard J. Meyers, Jr., Douglas C. Markley, Charles W. Allen, and Nevin P. Sherlock (The Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

Significant progress has been made at integrating single crystalline relaxor ferroelectrics into many types of SONAR transducers. For high-power projectors, PMN-28PT has offered efficient, broad band-width capability. For higher duty cycles, thermal limitations are reached resulting in power and duty cycle tradeoffs. Electrical bias is also required for crystal implementation when high-power operation is needed. Continuing research in compositional tailoring has resulted in several new modified relaxor crystal systems with improved temperature stability, lower loss, and higher coercive fields. PIN-PMN-PT ternary crystals are of particular interest. This ternary system offers 25–40 °C improvement in working temperature, reduced temperature sensitivity, and approximately two times increase in coercive field without sacrifice in electromechanical coupling or piezoelectric coefficients. To demonstrate the impact of this newer ternary crystal composition on high-power transduction, planar tonpilz arrays were fabricated and tested. This presentation will highlight important material properties and compare the results of high-power measurements of single transducer elements and arrays made from PMN-28PT and PIN modified ternary crystals. Acoustic tests were conducted as a function of ambient temperature, increased drive level, and increased duty cycle operation. [Work supported by ONR.]

9:45

2aEA6. Broadband characteristics of piezoelectric transducer bonded to a thick plate resonator. Iwaki Akiyama, Natsuki Yoshizumi (Dept. of Electric and Electron. Eng., Shonan Inst. of Tech., 1-1-25 Tsujido-nishikaigan, Fujisawa 251-8511, Japan, akiyama@elec.shonan-it.ac.jp), Shigemi Saito (Tokai Univ., Orito, Shimizu-ku, Shizuoka 424-8610, Japan), Katsumi Ohira, Osamu Takahashi (Japan Probe Co., Ltd., Minami-ku, Yokohama 232-0033, Japan), and Kentaro Nakamura (Tokyo Inst. of Tech., Yokohama 226-8503, Japan)

The piezoelectric transducer bonded to a thick plate resonator wholly vibrates as a single transducer. Since the distribution of stress in the thickness direction is asymmetric, the transducer resonates at even order harmonic frequencies as well as odd order ones. The back of the resonator is bonded to a quarter-wavelength matching layer to transmit backward the fundamental frequency wave through the layer. The front of piezoelectric transducer is bonded to another matching layer to transmit forward the ultrasonic wave of higher order frequencies. If both the piezoelectric transducer and the resonator are made of low Q materials, the admittance curve including troughs among resonance peaks is flattened and shows broadband characteristics. Such broadband characteristic transducers made of one to three composite piezoelectric materials were experimentally studied. The thickness of the piezoelectric transducer and the resonator were designed for resonance frequencies of 7 and 1 MHz, respectively. The back of matching layer was bonded to the absorbing material not to reflect forward the ultrasonic waves of 1 MHz. As a result, the ultrasound of center frequency of 5 MHz and fractional bandwidth of 100% were transmitted from the transducer driven by an impulsive signal.

10:00—10:15 Break

10:15

2aEA7. The design of quarter-wavelength impedance matching layers for cylindrical transducers. Douglas R. Heyden, Preston S. Wilson (Appl. Res. Labs. and Mech. Eng. Dept., Univ. of Texas at Austin, Austin, TX 78713-8029), and Richard H. Crawford (The Univ. of Texas at Austin, Austin, TX 78712-0292)

Impedance matching layers are commonly used in piezoelectric underwater acoustic projectors. The layer maximizes transmitted power from the ceramic into the water and also increases the bandwidth of the projector. For the design of cylindrical transducers, it is common in practice to utilize the familiar plane wave formulation of the quarter-wavelength impedance matching layer as a starting point for the design of the cylindrical layer. Material properties and thickness are then modified by trial and error, heuristic, or empirical methods to optimize the design. This practice is undertaken because, apparently, the quarter-wavelength impedance matching layer formulation is not readily available in the acoustics literature for the cylindrical coordinate system. To address this deficiency, the reflection and transmission coefficients for the cylindrical three-medium problem were derived. No general zero-reflection, perfect transmission condition was found, but the equations can be used to find the material properties and layer thickness required to maximize transmission at a given frequency. The results of the derivation are shown and used in the design of a layered cylindrical piezoelectric transducer.

10:30

2aEA8. Time reversal focusing for pipeline structural health monitoring. Joel Harley, Nicholas O'Donoughue, José M.F. Moura (Elec. and Comput. Eng., Carnegie Mellon Univ., Pittsburgh, PA 15213), and Yuanwei Jin (Univ. of Maryland, Eastern Shore, Princess Anne, MD 21853)

Guided wave technologies have become popular tools for nondestructive testing due to their potential to travel long distances. Unfortunately, analyzing data from guided waves is often difficult because of the numerous multimodal and dispersive effects that distort signals in solid media. Time reversal has been shown to be robust against these unwanted and adverse effects. Time reversal techniques are commonly used to focus ultrasonic waves across a medium and have been used to perform nondestructive testing using pulse-echo techniques. This paper investigates the use of time reversal processing techniques to compensate for multimodal and dispersive effects in a low-power structural health monitoring system for pipelines. Time reversal methods are demonstrated as a pitch-catch operation between two transducer arrays to illuminate changes caused by damage on a pipe. It is then shown and discussed how differences in the location and severity of damage affect the signals recorded at the receiving transducer array and how these results can be interpreted to measure those changes. The results are demonstrated experimentally and then compared with equivalent finite element simulations. [National Energy Technology Laboratory (NETL) is the funding source for this effort with Cost Share being provided by Carnegie Mellon University (CMU). CMU is funded under a Subcontract Agreement with Concurrent Technologies Corporation. N.O. is supported by a National Defense Science and Engineering Graduate Fellowship.]

10:45

2aEA9. A radial propagator for computing axisymmetric pressure fields using the angular spectrum method. Edward H. Pees (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841)

The notion of a propagator is central to the angular spectrum of plane wave formulation of diffraction theory, which expresses the pressure field diffracted by a two-dimensional aperture as a superposition of a continuum of plane waves. In the conventional form, an exponential term, known as a propagator, is multiplied by the wavenumber spectrum obtained from a two-dimensional spatial Fourier transform of the aperture boundary condition, to obtain the wavenumber spectrum in a plane parallel to the boundary, offset by some distance specified in the propagator. By repeated use of this propagator and Fourier inversion, one can completely reconstruct the homogeneous part of the pressure field beyond the aperture boundary. In this presentation, we draw upon earlier work relating the boundary condition to the axial pressure and show that when the aperture is axially symmetric, an alternative type of propagator can be derived that propagates an axial wavenumber spectrum away from the axis of the aperture. Use of this radial propagator can be computationally advantageous since it allows for field reconstruction using one-dimensional Fourier transforms instead of Hankel transforms or two-dimensional Fourier transforms.

11:00

2aEA10. A planar acoustic array for voice collection. David J. Gonski, Duong Tran-Luu, and Stephen Tenney (Army Res. Lab., 2800 Powder Mill Rd., Adelphi, MD 20783, dgonski@arl.army.mil)

A planar 7×7 acoustic array of microphones spaced 1 in. apart has been developed and tested. This array is intended for the collection of human speech. A simple analog electronic summation of the 49 microphone signals is carried out to form a preferential collection area in front of the array. We have conducted an analysis of the directional performance of the array to model its directivity. Two configurations of the array were tested. An array was populated with omnidirectional hearing aide microphones, and a second array was configured with cardioid microphones. Performance of each of these arrays was measured in an anechoic chamber and compared with the theoretical performance. To improve the front to back rejection a sound-absorbing pad was placed on the backside of the array. Two versions of the pad were characterized in the anechoic chamber, both were found to be effective. The array was packaged into a plastic box with open cell foam on the front for wind noise suppression and the sound-absorbing pad on the back for improved front to back rejection. Detailed acoustic and electronic design characteristics are presented.

11:15

2aEA11. Microphone array techniques using cross-correlations. Matthew B. Rhudy (Univ. of Pittsburgh, 560 Benedum Hall, Pittsburgh, PA 15261, mbr5002@gmail.com) and Brian A. Bucci (Dept. of Mech. Eng., Univ. of Pittsburgh, Pittsburgh, PA 15261)

Civilian noise complaints and damage claims have created a need to establish a detailed record of impulse noise generated at military training facilities. Wind noise is causing false positive impulse detections in the cur-

rent noise monitoring systems. Multiple channel data methods were investigated in order to distinguish the characteristics of noise events. A microphone array was used to collect four simultaneous channels of military impulse and wind noise data. Cross-correlation functions were then used to characterize the input waveforms. Three different analyses of microphone array data were developed. A new value, the min peak correlation coefficient, is defined from the minimum value of peaks of the cross-correlation coefficient functions among the different channels. This value is a measure of the likelihood that a given waveform originated from a correlated noise source. The angle of incidence of the noise source is calculated using a sound source localization technique based on the geometry of the array. A weighted averaging method was also developed to synthesize multiple channels of data into one single channel. This method preserves the correlated part of the overall signal, while reducing the effects of uncorrelated noise, such as wind.

11:30

2aEA12. On the radiation and wave propagation of sound within horns. Daniel Tengelsen (Dept. of Phys., Brig. Young Univ., N-283 ESC, Provo, UT 84602, danieltengelsen@gmail.com), Vianey Villamizar, Brian E. Anderson, and Timothy W. Leishman (Brigham Young Univ., Provo, UT 84602)

The horns used in loudspeaker systems are well known for their ability to increase radiation efficiency and control directivity. Because of the horn's ubiquitous nature, significant research efforts have been undertaken to explore both its directional and frequency response characteristics. The following research reported in this presentation incorporates several numerical methods (including finite difference and boundary element analysis) to better understand the nature of wave propagation through horns and the effects that geometrical and configurational changes have in better controlling directivity and improving frequency response.

11:45

2aEA13. On the acoustic impedance of a sealed loudspeaker enclosure. Timothy W. Leishman and Xi Chen (Acoust. Res. Group, Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602)

The acoustic impedance of a sealed loudspeaker enclosure is often oversimplified in loudspeaker models, which typically include only rough acoustical characteristics of the enclosed air volume and absorptive fill materials when no driver is present. The acoustical effects produced by sound transmission through and around the rear driver elements (e.g., the frame, magnet, voice-coil former, spider, pole-piece vent, etc.) may also be important for certain loudspeakers and applications. This presentation explores these effects through a discussion of enhanced modeling possibilities and the introduction of measurement techniques that may be used to assess the acoustic impedance of a loudspeaker enclosure when the driver is in place. The impedance more closely represents that actually seen by the rear portion of the driver diaphragm.

TUESDAY MORNING, 27 OCTOBER 2009 REGENCY EAST 1, 10:00 A.M. TO 12:00 NOON

Session 2aED

Education in Acoustics: Hands-on Experiments for High School Students

Uwe J. Hansen, Chair
Indiana Univ., Dept. of Chemistry and Physics, Terre Haute, IN 47809

Approximately 20 acoustics demonstrations will be set up, ranging in complexity from simple resonance on a string to ultrasonic levitation. Around 40 local high school students will perform these experiments with help from Acoustical Society of America (ASA) scientists and student members of ASA. Regular ASA conference participants are welcome as long as they do not interfere with student experimentation.

TUESDAY MORNING, 27 OCTOBER 2009 LIVE OAK, 9:00 TO 10:15 A.M.

Session 2aMUa

Musical Acoustics: Acoustics of Free-Reed Instruments I

James P. Cottingham, Chair
Coe College, Physics Dept., Cedar Rapids, IA 52402

Invited Papers

9:00

2aMUa1. Blown-closed free reeds with and without pipe resonators. James P. Cottingham (Phys. Dept., Coe College, Cedar Rapids, IA 52402, jcotting@coe.edu)

The Asian mouth-blown free reed instruments are of ancient origin and use a symmetric free reed coupled to a pipe resonator. The reed behaves as a blown-open or outward striking reed, with playing frequency below both the resonant frequency of the pipe and the natural frequency of the reed. Although these instruments were known in Europe when the Western free reed family originated about 200 years ago, it does not appear that the mechanism used in the Western instruments was copied from them. In Western free reed instruments the reed tongue is offset from the opening in the frame, permitting operation on only one direction of air flow. Pipe resonators are not required and generally not used. If one of these reeds is coupled with a pipe, the sounding frequency can, within certain

limits, be pulled to match the pipe frequency, and it behaves as a blown-closed or inward striking reed, with playing frequency below both the resonant frequency of the pipe and the natural frequency of the reed. This paper summarizes recent experimental research on the blown-closed free reed, with or without coupling to a pipe resonator, with emphasis on significant differences between the two situations.

9:30

2aMUa2. Free reeds coupling with either a vocal tract or rather small pipe. Laurent Millot (IDEAT UMR 8153, CNRS, Univ. Paris 1, ENS Louis-Lumière, 7 allée du Promontoire, BP 22, F-93161 Noisy-le-Grand Cedex, France, l.millot@ens-louis-lumiere.fr)

Free reed model used is based on a precise description of flow passing throw, the section defined between the shallot and the bended reed, and on the modelling of reed as a clamped-free cantilever beam. Classical description of loading pipe is based on wave reflection function, not designed to be excited by a flow but backward- and forward-traveling plane waves, so the free reed coupling with such a pipe model is physically irrelevant. Then, it is shown that an acoustical flow description is needed, and simple, for at least rather short pipe, giving more realistic sound simulations. It is then explained why a vocal tract description, physically relevant, is possible using only acoustical flow modeling for each element constituting the vocal tract model. In fact, flow acoustical modeling for elements is linked to electroacoustical analogies derived not as a low-pass approximations of input impedance but using the approximation of flow behavior for each element, valid for the whole audio range rather than only the low-pass register. Theoretical explanations should be given to explain why wave paradigma may be irrelevant and sound numerical simulations would be given for each coupling situation considered.

Contributed Paper

10:00

2aMUa3. Acoustical curiosities of an American reed organ. Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, One Coliseum Dr., Univ., MS 38677)

Free reed organs of the 19th and early 20th centuries have long provided interesting pursuits involving their acquisition, restoration, history and musicology, as well as performance, and for some—acoustical studies of the instruments. There is a preservation society with a journal, and the literature includes a treatise, books, and learned articles. Some acoustical curiosities of one such instrument are described, including stop types, reed spectra, waveforms, loudness, and intonation.

TUESDAY MORNING, 27 OCTOBER 2009 LIVE OAK, 10:30 TO 11:15 A.M.

Session 2aMUb

Musical Acoustics: Acoustical Measurements on Musical Instruments

Daniel O. Ludwigsen, Chair
Kettering Univ., Physics Dept., 1700 W. Third Ave., Flint, MI 48504

Contributed Papers

10:30

2aMUb1. Physics of water crotales. Randy Worland (Dept. of Phys., Univ. of Puget Sound, 1500 N. Warner, Tacoma, WA 98416-1031, worland@ups.edu)

Contemporary composers writing for percussion often incorporate unconventional playing techniques with the use of traditional instruments. Among these extended techniques is the lowering of gongs and crotales into water as they are being struck, resulting in a unique glissando effect that involves changes in both pitch and timbre. The orchestral crotale has a relatively simple geometry and overtone structure, making it an appropriate starting point for the study of this performance technique. Building on previously published work on the physics of crotales [B. M. Deutsch, C. L. Ramirez, and T. R. Moore, J. Acoust. Soc. Am. **116**, 2427–2433 (2004)], an experimental investigation of the overtone frequencies and frequency ratios of a crotale as a function of water depth is presented. Mode images created using electronic speckle-pattern interferometry are also shown.

10:45

2aMUb2. The auger shell whistle. Daniel Zietlow and Thomas Moore (Dept. of Phys., Rollins College, 1000 Holt Ave., Winter Park, FL 32789, dzietlow@rollins.edu)

We report on investigations of a musical instrument made from the shell *Terebra Turritella*, or auger shell. To our knowledge, this instrument has been developed only recently. The instrument is played from the large end using an edge tone, and a diatonic scale is achievable merely by drilling five tone holes in a straight line down the length of the spiral shell. It is surprising that a naturally occurring cavity can produce a diatonic scale, and implies that there is something special about the shape of the auger shell. However, experimental and theoretical investigations reveal that the auger shell does not actually have such a resonance structure.

11:00

2aMUb3. Why using complex stimuli for acoustical measurements may be necessary to get physically relevant analysis results. Laurent Millot (IDEAT UMR 8153, CNRS, Univ. Paris 1, ENS Louis-Lumière, 7 allée du Promontoire, BP 22, F-93161 Noisy-le-Grand Cedex, France, l.millot@ens-louis-lumiere.fr)

Classical methods for measurements and analysis in Musical Acoustics (and also Acoustics) rely on Fourier transforms, linear excitation superposition assumption, and monochromatic excitation or noise equivalents. It is shown that classical methods using measurements at a single point (input impedance measurement, for instance) may not be physically relevant. Indeed, only a time sampling and related time Fourier transforms are performed while physical phenomena are spatio-temporal. Even if one considers a monochromatic forward traveling wave excitation, the use of the

dispersion relation is not valid in spatio-temporal or spatio-frequency parameter space. It is shown that measurements need to be performed using also a spatial mapping to be physically relevant which implies at least Fourier bi-transforms (plane case). Then, it may be more simple and physically relevant to perform measurements with the whole complex excitation and use a frequency subband analysis tool, with a perceptive frequency subband mapping. Such a tool is described and it is shown that is permits the real-time switch between listening of partial to total resynthesis for any studied measurement signal, providing great surprises for the physical relevant subband(s) and the perceptive relevant ones. Audio performance will be given.

TUESDAY MORNING, 27 OCTOBER 2009 REGENCY EAST 3, 8:15 TO 11:50 A.M.

Session 2aPA

Physical Acoustics and Biomedical Ultrasound/Bioresponse to Vibration: 40th Anniversary of the Khokhlov-Zabolotskaya (KZ) Equation

Vera A. Khokhlova, Cochair
Moscow State Univ., Acoustics Dept., Leninskie Gory, 119992, Moscow, Russia

Mark F. Hamilton, Cochair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin, TX 78712

Chair's Introduction—8:15

Invited Papers

8:20

2aPA1. Historical aspects of the Khokhlov–Zabolotskaya equation and its generalizations. Oleg V. Rudenko, Vera A. Khokhlova (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Moscow 119991, Russia, rudenko@acs366.phys.msu.ru), and Mark F. Hamilton (The Univ. of Texas at Austin, Austin, TX 78712-0292)

Derivation of the Khokhlov–Zabolotskaya (KZ) equation provided a new approach to describing the combined effects of nonlinear propagation and diffraction in sound beams. In this paper, historical aspects of the KZ equation and its generalizations are presented. The interest in nonlinear acoustic beams of Academician Khokhlov and his colleagues at Moscow State University was inspired in the 1960s by emerging developments in laser physics and the corresponding models of nonlinear optical beams. The two cases, acoustical and optical, represent two limiting cases of nonlinear beams in weakly and strongly dispersive media, respectively, which required different theoretical approaches. The KZ equation and analogous nonlinear evolution equations of nonlinear wave physics are reviewed. It is illustrated how theoretical studies combined with numerical modeling resulted in predictions of new physical phenomena in nonlinear acoustic beams. Concurrently, newer applications of nonlinear acoustics such as parametric arrays, sonic booms, and medical acoustics stimulated the derivation of generalized KZ-type equations together with analytical and numerical methods to solve them. Modern applications and corresponding generalized KZ-type models that include effects such as frequency-dependent absorption, weak dispersion, scalar and vectorial inhomogeneities of the propagation medium, different orders of nonlinearity, and more accurate description of diffraction are presented.

8:40

2aPA2. Variations of the nonlinear equation for diffracting beams in fluids to study different modes of propagation in elastic media. Evgenia A. Zabolotskaya (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Four variations of the equation for nonlinear acoustic beam propagation derived originally for a fluid will be reviewed. These variations describe finite amplitude beam propagation in isotropic elastic solids and in crystals, influence of diffraction on nonlinear Rayleigh waves, and nonlinear shear wave beams. The distinguishing features of each case will be discussed. The equation for nonlinear longitudinal wave beams in isotropic solids is based on nonlinear theory of elasticity and is close to the original equation for fluids. The equation for sound beams in crystals takes into account that the direction of energy propagation does not coincide with the direction normal to wavefronts. Nevertheless, the general form of the equation is similar to that of the original equation for fluids. Distinguishing features of Rayleigh waves are that they are two-dimensional and their nonlinearity is nonlocal. Because of the nonlocal nonlinearity, the evolution equation for this case was derived in the frequency domain using Hamiltonian formalism. The equation for nonlinear shear wave beams was derived in the cubic approximation in terms of particle displacement and it accounts for any polarization. Solutions of these evolution equations will be presented to illustrate phenomena specific to each case.

9:00

2aPA3. Group analysis of the Khokhlov–Zabolotskaya type equations. Oleg A. Sapozhnikov (Dept. of Acoust., Phys. Faculty, Moscow State Univ., Leninskie Gory, Moscow 119991, Russia, and Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, olegs@apl.washington.edu)

Khokhlov–Zabolotskaya (KZ) equation is the basic equation of the theory of acoustic beams propagating in quadratically nonlinear media. Being generalized for the case of a medium with general-type nonlinearity and expressed in dimensionless notation, it has the form $[u_z+P(u)u_\tau]_\tau=u_{xx}+u_{yy}$, where u represents waveform, τ is retarded time, z is axial coordinate, x and y are lateral coordinates, and $P(u)$ characterizes a nonlinear addition to the linear wave velocity: quadratic nonlinearity corresponds to $P(u)=u$, cubic one to $P(u)=u^2$, etc. No general analytical solution has been obtained for this nonlinear KZ-type equation, so numerical methods or asymptotic approximations are usually employed. Additional powerful mathematical tool is Lie group analysis that enables to find general symmetries of differential equations. These symmetries help to generalize known analytical and numerical solutions, derive new solutions, and obtain conservation laws. Results of group analysis of KZ and KZK equations are discussed for various nonlinearities $P(u)$. Examples of obtaining of new solutions and deriving of reduced equations are presented. It is also shown that the generalized KZ equation can be written in Euler–Lagrange form, which makes it possible to apply Nöther theorem and derive new conservation laws for nonlinear acoustic beams.

9:20

2aPA4. Nonlinear acoustic wave propagation in inhomogeneous moving media. Philippe Blanc-Benon (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 69134 Ecully Cedex, France, philippe.blanc-benon@ec-lyon.fr), Mikhail V. Averiyanov (Moscow State Univ., Moscow 119991, Russia), Robin O. Cleveland (Boston Univ., Boston, MA 02215), and Vera A. Khokhlova (Moscow State Univ., Moscow 119991, Russia)

Extensive theoretical analysis, numerical studies, and both large-scale and laboratory-scale experiments have been dedicated to the problem of shock wave propagation in air during recent years. The current interest is motivated by supersonic civil transport which is necessarily affected by problems of sonic boom propagation in the atmosphere. The high-amplitude shock wave generated by a supersonic aircraft propagates through the atmosphere toward the ground and generates an acoustic field with non-uniform pressure distribution. Temporal characteristics and spatial structure of the sonic boom are influenced by aircraft trajectory, nonlinear effects, and diffraction and scattering by inhomogeneities. We review recent results from various teams based on a generalized KZK-type equation that includes the effects of a moving inhomogeneous media. Statistical analysis of the numerical solutions is performed, and the results are compared to experimental data obtained in the controlled laboratory-scale experiments conducted in the Ecole Centrale de Lyon anechoic wind tunnel.

9:40

2aPA5. Acoustic dissipation and finite-amplitude sound propagation in two-phase porous media. N. I. Pushkina, J. I. Osypik, and Ya. M. Zhileikin (Sci. Res. Comput. Ctr., Moscow State Univ., Vorobyovy Gory, Moscow 119992, Russia, n.pushkina@mererand.com)

Acoustic spectroscopy is known to be an important tool for studying various media. In the presented work, specific features of acoustic dissipation in two-phase porous media are theoretically studied and the propagation of a finite-amplitude attenuating sound beam described by the KZK-type equation is analyzed numerically for the case of a marine sediment. The KZK-type equation has been derived from the classical Biot equations, and it contains a dissipative operator corresponding to the frequency correction function $F(\kappa)$ in the Biot equations. In the present work, the properties of the correction function are studied and its well applicable representations are obtained. It is shown that the expansion of the correction function over κ^2 converges at $\kappa<5$. An asymptotic expansion of this function is obtained at large κ values. For high frequencies, simple dependence of viscous attenuation and phase velocity on parameters of a medium, useful for diagnostics of these parameters, has been found. The propagation of a finite-amplitude acoustic beam in a dissipative marine sediment has been numerically analyzed which showed that intense acoustic beam propagation can be accompanied by considerable non-linear phenomena while diffraction only weakly affects the process.

10:00—10:15 Break

10:15

2aPA6. Toward a better understanding of high intensity focused ultrasound therapy using the Khokhlov–Zabolotskaya–Kuznetsov equation. L. A. Crum, M. S. Canney, M. R. Bailey (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), O. V. Bessonova (Moscow State Univ., Moscow 119991, Russia), and V. A. Khokhlova (Univ. of Washington, Seattle, WA 98105)

High intensity focused ultrasound (HIFU) therapy is an emerging medical technology in which acoustic pressure amplitudes of up to 100 MPa are used to induce tissue ablation, often in combination with real-time imaging. The ultrasound energy is typically focused into a millimeter-size volume and used to thermally coagulate the tissue of interest while ideally sparing surrounding tissue. Nonlinear effects are important in HIFU as *in situ* intensities for clinical applications of up to 30 000 W/cm^2 have been reported. Since controlled experiments are often difficult to perform, especially *in vivo*, modeling can aid in understanding the physical phenomena involved in HIFU-induced tissue ablation. The Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation is applicable to HIFU because it includes all of the basic physical phenomena that are relevant to HIFU including acoustic beams, diffraction, focusing, nonlinear propagation, shock formation, and dissipation. In this paper, an overview of several recent advances in KZK modeling for HIFU applications are described. It is shown that shock-induced heating in tissue can cause localized boiling in milliseconds; furthermore, the bubbles associated with boiling can significantly alter HIFU treatments. [Work supported in part by NSBRI SMST01601, NIH EB007643, and RFBR 09-02-01530.]

10:35

2aPA7. Shocking stones with the Khokhlov–Zabolotskaya–Kuznetsov equation. Robin O. Cleveland (Dept. of Mech. Engin., Boston Univ., 110 Cummington St, Boston, MA 02215, robinc@bu.edu)

In shock wave lithotripsy (SWL) high-amplitude acoustic waves generated outside the body are focused onto kidney stones in order to fragment them into pieces that are small enough to pass naturally. In order to investigate mechanisms of fragmentation and collateral damage to the soft tissue it is necessary to understand the acoustic field delivered to the tissue in and around the kidney. This is not easily accomplished experimentally and motivates the development of numerical models. The application of the Khokhlov–Zabolotskaya–Kuznetsov (KZK) equation to the SWL problem is discussed as there are two underlying assumptions that are challenged: the peak pressures are on the order of 50 MPa resulting in an appreciable acoustic Mach number of 0.02, and large focusing gains are employed which violate the paraxial approximation. Predictions using the KZK equation, with a layered model of the tissue path to the kidney, demonstrate that the waveform shape, in particular, the risetime, is strongly affected by the tissue path. However, once the waveform enters the urine in the collecting space of the kidney, the waveform heals and a sharp shock results within about 5 mm of propagation. [Work supported in part by NIH DK-43881.]

10:55

2aPA8. Nonlinear pulsing schemes for diagnostic ultrasound. Michalakis Averkiou (Dept. of Mech. Eng., Univ. of Cyprus, 75 Kallipoleos Str., 1678 Nicosia, Cyprus, maverk@ucy.ac.cy)

With the introduction of ultrasound contrast agents and the development of tissue harmonic imaging (THI), nonlinear acoustics has become a major research direction in diagnostic ultrasound. In THI an image is formed from the nonlinear components (due to nonlinear propagation) of the backscattered signal. Pulsing schemes have been invented to specifically detect the nonlinear components. These schemes are pulse inversion (PI), power modulation (PM), and their combinations (PMPI2, PMPI3, and PMPI4). The KZK equation is used to investigate these pulsing schemes in conditions that closely resemble ultrasound imaging in order to fully understand the properties of these schemes. Measurements of nonlinear propagation of such pulses were compared with the KZK predictions and were found in good agreement. Pulse inversion isolates the even harmonic components only and extracts the *total* amount of nonlinearity in those components. With pulsing schemes with different amplitudes (PM, PMPI2, and PMPI3), the *differential* nonlinearity between pulses is detected, with PMPI2 extracting the largest total amount. In addition, the schemes with amplitude modulation have a *nonlinear fundamental* component that suffers less than the higher harmonics from thermoviscous absorption and thus offers more signal in penetration limited cases.

11:15

2aPA9. Statistical solutions of the Khokhlov–Zabolotskaya–Kuznetsov equation for analyzing mechanisms of image quality enhancement in tissue harmonic imaging. Xiang Yan (Dept. of Mech. Eng., The Univ. of Texas at Austin, Austin, TX 78712-0292) and Mark F. Hamilton (The Univ. of Texas at Austin, Austin, TX 78713-8029)

Statistical solutions of the Khokhlov–Zabolotskaya–Kuznetsov equation were obtained to analyze mechanisms contributing to image quality enhancement in tissue harmonic imaging (THI). The focus is on suppression of image clutter due to phase aberration and reverberation. Tissue heterogeneity is modeled with a random phase screen characterized by its variance and spatial correlation length. Solutions for the mean intensities of the linear (fundamental mode) and second-harmonic fields were derived from a focused Gaussian beam that is transmitted through a phase screen located an arbitrary distance from the source. The random phase variations of the screen are assumed to be small and described by a Gaussian autocorrelation function. The solutions are validated by comparison with ensemble averages of direct numerical simulations. A benefit of the analytical approach is separation of the different contributions to deformation of the beam by the phase screen, and the statistical approach is convenient for quantifying the merits of THI. The degree to which THI reduces beam deformation is assessed using a measure based on signal-to-clutter ratios introduced previously for this purpose by C. E. Bradley [Proceedings of 17th International Symposium on Nonlinear Acoustics (AIP, New York, 2006)]. Statistical solutions will also be presented for backscattering. [Work supported by NIH DK070618.]

Contributed Paper

11:35

2aPA10. Nonlinear acoustical beam formation and beam profiles in fluids. Cristian Pantea and Dipen N. Sinha (Mater. Phys. and Applications, MPA-11, MS D429, Los Alamos Natl. Lab., Los Alamos, NM 87545)

Nonlinear acoustical beam formation in fluids is investigated in several different configurations: (i) collinear mixing due to simultaneous excitation of a single transducer by two different high frequency signals, (ii) noncollinear mixing due to excitation from two separate transducers, (iii) nonlinear de-modulation of an amplitude modulated fixed frequency, and (iv) nonlinear mixing due to excitation of a single transducer by a fixed frequency and a chirped frequency signal. In all cases, the difference frequency acoustic beam generated by nonlinear mixing in the fluid is investigated. In contrast to using solids as the nonlinear medium, fluids have the advantage that the beam can be scanned in all directions allowing the determination of three-dimensional beam profile. The experiments were performed mainly in water and Fluorinert FC43. In the collinear measurement configuration, where the nonlinear down-converted beam is produced from a single transducer, the experimental results are compared with the predictions from the KZK equation. Electronics and transducer nonlinearity and the role of transducer "effective diameter" will also be discussed.

Session 2aPP

Psychological and Physiological Acoustics: Cognitive Aspects of Hearing

Craig A. Champlin, Chair
Univ. of Texas at Austin, Communication Sciences and Disorders, 1 University Station, Austin,TX 78712-0114

Contributed Papers

8:15

2aPP1. Signal-to-noise-ratio loss in hearing-impairment and feature enhancement. Woojae Han (Dept. of Speech and Hearing Sci., Univ. of Illinois at Urbana-Champaign, 901 S. Sixth St., Champaign, IL 61820, whan5@illinois.edu), Riya Singh, and Jont Allen (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The primary purpose of this study is to investigate detailed consequences of hearing loss and to scrutinize if individual hearing impaired (HI) listeners exhibit individual patterns in their consonant confusion scores. 16 English nonsense CV syllables, by 18 various talkers in five speech-weighted noise (−12, −6, 0, +6, and +12-dB signal-to-noise-ratios) and quiet conditions, were randomly presented to 11 HI subjects (18 HI ears) having sensorineural hearing losses at two different amplified levels: with and without NAL-R (similar to fitting hearing aids). The HI ears have sensorineural hearing loss with flat (mild-to-moderate-to-severe), ski-slop at low- and/or high-frequency regions, and cookie-bite configurations. To characterize the consonant loss, the performance of each HI ear was compared to average normal-hearing listeners in masking noise. Preliminary analysis shows that each HI ear has an individual profile, which is a characteristic of the impaired ear. Once the noise-sensitive consonants are identified, the aim is to use signal processing techniques to selectively enhance features to make them audible in noise. [Work supported by the NIH Grant RDC009277A.]

8:30

2aPP2. Comodulation masking release to detect cochlear dead regions in hearing impaired ears. Riya Singh (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, Everitt Lab, 1406 W. Green St., Urbana, IL 61801), Woojae Han, and Jont B. Allen (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The purpose of this study is to investigate if comodulation masking release experiments with hearing impaired (HI) individuals can be used as a reliable method to detect dead regions in the cochlea, if any. The experiment involves detection of a pure tone in a narrow band of noise (the on-signal band) in presence of four other narrow bands of noise that fall outside the auditory bandwidth at the center frequency of the on-signal band. These flanking bands are either all co-modulated so as to have the same amplitude envelope as the on-signal band or are randomly modulated and are presented monaurally to an impaired ear. Each band is level adjusted to be audible to the HI ear. Data are being collected for both these cases from normal-hearing listeners and HI listeners with no dead regions. Every HI individual also undergoes psychophysical tuning curve and threshold-equalizing noise tests which are currently used tests to detect cochlear dead regions. Each HI individual also takes speech perception tests with CV syllables under various noise conditions. The consonant loss profiles from the perception experiments help make a reasonable estimate of the possibility of a dead region. [Work supported by NIH.]

8:45

2aPP3. Infants' vowel discrimination in modulated and unmodulated noise. Lynne A. Werner (Dept. Speech & Hearing Sci., Univ. of Washington, 1417 NE 42nd St., Seattle, WA 98105-6246)

Adults detect and discriminate masked sounds better in temporally modulated maskers than in unmodulated maskers. It is not known whether the same is true of infants. In this study, 7–9-month-olds learned to respond when a repeated vowel changed from /a/ to /i/ or from /i/ to /a/ in speech spectrum noise. Discrimination was assessed using an observer-based procedure. In one condition the noise was unmodulated. In the other, the noise was modulated with the envelope of single talker speech. The level of the noise was 60-dB SPL. For infants in group 1, the level of the vowel in both conditions was set so that $d'=1$ was achieved in the unmodulated noise; discrimination in modulated noise was the second condition tested. For infants in group 2, the level of the vowel was fixed at the average level used to test the infants in group 1; discrimination in modulated noise was tested first. In both groups, d' in the modulated noise condition averaged only 0.5. Thus, infants were unable to take advantage of masker modulation to improve speech discrimination. Their speech discrimination was poorer in modulated than in unmodulated noise. [Work supported by NIH R01 DC00396 and P30 DC04661.]

9:00

2aPP4. Enduring non-verbal memory for spectral timbre. Denis McKeown and Tom Mercer (Dept. of Psych., Univ. of Leeds, Leeds, United Kingdom)

It is often assumed that auditory sensory memory disappears within a few seconds or is so smudged as not to permit fine discriminations between successive sounds beyond a few seconds. But studies have been confounded by proactive interference, task difficulty, and verbal labeling. In Experiment 1, participants heard two complex tones separated by a silent retention interval of 2, 4, 8, 16, or 32 s. Tones varied widely in pitch from trial to trial (so no standard memory could be formed) and were non-speech-like (difficult to verbalize), and very long inter-trial intervals were used (32 s) to reduce interference from prior sounds. After the tone-pair presentation, listeners made a same-different response based on slight differences in spectral timbre. Discrimination performance expressed as d', at each retention interval revealed that the persistence of memory for spectral timbre surpassed the usually quoted lifetime of some few seconds and for certain participants it endured for 32 s. In Experiment 2, overall stimulus levels were roved within trials to rule out loudness cues in the discrimination. It is proposed that reasonably robust non-verbal memory traces for spectral timbre persist for several seconds allowing listeners to make fine discriminations of changes to a spectrum.

9:15

2aPP5. Efficient coding of correlated complex acoustic dimensions through active listening. Christian E. Stilp, Timothy T. Rogers, and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, cestilp@wisc.edu)

Efficient coding extracts and exploits redundancy to optimize information processing in sensorineural systems. Stilp *et al.* [J. Acoust. Soc. Am. **124**, 2496 (2008)] reported evidence for rapid and efficient adaptation to correlation among complex acoustic attributes. In the study, following passive exposure to highly correlated stimulus features, discriminability of sound pairs violating the correlation is temporarily lost before subsequent recovery via active testing with stimuli whose features were poorly correlated. The present study examines listeners' ability to extract and exploit correlation between stimulus attributes exclusively through active testing. Listeners discriminated stimuli (AXB) for which two complex, independent dimensions, attack/decay (AD) and spectral shape (SS), were

highly correlated (r^2=0.96). In the first testing block of 128 trials, discrimination of sound pairs respecting the correlation is superior to sound pairs violating the correlation. Only through successive testing blocks, listeners discover variance orthogonal to the otherwise perfect correlation between AD and SS, and discrimination recovers to baseline levels. Listener performance will be discussed within the context of maximum likelihood and connectionist models. [Work supported by NIDCD.]

9:30

2aPP6. Neural representation of speech sounds: Study using frequency following response. Radhika Aravamudhan (George S. Osborne College of Audiol., Salus Univ., 8360 Old York Rd., Elkins Park, PA 19027, raravamudhan@salus.edu)

The neural encoding of an acoustic signal begins in the auditory nerve and travels to the auditory brainstem and further to the auditory cortex. Previous studies have used nonspeech signals like tones and clicks to evaluate the integrity and synchrony of the auditory pathway. Most of the previous research has focused on how any acoustic signal is perceived by using behavioral methods. One of the main areas in understanding how we hear the signal depends on how this acoustic signal is represented in the auditory pathway. A number of studies have studied the acoustic parameters that influence the perception of speech, but very few of them have focused on how these acoustic changes are represented in the auditory pathway. Since the signal representation in the auditory pathway is very crucial to how the signal is perceived, studies that focus on the relationship between the changes in the input acoustics and its influence on neural representation become very essential. In the current project the neural representation of isolated vowels and vowels in CV context was recorded using human frequency following response. The results will be discussed in the paper.

9:45

2aPP7. How can a video game cause panic attacks? I. Effects of an auditory stressor on the human brainstem. Judith L. Lauter, Elizabeth Mathukutty, and Brandon Scott (Dept. of Human Services, Stephen F. Austin State Univ., Box 13019 SFA Station, Nacogdoches, TX 75962, jlauter@sfasu.edu)

The auditory brainstem response (ABR) was recorded during simultaneous binaural presentation of two types of sounds: (1) condensation clicks presented through in-the-ear earphones at 43.1/s, 60 dB nHL; and (2) recordings of breathing sounds, presented through supra-aural headphones, at levels adjusted by participants to be equivalent to the clicks. In alternate blocks, the breathing sounds were either: (1) a recording of quiet breathing (blocks 1, 3, and 5) or (2) a recording of erratic (stressed) breathing (blocks 2 and 4). The erratic breathing was modeled on a video game soundtrack in which the character was represented as running, wounded, and frightened. Four 2048-sweep ABR waveforms were collected in each of the five blocks, and the mean amplitude of ABR peak V was calculated over each set of four waveforms. Results indicate a significant decrease in the amplitude of ABR peak V during erratic breathing versus quiet breathing. Implications include (1) possible new evidence of the effect of selective attention on the ABR, (2) the potential for using auditory stressors to study the central physiology of emotional responses in humans, and (3) clues to physiological correlates of the effects of certain video games known to evoke panic attacks in susceptible players.

TUESDAY MORNING, 27 OCTOBER 2009 REGENCY WEST 1 & 2, 9:00 TO 11:50 A.M.

Session 2aSC

Speech Communication: Advances in Speech Synthesis

Norma S. Antonanzas-Barroso, Chair
UCLA School of Medicine, Head and Neck Surgery, 31-24 Rehab Center, Los Angeles, CA 90095-1794

Invited Papers

9:00

2aSC1. Talking heads: Speech synthesis and embodied cognition. Philip Rubin, Gordon Ramsay (Haskins Labs., 300 George St., New Haven, CT 06511 and Yale School of Medicine, 333 Cedar St., New Haven, CT 06510, rubin@haskins.yale.edu), and Eric Vatikiotis-Bateson (Univ. of British Columbia, Vancouver, BC V6T 1Z4, Canada, evb@interchange.ubc.ca)

This presentation provides a brief overview of 300 years of effort toward the creation of talking heads: mechanical, electronic, and/or computational models of human speech. Speech, language, communication, and cognition are fundamentally shaped, in part, by both biological and physical factors. To understand this grounding and how to effectively replicate its most salient aspects in synthesis systems requires us to pay serious attention to the structure, kinematics, and dynamics of the articulators; the organization and characterization of complex, emergent behavior in multimodal systems; and the consideration of how events within such systems unfold over multiple time scales. New approaches that will help advance our knowledge and improve our synthesis tools and techniques will be discussed.

9:20

2aSC2. Challenges in evaluating the intelligibility of text-to-speech. Ann Syrdal (AT&T Res., 180 Park Ave., Rm. D159, Florham Park, NJ 07932-0971, syrdal@research.att.com), Murray Spiegel (Telcordia Technologies, Piscataway, NJ 08854-4151), Deborah Rekart (AT&T Services, Dallas, TX 75287), Susan R. Hertz (Nova Speech LLC and Cornell Univ., Ithaca, NY 14850), Tom Carrell (Univ. of Nebraska, Lincoln, NE 68583-0738), H. Timothy Bunnell (Alfred I duPont Hospital for Children, Wilmington, DE 19803), and Corine Bickley (Gallaudet Univ., Washington, DC 20002)

Text-to-speech (TTS) technology imposes different constraints on intelligibility than those sufficient for the evaluation of other speech communication systems. For example, the newly revised standard S.2-2009 explicitly excludes TTS from the speech communication systems it covers. Since there is no current standard appropriate for evaluating TTS intelligibility, the ASA Standards Bioa-

coustics (S3) working group on Text-to-Speech Technology (WG91) was formed with the initial goal of developing such standard. We describe several ways in which standard methods of testing speech intelligibility are unsuitable for TTS technology and outline our approach to overcoming these limitations. We present an overview of our proposed standard, which is currently nearing its final draft stages.

9:40

2aSC3. Recombinant speech synthesis: Natural text-to-speech synthesis with prosodic control. Esther A. Klabbers, Taniya Mishra, and Jan P.H. van Santen (Div. of Biomedical Comput. Sci., Oregon Health & Sci. Univ., 20000 NW Walker Rd., Beaverton, OR 97006, klabbers@cslu.ogi.edu)

Unit selection text-to-speech synthesis methods rely on large corpora to cover all phoneme sequences in as many prosodic contexts as possible. This coverage is rarely complete except in limited domains. This becomes particularly salient when using prosodic markup to generate specific prosodic patterns (e.g., emphatic stress). An architecture is proposed combining the naturalness of unit-selection synthesis with the requirement of prosodic control. The speech corpus consists of multiple sub-corpora, each optimized to cover a "linguistic subspace"; subspaces include phoneme sequences, left-headed feet, sentence structures, and paralinguistic categories. The system relies on the superpositional model of intonation to decompose natural pitch contours into component contours, e.g., phrase curves (corresponding to phrases) and accent curves (corresponding to left-headed feet); on analogous methods for timing; and on hybridization methods to implement paralinguistic features. During synthesis, phoneme sequences, curves, and parameters are generated from the sub-corpora, optionally modified as per prosodic control tags, and "re-combined." The explicit representation in terms of component curves allows for complete prosodic control, while the naturalness of the prosodic patterns is guaranteed by extracting these curves from natural speech and smoothly modifying them, thereby preserving important natural detail. [Research supported by NSF grant 0205731, "Prosody generation in child-oriented speech."]

10:00—10:30 Break

10:30

2aSC4. Speech transformation: Increasing intelligibility and changing speakers. Alexander Kain (Div. of Biomedical Comput. Sci. Oregon Health & Sci. Univ., 20000 NW Walker Rd., Beaverton, OR 97206, kaina@ohsu.edu)

Speech transformation changes an aspect of speech without changing its message, typically using a process of analysis, feature modification, and synthesis. The feature modification step can consist of a (trainable) mapping function or a hybridization of several feature sets for the purposes of perceptual experimentation. First, speech transformation approaches are presented that aim to increase the intelligibility of speech. One approach is used in the context of increasing the intelligibility of conversationally spoken speech for hearing-impaired listeners. Another approach aims to increase the intelligibility of speaking-impaired individuals by the general population. Second, transformation approaches are presented which aim to change a source speaker's speech (natural or TTS-generated) to sound as if a specific target speaker had spoken it (also known as voice transformation).

10:50

2aSC5. Quasi-articulatory synthesis as a tool for basic science and education. Helen M. Hanson (ECE Dept., Union College, 807 Union St., Schenectady, NY 12308)

HLsyn is a quasi-articulatory speech synthesis system based on Klatt's formant synthesis model [D. K. Klatt and L. C. Klatt, J. Acoust. Soc. Am. **87**, 820–857 (1990)]. The elements of a circuit model of the vocal tract are derived from cross-sectional areas of constrictions and subglottal pressure. Solution of the circuit results in pressure drops across constrictions, which in turn lead to source characteristics (both periodic and noise). The circuit model imposes constraints on the relations between the pressures and flows in the circuit, resulting in more natural variation in the amplitudes of two simultaneous sources, e.g., a periodic source and a noise source. While the virtues of HLsyn as a speech synthesis system have often been sung [e.g., Hanson and Stevens, J. Acoust. Soc. Am. **112**, 1158–1182 (2002)], HLsyn has other uses that may go unnoticed. Using HLsyn to explore or simulate the production of certain sounds often leads to insights about how speech is produced. That is, this synthesizer can be used to explore the basic science of speech production and perception. Likewise, HLsyn as a tool in basic speech science courses brings equations and circuits to life for students. Examples of insights gained through HLsyn will be described. [Work supported by NIH.]

11:10

2aSC6. Advances in simulation of sentence-level speech production with kinematic models of the vocal tract and vocal folds. Brad H. Story (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, 1131 E. 2nd St., Tucson, AZ 85721)

Speech simulation refers to a system in which computational models are used to simulate the physical processes of human sound production. These processes consist primarily of vocal fold vibration and acoustic wave propagation in the tracheal, nasal, vocal tract systems, and radiated acoustic output. Although simulations of various aspects of speech production are typically limited to short time scales, this presentation will focus on using kinematic models of the vocal folds and vocal tract shape to generate speech at the level of words and sentences. Specifically, the vibrating medial surface of the vocal folds is represented by a kinematic model which allows for direct time-dependent specification of fundamental frequency, amplitude of vibration, glottal geometry, and glottal symmetry. The vocal tract shape is also governed by a kinematic model, based on data collected with MRI and x-ray microbeam techniques, that allows for specification of time-dependent cross-sectional area changes in an acoustic waveguide. When these models are coupled, the result is an acoustically-interactive simulation of the sound production process from which pressures and airflows are generated. The components of the system will be presented and then used to demonstrate samples of simulated speech. [Work supported by NIH R01-DC04789.]

11:30—11:50 Panel-Discussion

Session 2aSP

Signal Processing in Acoustics: Time-Frequency Theory and Applications

Edmund J. Sullivan, Chair
EJS Consultants, 46 Lawton Brook Ln., Portsmouth, RI 02871-1032

Chair's Introduction—7:55

Invited Papers

8:00

2aSP1. Wave propagation in phase-space. Leon Cohen (Dept. of Phyiscs, City Univ.-Hunter College, 695 Park Ave., New York, NY 10021)

We derive exact expressions for the spreading of a propagating pulse in a dispersive medium. We address both the deterministic and the random case, and also allow for the possibility of frequency-dependent attenuation. The conditions for contraction of a pulse before it eventually spreads to infinity are derived. In the interpretation of the expressions, we show that considerable insight may be achieved if the pulse is analyzed in phase-space where the phase-space distribution may be time-frequency or position-wavenumber. Applications to a train of pulses relevant to reverberation and clutter is also discussed. [Work supported by ONR 321US.]

8:20

2aSP2. Time-frequency and position-wavenumber acoustic signal analysis. Patrick J. Loughlin (745 Benedum Hall, Univ. of Pittsburgh, Pittsburgh, PA 15261, loughlin@pitt.edu)

The spectrogram, which is a plot of the spectral intensity of a signal over time, has been widely used in acoustic signal processing because many signals, such as speech, animal vocalizations, music, and the sonar backscatter from elastic objects, have frequencies that change over time, which convey important information about the signal or source from which it originated. Since the development of the spectrograph at Bell Laboratories in the 1940s, more modern methods for time-frequency analysis, such as the Choi–Williams and Zhao–Atlas–Marks distributions, have been developed, which overcome some of the limitations of the spectrogram and, in particular can show time-frequency detail in the signal that is obscured by the spectrogram. We will discuss these methods and the general area of time-frequency acoustic signal analysis with examples drawn from a variety of applications. A particular focus will be made on showing how time-frequency analysis, and also position-wavenumber analysis, can be used to formulate and gain physical insights into dispersive pulse propagation. We will also comment on and illustrate the use of wavelets for time-frequency analysis. [Work supported by ONR 321US.]

8:40

2aSP3. Gibbs sampling for modal arrival time and amplitude estimation from time-frequency representations of acoustic signals. Zoi-Heleni Michalopoulou (Dept. of Math. Sci., New Jersey Inst. of Technol., Newark, NJ 07102, michalop@njit.edu)

A Gibbs sampling-maximum *a posteriori* approach is developed for modal arrival time and amplitude estimation from time-frequency representations of broadband acoustic signals propagating in underwater media. The goal is to obtain accurate estimates of arrival times of propagating modes and corresponding amplitudes, which can then be employed for source localization and geoacoustic inversion. The method provides uncertainty information on both modal arrival time and amplitude estimates, typically unavailable when traditional methods are used. Estimates and uncertainty are propagated through an inversion process, generating posterior probability distributions of source range and geoacoustic properties. [Work supported by ONR.]

9:00

2aSP4. Adaptive noise and interference removal from speech based on a modified short time Fourier transform. Douglas J. Nelson (Natl. Security Agency, 9800 Savage Rd., Fort Meade, MD 20755-6214)

A simple linear adaptive method for removing noise and interference from speech is presented. The method is based on a linear time-frequency representation in which the short time Fourier transform (STFT) is concentrated (reassigned) in frequency. The concentration process results in a complex-valued surface in which speech components and narrowband AM/FM interference components are essentially time-varying spectral impulses. Unwanted components may be removed from the surface, and the clean signal with these components removed may be reconstructed as a linear time marginal, computed by integrating the surface with respect to frequency. This process is linear and results in a nearly distortion-free reconstructed signal. Moreover, unlike processes like methods such as spectral subtraction, the process requires no Fourier inversion. This process is computationally efficient since the STFT is implemented as a bank of recursive two tap filters.

9:20

2aSP5. Estimation of seismic interface wave dispersion in geoacoustic inversion. Hefeng Dong (Dept. of Elec. & Telec., Norwegian Univ. of Sci. & Techn., NO-7491 Trondheim, Norway) and Lanbo Liu (Univ. of Connecticut, Storrs, CT 06269-2037)

Shear wave velocity variation as depth in underwater environment is closely related to dispersion of seismic interface waves traveling along the water/sediment boundary. An estimate of the dispersion of seismic interface waves can be obtained by time-frequency analysis. There are different methods for the time-frequency analysis to estimate dispersion of the interface waves. Phase velocity dispersion can be estimated by multi-sensor method, while group velocity dispersion can be estimated by single-sensor method. Multi-sensor method uses array data by knowing sensor-spacing, which gives the average estimate of the phase velocity within the ranges covered by the array. Single-sensor method uses one trace at a time by knowing the distance between source and the sensor, and gives the local group velocity estimate. In this paper, time-frequency analysis methods for both phase- and group-velocity estimations are presented. An experimental example for estimation of shear wave velocity variation as depth in the upper layers of the sediment is presented by inverting the estimated dispersion relations of seismic interface waves [Work supported by NFR under No. 186923/I30].

9:40

2aSP6. Space-time-frequency analysis of the scattered acoustic wave field from elastic shells recorded by bistatic sonar systems. Karim G. Sabra and Shaun D. Anderson (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Graduate Box 1000, Atlanta, GA 30332, karim.sabra@me.gatech.edu)

An ongoing challenge for underwater sonar systems is to discriminate a man made target (shell) from surrounding clutter returns and to provide robust classification features for the estimation of the physical target characteristics (e.g., shell thickness and material properties). To this end, time-frequency analysis, and, in particular, Wigner–Ville analysis, has been shown to provide a robust processing tool for interpreting the evolutional time dependent aspect of the scattered acoustic wave field from elastic shells. The design of a robust space-time-frequency bistatic sonar system to enhance the target detection of shells with the use of a sensor array will be presented. Practical implications for bistatic mine countermeasure sonar systems, using a network of autonomous underwater vehicles, will be discussed.

10:00—10:15 Break

Contributed Papers

10:15

2aSP7. Application of compressive sampling to passive sonar signals. R. Lee. Culver (Grad. Prog. Acoust. & Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804) and N. K. Bose (Penn State Univ., State College, PA 16804)

Recently the compressive sampling (CS) paradigm has generated considerable interest in the signal processing community because it offers the potential to fully characterize signals without satisfying the Nyquist requirement (sampling frequency must be more than twice the highest frequency in the signal). Signal compression itself is not new; it is used in all file compression algorithms. However, it generally requires that all coefficients be generated, many or most of which are discarded and only a few are transmitted. The theoretical basis for CS has been presented in a number of signal processing and statistics journal articles [e.g., Candès and Wakin, IEEE SP (March 2008)]. CS is fixed (non-adaptive) and is efficient in that only the coefficients required for signal characterization are calculated. The key requirement of CS is to find a basis in which the signal representation is sparse and thus can be represented with a minimum number of coefficients. Here we explore possible benefits of applying CS to sonar signals, including signal compression, bandwidth reduction, and exploitation of sparse sampling geometries. [Work sponsored by ONR Undersea Signal Processing.]

10:30

2aSP8. Time-frequency analysis techniques for long range sediment tomography. Gopu R. Potty and James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02879)

Long range sediment tomography technique uses travel-time of acoustic normal modes at different frequencies to invert for the geoacoustic properties. The modal travel-times are calculated from the time-frequency analysis of the acoustic time series. Travel times in the frequency range 1 Hz to less than 1 kHz modes 1–6 obtained using broadband sources are typically used as data for the inversion. Various time-frequency analysis techniques used in the past in the context of long range sediment tomography will be presented. These techniques include short time Fourier transform, wavelet transform, matching pursuit algorithms, and dispersion based short time Fourier transforms. Performance of these techniques will be compared using data from various types of broadband sources deployed during different field experiments. [Work supported by Office of Naval Research.]

10:45

2aSP9. Using time corrected instantaneous frequency to detect source motion of a towed projector. Jack A Shooter (Appl. Res. Labs, Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, shooter@arlut.utexas.edu)

The TCIF algorithm [S. Fulop and K. Fitz, J. Acoust. Soc. Am. 119, 360–371 (2006)] was applied to a narrow band projector from the Shallow Water 2006 Hudson Canyon experiment. To observe Doppler shifts near CPA, short overlapping FFTs were formed from the measured time series data. The appearance of a secondary Doppler component indicates that the source was wobbling or fishtailing during the tow event. Analysis indicated the source was moving back and forth with amplitudes of about 6 in., at a rate of 1–2 cycles/s. Detecting these source dynamics required a basic FFT length less than 18% of the variable period and signal-to-noise ratio greater than 18 dB. Narrowband filtering the data significantly reduced interference from outside the band of interest. The FFT lengths were so short the frequency bin spacing was much greater than the wobble frequency variation, yet TCIF permitted the source wobble period to be observed. [Work supported by IR&D.]

11:00

2aSP10. Temporal analysis of human motion signatures. Alexander Ekimov, Marshall Bradley, and James Sabatier (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, aekimov@olemiss.edu)

Passive and active signals characterize human motion. Passive signals produced by footsteps are recorded with microphones and geophones. Active signals are measured with Doppler techniques and characterize the oscillatory motion of human body parts with different acoustic cross-sections and velocities. Simultaneous measurements of human passive and active signatures show temporal synchronization. This synchronization results in equal cadence frequencies for human motion signatures. These signatures are compared to a mathematical model of human motion [Boulic *et al.*, "A global human walking model with real-time kinematic personification," Vi-

sual Comput. (1990)]. This empirical model predicts the motion of key body parts including shoulders, elbows, fingers, hips, knees, and toes for an "average" human walking at a constant velocity. Speeds of the body parts depend on the walking velocity and the size of the individual as parametrized by thigh height. Comparison of experimental data with the predictions of the mathematical simulations enhances the capability to characterize humans. Measurements of the motion signatures of walking humans using a multimodal sensor suite (seismic, passive, and active ultrasonics, and radar sensors) are presented and discussed. [Work supported by Department of the Army, Army Research Office Contracts W911NF-04-1-0190 and W911NF-08-1-0389.]

11:15

2aSP11. Blind deconvolution of quadratic time-frequency representations of musical signals for reverberation feature extraction. Gang Ren (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627, garen@ece.rochester.edu), Mark F. Bocko, and Dave Headlam (Univ. of Rochester, Rochester, NY 14627)

An acoustic space is uniquely characterized by its reverberant behavior. Due to the complexity of the multiple reflections and diffraction of sound in an enclosure, currently available system identification algorithms cannot effectively estimate the impulse response of a concert hall simply from musical recordings in that space. This paper reports the use of blind image deconvolution methods to construct echo patterns from quadratic time-frequency representations of reverberant recordings of music. First, a quadratic time-frequency analysis is performed to decompose the musical signal into its constituent harmonic components. Quadratic time-frequency analysis methods are known to give enhanced resolution in the time-frequency plane in comparison to conventional Fourier analysis. Reverberant features then appear as blur in the time-frequency plane, which can be estimated by employing the methods of blind deconvolution of this "image" of the sound. The proposed algorithm retrieves both the blur pattern, which corresponds to the reverberation echogram, and the direct acoustic signal. By choosing the time-frequency frame scale and smoothing window, a multi-resolution analysis of the underlying acoustic impulse response is obtained. Various quadratic time-frequency analysis methods are evaluated and their relative performance is reported. The proposed methods are also compared to existing dereverberation algorithms.

11:30

2aSP12. Speech localization in any direction using power and frequency signatures, gradients, and differences. Colin L. Barnhill and James West (ECE Dept., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218, cb@jhu.edu)

Speech localization is a relatively simple task for a human but, often, a difficult task for acoustic arrays. Although arrays can localize impulsive and narrow-band sources through the use of cross-correlation and subspace methods (like MUSIC), these methods break down when there is low SNR, reverberant conditions, or the source is neither impulsive nor narrow-band. Interesting speech sources (for teleconferencing or surveillance) are most often in environments where many of the breakdown conditions exist. New or multiple methods are necessary to reliably localize speech. A new algorithm will be presented that makes use of multiple localization methods to locate the desired speech signal. The localization methods are based on the time and frequency power signatures, gradients, and differences of third order supercardioid beams. Multiple beams are used to spatially segregate power signatures, and the results are clustered to determine position. Using a spherical array, it is possible to localize speech in any direction using these beams. Experimental acoustic and mathematical results in real room situations will be presented.

TUESDAY MORNING, 27 OCTOBER 2009 RIO GRANDE CENTER, 10:20 TO 11:45 A.M.

Session 2aUW

Underwater Acoustics: Reverberation Measurements and Modeling I

Eric I. Thorsos, Cochair
Univ. of Washington, Applied Research Lab., 1013 NE 40th St., Seattle, WA 98105

John S. Perkins, Cochair
Naval Research Lab., Code 7140, Washington, D.C. 20375

Jason D. Summers, Cochair
SAIC, Acoustic and Marine Systems Operation, 10401 Fernwood Rd., Bethesda, MD 20817

Chair's Introduction—10:20

Invited Paper

10:25

2aUW1. Update on the reverberation modeling workshops. John S. Perkins (Naval Res. Lab., Washington, DC 20375, john.perkins@nrl.navy.mil) and Eric I. Thorsos (Univ. of Washington, Seattle, WA 98105-6698)

In order to investigate the status of reverberation modeling and establish well-defined benchmark problems, two reverberation modeling workshops have been jointly sponsored by the Office of Naval Research and the Navy Program Executive Office C4I (PMW 120). The first workshop was held 7–9 November 2006, and the second during 13–15 May 2008. This paper presents the approach used in formulating the reverberation problems posed to workshop participants and shows examples of the modeling agreement obtained for some of the workshop problems. Issues that have arisen as workshop participants have collaborated in their attempts to produce consensus solutions will be pointed out. Finally, some of the problems to be addressed at a proposed third workshop will be discussed.

10:45

2aUW2. Ray versus mode differences in reverberation modeling solutions for environments with high boundary scattering loss. Eric I. Thorsos, Frank S. Henyey, Jie Yang, and Stephen A. Reynolds (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, eit@apl.washington.edu)

Several of the problems for the first Reverberation Modeling Workshop yielded interesting differences between solutions obtained with ray and normal mode methods. These particular problems were defined with high boundary scattering loss. A bottom reverberation case at 3.5 kHz with a down-refracting sound speed profile (Problem VI) will be considered as a case in point. The ray solutions show a "direct path" contribution unaffected by the bottom scattering loss as long as a direct path can reach the bottom, while the mode solutions obtained to date show a lower reverberation level during this period due to modal attenuation. These differences occur in both incoherent and coherent reverberation solutions for both rays and modes. Arguments will be presented that indicate the correctness of the ray solutions for this case. Suggestions will also be made on how the mode approach can be used to obtain solutions in agreement with the ray method. [Work supported by ONR and the Navy Program Executive Office C4I (PMW 120).]

11:00

2aUW3. Reverberation versus time or reverberation versus range? A definitive relationship. Chris Harrison (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

Strictly reverberation consists of all the back-scattered contributions from many locations that arrive at one absolute time regardless of their respective travel times. However, for calculation purposes, particularly at long range, it is often more convenient to assume the scatterers to be close together at a single location with corresponding propagation loss for the outward and return paths. It would be useful to know whether the latter approximation is a safe one or not. An analytical method of calculating reverberation either at a fixed time or at a fixed range will be presented. The ratio of these quantities is found to depend on the (vertical) angular spread of the contributing paths. In isovelocity water at long ranges, mode-stripping reduces the angular spread to the point where fixed-range reverberation is indistinguishable from fixed-time. The greatest effect is at shorter ranges where there is no significant mode-stripping, and the ratio rises to roughly 1 plus 0.855 times the square of the critical angle (in radians). Thus the effect is typically less than 1 dB at any range.

11:15

2aUW4. Modeling acoustic propagation in shallow water using finite elements. Marcia J. Isakson and Preston S. Wilson (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713, misakson@arlut.utexas.edu)

Finite element models approach an exact solution of the Helmholtz equation as discretization density increases. Therefore, these solutions are an excellent benchmark model for reverberation studies. This study will present finite element reverberation solutions for two-dimensional shallow water waveguides for times up to 4 s. Five waveguides are considered: rough bottom only, rough surface only, rough surface and rough bottom, rough interfaces with summer (downward refracting) profile, and rough interfaces with winter (upward) refracting profile. [Work sponsored by Office of Naval Research, Ocean Acoustics.]

11:30

2aUW5. High-frequency broadband coherent reverberation predictions for the reverberation modeling workshop. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

The reverberation modeling workshop included cases with strong attenuation to the forward propagation caused by multiple forward scattering. The effects of this attenuation were consolidated into effective reflection coefficients using the small slope approximation and provided to workshop participants. In this talk the coherent modeling of reverberation for these strongly attenuated cases using complex modes as computed by KRAKENC is described, and the results of these coherent modal predictions are compared to results obtained by other participants using ray theory. The importance of including both the complex mode shapes and eigenvalues, as well as the coherent interactions between modes, to obtain accurate predictions is discussed.

TUESDAY AFTERNOON, 27 OCTOBER 2009 RIO GRANDE WEST, 2:00 TO 4:00 P.M.

Session 2pAB

Animal Bioacoustics: Emotion-Related Mechanisms of Mammalian Vocalizations

Michael J. Owren, Chair
Georgia State Univ., Dept. of Psychology, P.O. Box 5010, Atlanta, GA 30302-5010

Chair's Introduction—2:00

Invited Papers

2:05

2pAB1. Emotion-related acoustic communication in bats. Sabine Schmidt (Inst. of Zoology, Univ. of Veterinary Medicine Hanover, Buenteweg 17, 30559 Hanover, Germany, sabine.schmidt@tiho-hannover.de)

Some features of emotional prosody in human speech may be rooted in mechanisms common to mammals. The role of vocal communication in social interactions was studied in bats, a highly vocal group evolutionarily remote from primates. The present paper focuses on communication during agonistic encounters in the Indian False Vampire bat. Three call types with distinct time-frequency contours occurred; aggression calls, whistles, and response calls. In a first experiment, agonistic approach situations were analyzed to assess the extent to which these call types reflected the specific part of the caller in the interaction and the intensity of the agonistic display. A frame-by-frame video analysis followed by a sound analysis revealed that call type indicated the part of the respective caller while interaction intensity was encoded in similar parameter changes across call types. The systematic change in vocal parameters with

affect intensity corresponded to prosodic changes in human speech. A playback experiment based on a habituation-dishabituation paradigm investigated how the bats categorized the vocalizations emitted during third-party agonistic interactions. The bats were able to discriminate between call types; however, they did not necessarily form categories corresponding to call type in reciprocal experiments, an evidence for a context-dependent evaluation of social calls.

2:25

2pAB2. Vocal production, affect expression, and communicative function in African elephant rumble vocalizations. Joseph Soltis (Education and Sci. Disney's Animal Kingdom, Box 10000, Bay Lake, FL 32830, joseph.soltis@disney.com)

African elephant rumbles are harmonically rich vocalizations with clear formant structure. The low-frequency, high amplitude nature of rumbles is consistent with the observed large vocal production anatomy. Examination of formant locations and inter-formant dispersion implicates the trunk in addition to the oral cavity in rumble production. Adult female African elephants express affect intensity in ways similar to other mammals. When interacting with social superiors, subordinate rumbles exhibited increased and more variable fundamental frequencies, and increased amplitudes and durations, compared to periods of social calm. Filter features did not change across behavioral contexts. Acoustic responses originally related solely to affect may come to serve communicative functions. During dyadic approaches by adult females, rumbles resembling the "fear response" of subordinates (i.e., those with increased durations and more variable fundamental frequencies) increased the likelihood of friendly social interaction, and these signals were produced by both subordinate and dominant animals. Two top-ranking females were co-dominant and competed for alpha status. When interacting with each other, they produced rumbles with decreased fundamental frequencies, increased amplitudes and durations, and a decrease in inter-formant dispersion. This acoustic response is not generally consistent with the expression of affect but may signal large body size to social competitors.

2:45

2pAB3. Affect cues in communication calls of tree shrews. Elke Zimmermann (Inst. of Zoology, Univ. of Vet. Medicine, Buenteweg 17, 30559 Hannover, Germany, elke.zimmermann@tiho-hannover.de)

Comparative studies on affective prosody in human speech revealed remarkable cross-cultural similarities suggesting that affective prosody may have originated from a prehuman basis. To explore this hypothesis, the acoustic variation in communication calls and its perception in tree shrews were examined. Tree shrews are small diurnal mammals, genetically closely related to primates, living in dispersed pairs in the dense tropical forests of south-east Asia. Calls were induced experimentally in a social encounter and a disturbance paradigm, respectively, and related to two behaviorally defined arousal states within specific behavioral contexts. Context and arousal state of the caller reliably predicted spectral and temporal variations in call structure. Whereas context is closely associated with the frequency-time contour of calls (call type), arousal is expressed in shifts of fundamental frequency and the rate of call production. In a habituation-dishabituation paradigm, testing the effect of arousal-related variation within the same call type, tree shrews were able to discriminate acoustically between two arousal states. All in all, these findings document the relevance of affect cues in the vocal communication system of a non-primate mammal, the tree shrew, and support that mechanisms involved in the acoustical expression and perception of emotions are deeply rooted in mammals. [Work supported by DFG FOR 499.]

3:05

2pAB4. Infant mouse isolation ultrasounds: Production, perception and neural plasticity. Robert C Liu (Dept. of Biology, Emory Univ., 1510 Clifton Rd. NE, Atlanta, GA 30322, robert.liu@emory.edu)

Mouse pups emit ultrasonic whistles when they are isolated from their nest, likely reflecting an increased state of arousal that may be modulated by emotional and motivational processes [G. Ehret, Behav. Genet. **35**, 19–29 (2005)]. Mouse mothers recognize the behavioral relevance of these calls and search out these pups to retrieve them to the nest. On the other hand, pup-naive virgin females do not prefer to approach these calls over a neutral sound, suggesting they do not recognize their significance. Whether this difference in recognition correlates with any changes in how the auditory system processes these vocalizations is being investigated. Evidence from electrophysiological recordings in the auditory cortex of both anesthetized and awake female mice demonstrate that the cortical activity evoked by natural ultrasonic calls is changed in a way that could functionally improve the ability to detect and discriminate them. These results indicate that irrespective of a possible difference between virgins and mothers in the neural circuitry for deciding to respond to pup calls, the sensory system itself is plastic in its representation of these vocalizations.

Contributed Paper

3:25

2pAB5. About phonetic and perceptual similarities in laughing, smiling, and crying speech. Donna Erickson (1-11-1 Kamiasao, Asao-ku, Kawasaki City, Kanagawa 215-8558, Japan), Caroline Menezes (Univ. of Toledo, Toledo, OH), and Ken-ichi Sakakibara (Univ. of Hokkaido, Sapporo, Japan)

Acoustic, articulatory (EMA), and perceptual characteristics of recordings of spontaneous happy laughing/smiling, sad crying, and neutral speech were examined. Listeners were asked to (1) rate the emotional intensity of the speech and (2) categorize it as happy, sad, not-emotional, unknown, or other. Results show distinctive acoustic, articulatory, and perceptual characteristics for listeners' ratings of "emotional" versus "not-emotional" speech. Emotional speech has significantly higher F0, higher F2, lower H2, raised /retracted upper lip, and lowered tongue. Between happy laughing/smiling speech and sad crying speech, phonetic and perceptual similarities were observed, with mismatches between the emotion intended by the speaker and that perceived by listeners. The results of this study highlight how underlying emotions interact with temporal emotion in a complex way, affecting the perception of emotion. Laughing and crying to bring about a type of catharsis or balance to the system are also discussed. [This work was supported in part by Japanese Ministry of Education, Science, Sport, and Culture, Grant-in-Aid for Scientific Research (C), (2007–2010):19520371 and SCOPE (071705001) of Ministry of Internal Affairs and Communications (MIC), Japan.]

Invited Paper

3:40

2pAB6. Emotion and vocalization in mammalian perspective: What do we know, what do we need to know? Michael J. Owren (Dept. of Psych., Georgia State Univ., P.O. Box 5010, Atlanta, GA 30302-5010, owren@gsu.edu)

Emotional processes can play a central role in mammalian vocal communication, including both in triggering vocalizations and in mediating listener behavioral responses. However, the processes involved have received relatively little attention within the field of animal bioacoustics, at least in comparison to topics such as vocal anatomy, sensory processing, and the possible importance of higher-order cognition. It is therefore useful to take stock of what we currently know about the role of emotion in mammalian vocalization across species. Common elements of the papers in this Special Session will be highlighted, with the goal of identifying commonalities and differences in underlying approaches and overall findings. This overview will focus on key areas and issues that should be addressed both in reaching a better understanding of the role of emotion in vocalization and in raising the profile of this critical component of organismal function in the larger arena of the science of vocal communication.

TUESDAY AFTERNOON, 27 OCTOBER 2009 RIO GRANDE EAST, 1:35 TO 3:00 P.M.

Session 2pAOa

Acoustical Oceanography: Acoustics and Ocean Acidity II

Timothy F. Duda, Cochair
Woods Hole Oceanographic Inst., 98 Water St., Woods Hole, MA 02543-1053

Peter F. Worcester, Cochair
Univ. of California, San Diego, Scripps Inst. of Oceanography, 9500 Gilman Dr., La Jolla, CA 92093

Invited Paper

1:35

2pAOa1. Ocean acidification and its impact on ocean noise level: An analysis using empirical and physical models of acoustic transmission loss. D. Benjamin Reeder and Ching-Sang Chiu (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943, dbreeder@nps.edu)

Over the next several hundred years, ocean acidification is expected to reduce average ocean *p*H. It has been reported that, given a 0.3 reduction in *p*H from 8.1 to 7.8, a 40% reduction in the acoustic absorption coefficient at low frequencies (less than 1 kHz) could result, suggesting a significant increase in ocean noise. Such extrapolations are based on the corresponding percentage change in the absorption loss only and ignore other dominant acoustic energy loss mechanisms, such as seabed attenuation, as well as the effect of ocean waveguide physics and the depth dependence of ocean *p*H. Presented here is an analysis of the impact of the reduction in ocean *p*H on ocean noise level using empirical and physical models of acoustic transmission loss. Three ocean acoustic environments are analyzed to elucidate the expected change in ocean noise level as a function of frequency: shallow water, a surface duct, and the deep ocean. Results show a negligible change in ocean noise in the shallow water and surface duct environments for all frequencies. In the deep ocean, the maximum change in ocean noise level occurs at approximately 2 kHz, with no significant change below 800 Hz.

Contributed Papers

2:00

2pAOa2. Ocean noise level change in response to ocean acidification. Ilya A. Udovydchenkov and Timothy F. Duda (AOPE Dept., Woods Hole Oceanograph. Inst., 266 Woods Hole Rd., Woods Hole, MA, 02543)

It has been known for several decades that reduction in ocean *p*H (increase in acidity) would have an impact on ocean sound content and propagation loss. Net flux of carbon dioxide into the ocean has already reduced *p*H in the upper ocean, and as a result has reduced absorption in those same waters. This trend will likely continue. Based on forecasts of oceanic *p*H profiles and sound sources, changes in ocean sound content (ambient noise) can be computed. In general, sound energy loss can be partitioned into intrinsic dissipation (the chemical relaxation absorption process) and dissipation in the seafloor, with the boundary loss being analogous to sub-unity reflectivity at the boundaries of a room or concert hall. Here, expected deep ocean noise level changes in various frequency bands from 0.1 to 10 kHz are presented. A model is used that includes basic propagation physics (including the sound channel), depth variability of absorption, loss of sound energy into seafloor, and distributed acoustic sources. At low frequency, little sound is absorbed, so absorption effects are muted. At higher frequencies, sound does not propagate far, also minimizing absorption effects. In the center of the band, there can be sizable effects.

2:15

2pAOa3. A computational assessment of the sensitivity of ambient noise level to ocean acidification. John E Joseph and Ching-Sang Chiu (Oceanogr. Dept., Naval Postgrad. School, 833 Dyer Rd., Monterey, CA 93943, jejoseph@nps.edu)

Low-frequency (<2 kHz) sound propagating through the ocean is attenuated by the *p*H-dependent boric acid chemical relaxation process. This well-studied process shows that lower seawater *p*H results in lower sound attenuation at low frequencies. Thus, the uptake of anthropogenic CO_2 by seawater, leading to ocean acidification and lowering *p*H of seawater, has potential to change ambient noise levels. An important question is: By how much? Here, changes in ambient noise level due to theoretical changes in

seawater *p*H have been calculated at three different locations, the Pacific, Atlantic, and Arctic oceans, using a hypothetical spatial distribution of *p*H based on the literature and forecasted to the year 2100 [Cao *et al.*, (2007); McNeil and Matear, (2006)]. A range-dependent acoustic model has been applied using sound speed and winds based on climatology, and a historical shipping database to make ambient noise level predictions. Results of our model calculations are presented and discussed.

2:30

2pAOa4. Sound absorption and *p*H quantification in the presence of signal fluctuations. Timothy F. Duda (AOPE Dept. MS 11, Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The technique of using differential sound attenuation over a known path in two or more frequency bands to measure *p*H-dependent sound absorption effects, and by extension integrated ocean *p*H, was introduced in the 1980s by Jin and Worcester. The method is complicated by the fact that the mean levels of fluctuating signals with additive noise must be determined. The gain of the measurement increases with propagation distance because absorption (attenuation) accumulates with distance, but fluctuations tend to also increase with distance, and signal to noise ratio decreases, inhibiting the measurement. The fluctuation effects dominate. Utilizing intensity fluctuation data (from long-range experiments of recent decades) plus estimated signal and noise levels, the quantity of independent differential attenuation data points needed (N) to make measurement (*p*H accuracy of 0.05) is estimated and presented, as a function of center frequency, band separation, and propagation distance. One such estimate is $N=500$ for 450-Hz, 200-Hz band separation, 700 km. Reducing to 100-Hz separation quadruples N for a given distance and center frequency. Increasing distance slightly increases N and shifts the optimal frequency (least N required) downward, while reducing it has the opposite effect.

2:45

2pAOa5. Effect of ocean acidity on the mid-frequency ambient noise field in shallow water. Daniel Rouseff and Dajun Tang (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

Due to increased acidification, Hester *et al.* [Geophys. Res. Lett. **35**, L19601] estimate that by midcentury the sound absorption of seawater will decrease by almost 40%. In the present study, the expected relative increase in the mid-frequency ambient noise field is estimated. Emphasis is placed on shallow water scenarios with typical summertime sound speed profiles and frequencies from 1 to 5 kHz. A model is developed that assumes shallow water internal waves will redistribute energy from high-order acoustic modes excited by noise sources near the sea surface into ducted, low-order modes that propagate with low loss. Calculations are made as a function of *p*H level, the strength of the internal waves, and the bottom composition. The modeled noise field is beamformed and the effect on the so-called ambient noise notch is estimated.

TUESDAY AFTERNOON, 27 OCTOBER 2009 RIO GRANDE EAST, 3:45 TO 4:45 P.M.

Session 2pAOb

Acoustical Oceanography: Munk Award Lecture

Martin F. Siderius, Chair
Portland State Univ., Electrical and Computer Engineering Dept., 1900 SW Fourth Ave., Ste. 160-11, Portland, OR 97207

Chair's Introduction—3:45

Invited Paper

3:55

2pAOb1. Acoustical oceanography and shallow water acoustics. James F. Lynch (Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 02543, jlynch@whoi.edu)

In acoustical oceanography, we use sound to help infer the physical characteristics of the ocean and seabed. There is a rich store of oceanographic, geological, and biological phenomena to be explored in shallow water, and we look at some of these phenomena first, with an eye toward what the current interests and challenges are. In seabed acoustics, the continental slopes and the canyons that cross-cut them are new and challenging areas to address. In physical oceanography, the full four-dimensional description of the internal wave field on the continental shelf is still a challenge. Also, the physical oceanography of the shelfbreak, continental slope, and canyons is another exciting new area. Biologically, the fact that marine life concentrates near the shelfbreak and in canyons is known, but is not well quantified. In looking at these topics, we will first address the forward propagation problem, and then decide how and if the acoustic inverse problem should be addressed. We will also discuss what specific ocean, biological, and seabed properties most need to be measured, and also how to best measure them—whether acoustically or otherwise.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration: Imaging, Therapy, and Propagation

Mark S. Wochner, Chair
Univ. of Texas at Austin, Applied Research Lab., 10000 Burnet Rd., Austin, TX 78758

Contributed Papers

1:00

2pBB1. Kidney stone localization *in vitro* using multiple-signal-classification. Jonathan M. Kracht, Paul E. Barbone, and Robin O. Cleveland (Boston Univ., 110 Cummington St., Boston, MA 02215)

During shock wave lithotripsy, respiration and patient movement result in motion of a kidney stone. It has been estimated that 50% of the shock waves miss the stone. The misshots result in damage to the tissue with no therapeutic benefit. Here we employed a ring array of seven piezo-electric elements with center frequency of ~600 kHz to track stones *in vitro*. Each element was used to transmit in turn, and the individual waveforms received on all elements were recorded. The Fourier transform of these data gives the multistatic response matrix, whose largest eigenvalues represent the dominant scatterers in the medium. The associated eigenvectors were employed by the multiple-signal-classification (MUSIC) method to generate a scattering indicator image from which scatterer locations were determined. We tested MUSIC's applicability to track artificial kidney stones in water or with a tissue phantom. Single targets were tracked with an accuracy of 4 mm. A single stone resulted in three eigenvalues and therefore the seven element array could track two targets. Two targets were resolved for separations of 10 mm. The results indicate that MUSIC can be employed to track kidney stones although more elements would be required to track multiple fragments. [Work supported in part by NIH DK-43881].

1:15

2pBB2. Radiation pressure from ultrasound to help kidney stones pass. Wei Lu, Bryan W. Cunitz, Peter J. Kaczkowski (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th Seattle, WA 98105, weilu@u.washington.edu), Anup Shah (Univ. of Washington School of Medicine, Seattle, WA 98195), Oleg A. Sapozhnikov (Phys. Faculty, Moscow State Univ., Moscow 119991, Russian Federation), and Michael R. Bailey (Univ. of Washington, Seattle, WA 98105)

Residual kidney stone fragments often remain months after treatment. These fragments may nucleate new stones and contribute to a 50% recurrence within 5 years. Here, a research focused ultrasound device was used to generate fragment motion with the goal of facilitating passage. Natural and artificial stones 1–8 mm in length were surgically placed in the urine space in pig kidneys. The ultrasound source was a 2.75-MHz, eight-element annular array with a 6-cm radius of curvature. At adjustable focal depths of 5–8 cm, the focal pressure beam width in water was about 2 mm, and peak pressure was about 4 MPa. Targeting was done by ultrasound using B-mode and twinkling artifact that stones produce in Doppler mode. The commercial imaging probe was placed within and oriented down the axis of the therapy probe. Ultrasound and fluoroscopy showed the stones moving in real-time under the influence of the focused ultrasound. Stones moved on the order of 1 cm/s away from the source and several stones moved several centimeters down the ureter. It appeared that stones were affected only when directly in the focal beam, perhaps indicating that radiation pressure not streaming caused the motion. [Work supported by NIH DK43881 and NSBRI SMST001601.]

1:30

2pBB3. Dependency of attenuation slope of tumor tissue on animal viability. Alexander Haak, Zachary T. Hafes (Dept. of Elec. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801, ahaak@uiuc.edu), Timothy J. Hall (Univ. of Wisconsin-Madison, Madison, WI 53705), and William D. O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801, ahaak@uiuc.edu)

The attenuation slope (AS) is an acoustic parameter that can be utilized to characterize tissue. Previously, an algorithm was developed and validated on tissue mimicking phantoms that utilized the backscattered ultrasound signal to estimate the AS from single element pulse/echo data. Most AS data reported in literature have been acquired from dead tissue samples. There may be differences in the AS between live and dead tissues due to blood flow. Therefore, the dependency of the AS on animal viability is investigated. Fibroadenomas that developed spontaneously in Sprague Dawley rats were scanned with two single element transducers (3.5 and 7.5 MHz). The animal was then euthanized and the tumor was scanned again using the same parameters as in the previous scan. The data were postprocessed and the AS were estimated. The estimated AS of fibroadenomas for live animals ranged from 0.34–1.4 and 0.53–1.48 dB/cm MHz for 3.5 and 7.5 MHz, respectively. Similar AS results (0.3–1.1 and 0.5–1.4 dB/cm MHz for 3.5 and 7.5 MHz, respectively) were obtained for the euthanized animals. An analysis of variance was performed on the AS data and no significant difference in the AS estimated between live and dead animals was obtained. [Work supported by NIH Grant No. R01CA111289.]

1:45

2pBB4. Comparison of ultrasonic backscatter coefficient estimates from rat mammary tumors using single element and array transducers. Lauren A. Wirtzfeld, Goutam Ghoshal, Zachary T. Hafez, Rita J. Miller, Sandhya Sarwate (Bioacoustics Res. Lab., Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801), Timothy J. Hall (Univ. of Wisconsin, Madison, WI 53705), Michael L. Oelze, and William D. O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

The ultrasonic backscatter coefficient (BSC) is the fundamental quantitative estimate from measurements that can be parametrized to yield the effective scatterer diameter and acoustic concentration. The ability to accurately estimate the BSC using different imaging systems (i.e., a system independent estimate) is significant for clinical application of QUS imaging. In this study, BSCs were estimated from spontaneous mammary tumors in rats using both single-element transducers and linear arrays from a clinical imaging system. The BSC as a function of frequency was computed from the rf backscattered signals from Sprague Dawley rats that developed either fibroadenoma or carcinoma tumors. The tumors were scanned using three single-element transducers with a collective −10-dB bandwidth of 1.4–18 MHz and two linear arrays from the Ultrasonix RP system with a collective −10-dB bandwidth of 2–8 MHz. For the single-element transducers, a smooth Plexiglas plate was used to acquire a reference scan. For the linear arrays, a well characterized tissue mimicking phantom containing spherical glass beads was used as a reference. Based on the respective reference scans, the BSCs were extracted from the data. The results between a laboratory system and a clinical imaging system showed good agreement. [NIH Grant R01CA111289.]

2:00

2pBB5. A low-frequency array system for transcranial foreign body detection. Caleb H. Farny, Sai Chun Tang, and Greg T. Clement (Dept. of Radiology, Brigham and Women's Hospital, Harvard Med. School, 221 Longwood Ave., Boston, MA 02115, cfarny@bwh.harvard.edu)

There has been recent interest in developing portable devices for advanced battlefield regions to detect foreign bodies that have penetrated the skull. Here we present on a compact ultrasound array system for transcranial detection of foreign bodies embedded in a tissue-mimicking phantom. The system is designed to be powered from a laptop via USB and present information regarding the presence of foreign targets within the skull. Design parameters investigated the influence of transducer center frequency and diameter, with particular importance placed on minimizing the number of transducers employed in the array for weight reduction. Frequencies ranging from 320 to 800 kHz were examined as well as the tradeoff between longitudinal and shear mode propagations through the skull bone layer. *Ex vivo* human calvaria were used to mimic application conditions. Measurement consisted of successive excitation of each transducer in the array by a three-cycle bipolar pulse from a single-channel multiplex pulse-receive system, and the scattered signal was successively received by each transducer in the array. The maximum volume of interrogation and resolution of detection relative to the skull bone were determined from placement of stainless steel and ceramic targets embedded in the phantom positioned inside the skull. [Work supported by USAMRMC:NPI-0704-00710.]

2:15

2pBB6. In vivo imaging of developing mouse embryos using a high-frequency annular-array and chirp-coded excitation. Jonathan Mamou, Jeffrey A. Ketterling (F. L. Lizzi Ctr. for Biomedical Eng., Riverside Res. Inst., 156 William St., New York, NY 10038, mamou@rrinyc.org), Orlando Aristizábal, and Daniel H. Turnbull (New York Univ. School of Medicine, New York, NY 10016)

High-frequency ultrasound (HFU) offers great potential for *in vivo* imaging of developing mouse embryos. Nevertheless, mouse-embryo imaging remains challenging because of the random orientation of the embryos and of the limited acoustic penetration depth of HFU. In this study, a 34-MHz, five-element annular array was excited using conventional mono-cycle and chirp-coded excitations. Synthetic focusing and pulse compression (for the chirp excitation) were used to form images with increased depth of field and penetration depth (for the chirp excitation). Three-dimensional (3-D) datasets were acquired from 11-, 12-, and 13-day-old mouse-embryo heads using both excitation schemes and a respiratory-gating algorithm that limited motion artifacts. 3-D reconstructions of the brain ventricles were formed and compared to MRI-based reconstructions. The chirp-based reconstructions were more morphologically accurate than the mono-cycle-based reconstructions. Furthermore, only chirp datasets provided sufficient penetration depth to correctly segment the brain ventricles lying deep inside the mother. Differences between the calculated brain-ventricle volumes obtained using chirp and mono-cycle excitations were as high as 80%. The results of this study suggest that high-frequency chirp-annular-array imaging could become a valuable tool for translational studies and non-invasive efficient phenotyping of the early mouse central nervous system. [Work supported by NIH Grants EB006509 and EB008606.]

2:30

2pBB7. Shannon entropy can detect myocardial anisotropy without specular echo gating *in vitro*. Ya-Jian Cheng, Michael S. Hughes, Jon N. Marsh, Kirk D. Wallace, and Samuel A. Wickline (Dept. Biomedical Eng., Washington Univ. in St. Louis, Forest Park Ave., St. Louis, MO 63124, primase@gmail.com)

Ultrasonic images are often corrupted by specular echoes that obscure features near interfacial boundaries. This problem can be especially severe for *in vitro* examination of excised tissues (e.g., myocardium) which must often be sliced into thin sections to permit examination by other techniques in addition to ultrasound (e.g., optical, MRI, and histology), in order to characterize infarct region size and myofiber orientation. In many of these studies, myocardial anisotropy is obscured without removal of specular echoes by gating. Here, it is demonstrated that entropy imaging enables visualization of myocardial anisotropy without gating. Short axis slice of mouse heart ($n=4$) was C-scanned on a 55×55 grid, with 0.1 mm stepsize. At each point one 8-bit, 2048 point, time-averaged A-line was acquired. These were analyzed using three signal receivers, log variance, log energy (sum of squared), and Shannon entropy, to produce one pixel value at each location in the C-scan. Diffusion-tensor-weighted magnetic resonance at 12 T was used to validate the ultrasonic result. Among the three signal receivers, only entropy clearly resolves myocardial anisotropy without specular echo removal. [This work supported by NIH CA119342, and NSC-095-SAF-I-564-051-TMS.]

2:45

2pBB8. Real-time Renyi entropy processing for molecular imaging using targeted nanoparticles. Kirk Wallace, John McCarthy, Victor Wickerhauser, Jon Marsh, Gregory Lanza, Samuel Wickline, and Michael Hughes (Washington Univ. in Saint Louis, St. Louis, MO 63108)

Previously, improvements in *in vivo* molecular imaging sensitivity were obtained using Renyi entropy, $I_f(r)$ with values of r near 2, specifically $r=1.99$. This result raised the possibility of further improvements in sensitivity even closer to the limit $r\rightarrow2$ [at $r=2$, $I_f(r)$ is undefined]. However, such an investigation was not feasible due to excessive computational time required to calculate $I_f(r)$ near this limit. In this study, an asymptotic expression for the limiting behavior of $I_f(r)$ as $r\rightarrow2$ is derived and used to present results analogous to those obtained with $I_f(1.99)$. Moreover, the limiting form, $I_{f,\infty}$ is computable directly from the experimentally measured waveform, $f(t)$ by an algorithm suitable for real-time implementation. To test our approach, five mice were injected with $\alpha_v\beta_3$-targeted nanoparticles, and ultrasound images obtained at 0-, 15-, 30-, and 45-min post-injection. Two control groups ($N=5$, injected with untargeted-nanoparticles, or no injection) were also imaged. Renyi images were able to differentiate the groups ($p<0.05$) at 15 min post-injection. This outcome agrees with previous studies using targeted-nanoparticles and demonstrates the ability of entropy-based signal receivers when used in conjunction with targeted-nanoparticles to elucidate the presence of $\alpha_v\beta_3$-integrins in primordial neovasculature, particularly in acoustically unfavorable environments. [Work supported by NIH EB002168.]

3:00—3:15 Break

3:15

2pBB9. Assessment of shear elasticity and viscosity of tissue using acoustic radiation force applied to a spherical acoustic inhomogeneity. Andrei B. Karpiouk, Salavat R. Aglyamov, and Stanislav Y. Emelianov (Dept. of Biomedical Eng., Univ. of Texas at Austin, r. 3.314, 107 W Dean Keeton St., Austin, TX 78712, andrei.karpiouk@engr.utexas.edu)

An accurate assessment of mechanical properties of tissue is desired in many applications ranging from biomedical research to medical diagnostics and surgery. Currently, several methods to assess shear elasticity of tissue are available while the techniques to measure tissue viscosity are not yet fully developed. However, shear viscosity is an indicative functional parameter of soft tissue. In addition, the local changes of shear viscosity can affect the accuracy of shear elasticity assessment. Therefore, in this study, an acoustic radiation force approach to assess both shear elasticity and viscosity of tissue is developed. Such approach is based on monitoring of the spatio-temporal displacement of an ultrasound inhomogeneity in response to applied impulsive radiation force. In experiments, a 2-mm-diameter solid sphere was embedded into the gel phantoms with varying shear elasticity and viscosity. The sphere was then perturbed using acoustic radiation force produced by a single-element focused ultrasound transducer operating at 3.5 MHz. The displacement of the sphere was monitored using another single-element focused ultrasound transducer with center frequency of 25 MHz. The results indicate that the temporal characteristics of the displacement can be used to assess both shear elasticity and viscosity of tissues simultaneously and independently. [Work supported by NIH.]

3:30

2pBB10. Quantitative estimation of complex shear modulus of liver and hydrogels through shear-wave imaging. M. Orescanin (Dept. of Elec. and Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews Ave., Urbana, IL 61801, moresca2@uiuc.edu), M. A. Qayyum, K. S. Toohey, and M. F. Insana (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

A Doppler-based shear-wave imaging technique is described for estimating the complex shear modulus ($\mu+i\omega\eta$) at shear-wave frequencies between 50 and 450 Hz. The developed technique involves a mechanical actuator that harmonically drives a stainless steel biopsy needle placed in the medium. Narrowband cylindrical shear waves are imaged using pulsed Doppler techniques. Shear moduli were computed from Doppler detected velocity. A phase gradient technique is applied to measure shear-wave speed. Spatially averaged speeds were numerically fit to a mathematical model relating dispersion and complex modulus. The proposed method was used to estimate complex shear modulus of the homogeneous three dimensional collagen hydrogels and of fresh and thermally-damaged porcine liver. The elastic shear modulus of 4% collagen hydrogel was measured, $\mu=640\pm14$ Pa, using a commercial rheometer as a standard. The phase gradient approach yielded frequency-independent moduli $\mu=570\pm67$ Pa and $\eta=0.16\pm0.09$ Pa s. For fresh liver shear-wave imaging between 50–300 Hz yielded frequency-independent moduli $\mu=1757\pm733$ Pa and $\eta=1.72\pm0.48$ Pa s and $\mu=3851\pm3233$ Pa and $\eta=8.9\pm2.9$ Pa s for thermally-damaged liver. Good agreement between the rheometer measurements and the shear-wave imaging approach shows that we can quantitatively estimate viscoelastic properties of the hydrogels and liver.

3:45

2pBB11. Generation of long pulses of focused ultrasound by time reversal system. Laurent Fillinger (Artann Labs., 1459 Lower Ferry Rd., Trenton, NJ 08618), Yegor Sinelnikov (126 Liberty Ave., Port Jefferson, NY 11777), Alexander Sutin, and Armen Sarvazyan (Artann Labs., Trenton, NJ 08618)

Ultrasound focusing based on time reversal acoustics (TRA) principles typically produces short ultrasound pulses having relatively low time-average intensity, which highly limits the therapeutic applications of these systems. The aim of this study was to explore theoretically and experimentally the possibility of generating long ultrasound signals by the TRA focusing. The TRA focusing of long signals requires application of broadband signals and depends on the signal bandwidth and the number of transducers. The radiation of the sequence of the TRA focused long signals makes it possible to focus ultrasound in a continuous mode. The theoretical model comprised several point sources located inside a liquid-filled reverberator. The model allowed calculation of the spatial structure of ultrasound field for the various focused signals. The dependence of the focused field spatial structure on the bandwidth of the signal as well as onto the number of ultrasonic transducers was investigated. Experiments were conducted using a water filled plastic bottle reverberator with five attached transducers. The effective TRA focusing of long signals was demonstrated for linear sweep and band-limited noise in the 200–800-kHz range.

4:00

2pBB12. Fast MATLAB-based ultrasound simulation software: Time harmonic calculations. Robert J. McGough, Donald VanderLaan, and Joshua Wong (Dept. of Elec. and Comput. Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824)

A library of fast C++ routines is created for MATLAB simulations of time harmonic pressures generated by ultrasound phased arrays. These programs are included in the recently released FOCUS software package, where FOCUS is an acronym for "fast object-oriented C++ ultrasound simulator." FOCUS is intended for large scale phased array simulations, characterizations of beamforming strategies, and analysis of new phased array designs. This software supports nearfield simulations of individual flat circular and rectangular pistons and focused spherical shells. Routines for pressure calculations with flat and curved phased array geometries consisting of circular and rectangular elements are also included. Within the FOCUS program, time harmonic simulations are performed with a combination of the fast nearfield method and the angular spectrum approach. Comparisons of errors and computation times with competing programs are evaluated for single element and phased array calculations, and results show that, for the same peak error value, FOCUS achieves an order of magnitude or more reduction in the computation time in large phased array simulations.

4:15

2pBB13. Fast MATLAB-based ultrasound simulation software: Transient calculations. Robert J. McGough, Donald VanderLaan, and Joshua Wong (Dept. of Elec. and Comput. Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824)

The FOCUS software package quickly calculates transient pressure fields generated by ultrasound phased arrays. This software is useful for visualizing pressure wave propagation in the nearfield region, and the results are directly applicable to beamforming evaluations and phased array design for diagnostic and therapeutic applications. FOCUS includes support for transient pressure calculations with single transducers, including flat circular and rectangular pistons as well as spherically focused shells, and arrays of flat circular and rectangular transducers. Routines to calculate time delays for focusing the arrays are also included with the software. Transient calculations are performed with a combination of the fast nearfield method and time space decomposition, which achieves rapid convergence to reduce the numerical error while significantly reducing the computation time. Within the FOCUS software, time space decomposition is implemented with a simplified approach that decreases the overhead and reduces the computational complexity. The advantages of FOCUS over competing approaches include short computation times and small numerical errors. Examples of the supported element and array geometries will be shown, and sample codes will be provided. Transient pressure fields computed with FOCUS and with FIELD II are compared, and results show that, for the same peak error value, FOCUS is several times faster for calculations in the nearfield region.

Session 2pMUa

Musical Acoustics: Acoustics of Free-Reed Instruments II: Accordion Acoustics

James P. Cottingham, Chair
Coe College, Physics Dept., Cedar Rapids, IA 52402

Invited Paper

1:00

2pMUa1. The adoption of the accordion and other bellow-blown free-reed instruments in world cultures. Paul A. Wheeler (Utah State Univ., 4120 Old Main Hill, Logan, UT 84322-4120)

Free-reed mouth organs, such as the Laotian khaen or the Chinese sheng, have been known for centuries in Asia. Several free-reed instruments (such as the accordion, concertina, and harmonium) using bellows rather than the mouth as a wind supply were developed in Europe during the 19th century. As European influence spread throughout the world, so did the adoption of the accordion-like instruments that penetrate local musical cultures around the world. This paper presents an overview of the usage of bellows-blown free-reed instruments in musical cultures of the world. It includes the zydeco accordion of New Orleans, the bandoneon (concertina) of the Argentine tango, and the harmonium of Hindustani music in India.

1:30

2pMUa2. Characteristics and construction of the accordion. John J. Stankus (11308 Pickfair Dr., Austin, Texas, 78750-2528, squeeze@accordion.com)

The accordion family has a broad spectrum of members including piano accordions, chromatic button accordions, diatonic button accordions, bayans, buttonbox accordions, concertinas, and bandoneons. The common sound generator in all of these instruments is the free reed. This presentation will focus on the characteristics and construction of the accordion with emphasis on the tunings and the reeds. The configuration of the keys and buttons will be explained using the piano accordion. The application of the concert tuning and musette tuning will be contrasted. The versatility of the accordion will be shown as the combination of the four sets of reeds emulates the various instruments of the orchestra. The effect of tone chambers will be discussed. Accordion construction and its materials will be explained with photos and hands on parts. The materials and construction processes will be described with focus on the reeds and reed blocks. The 19th century construction of the accordion has not really changed. Some ideas will be presented for building a modern accordion utilizing the latest materials and processes. Examples of this are the use of carbon fiber molded case and reed manufacture using electrostatic discharge machining.

Contributed Papers

2:00

2pMUa3. Vibrational modes of accordion reeds. Sarah L. Behrens (Coe College, 1220 First Ave. NE, Cedar Rapids, IA 52402, slbehrens@coe.edu), Whitney L. Coyle (Murray State Univ., Murray, KY 42071), Nicholas P. Goodweiler (Univ. of Iowa, Iowa City, IA 52242), and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

Some new measurements made of the oscillation of air-driven accordion reeds show that higher transverse modes through the fourth mode are present as well as the first torsional mode. The second and third transverse modes are observable even at low amplitudes of oscillation. All of these have been previously observed in reed organ reeds [Paquette *et al.*, J. Acoust. Soc. Am. **114**, 2348 (2003)]. Additionally, for the first time a lateral mode of vibration (transverse vibration perpendicular to the usual first transverse mode) has been observed. For airflow in a given direction, only one of the two reeds mounted in each wind chamber is the primary source of sound production, but the vibration of the secondary reed has also been studied. The amplitudes of higher-frequency modes relative to the fundamental are observed to be higher in the secondary reed than in the primary reed. Finite element calculations of the reed modes have been made, and the calculated mode frequencies and node locations were used to verify mode identification. [Work partially supported by National Science Foundation REU Grant No. PHY-0354058.]

2:15

2pMUa4. Influence of accordion reed chamber geometry on reed vibration and airflow. Whitney L. Coyle (Murray State Univ., Murray, KY 42071, wlcoyle@hotmail.com), Sarah L. Behrens, and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

An experimental study has been made on the motion of air-driven accordion reeds in which some measurements were made with the reed block removed from the instrument, but most were made with the reeds in the instrument with the bellows replaced by a clear acrylic wind chest driven at appropriate blowing pressure by a small organ blower. Measurements of reed displacement and velocity as a function of time were made using a laser vibrometer system, and corresponding sound pressure waveforms were obtained from a probe microphone near the reed opening. Airflow waveforms were calculated by integrating the pressure waveform and using a computed area function. Effects of changes in reed chamber geometry on the reed vibration spectrum and the airflow waveform have been investigated, along with effects on the spectrum of the radiated sound and sounding frequency. Results of changes in the position of the pallet valve, such as partial opening, have also been studied. Additional measurements made for each reed include the variation in sounding frequency and amplitude of vibration with blowing pressure. The results are compared with previous measurements and calculations from theoretical models. [Work partially supported by National Science Foundation REU Grant No. PHY-0354058.]

Invited Paper

2:30

2pMUa5. Accordion reeds, cavity resonance, and pitch bend. Thomas Tonon (11 Bolfmar Ave., Princeton Junction, NJ 08550)

In keyed free reed instruments such as accordions and concertinas, the reeds are mounted over cavities that have little effect on the vibration of the reed itself, because resonances between the reed and cavity are rarely encountered. In fact, in conventional instruments, resonances can interfere with the self-excitation mechanism of the reed and become problematic to the builder. On the other hand, one can exploit such resonances in order to produce pitch bend and other acoustic effects, by intentionally designing the cavity for near resonance and by providing a mechanism that permits the musician to engage resonance at will. First, conventionally designed cavity resonance is investigated over the frequency range of the instrument, illustrating the thoroughness and facility with which cavity resonance can be designed out of conventional instruments. Finally, cavity designs and linkage mechanisms to exploit desirable acoustic effects of cavity resonance are illustrated.

TUESDAY AFTERNOON, 27 OCTOBER 2009 LIVE OAK, 3:15 TO 4:15 P.M.

Session 2pMUb

Musical Acoustics: Pitch Bending Accordion Demonstration and Concert

James P. Cottingham, Chair
Physics Dept., Coe College, Cedar Rapids, IA 52402

Accordionist Ruben Coe of Laredo, Texas, will demonstrate pitch bending on a specially designed accordion.

TUESDAY AFTERNOON, 27 OCTOBER 2009 RIO GRANDE CENTER, 5:00 TO 6:00 P.M.

Session 2pMUc

Musical Acoustics: Concert Featuring Juan Tejéda and the Band Conjunto Aztlan

Paul A. Wheeler, Chair
Utah State Univ., Logan, UT 84341

Conjunto Aztlan, founded in Austin in 1977, performs in the Austin and San Antonio areas. The instrumentation features Tejano button accordion, two- and three-part harmonies, and original songs. The music includes traditional conjunto as well as a variety of other styles.

TUESDAY AFTERNOON, 27 OCTOBER 2009 REGENCY EAST 1, 1:30 TO 4:15 P.M.

Session 2pNS

Noise: Advancements in Noise Control, Hearing Conservation, and Speech

Pamela J. Harght, Chair
BAi LLC, 4006 Speedway, Austin, TX 78751

Contributed Papers

1:30

2pNS1. Comparison of noise levels between four hospital wings with different material treatments. Cassandra H. Wiese, Lily M. Wang, and Lauren M. Ronsse (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, cwiese@mail.unomaha.edu)

Noise levels in four hospital areas of the University of Nebraska Medical Center in Omaha, NE, have been measured and compared to study the relationship between materials present in each of the zones and the resulting noise levels. The four areas evaluated, including three hospital wings and a neonatal intensive care unit, have varying levels of acoustic floor and ceiling treatments in the hallways and nurses' stations. Sound pressure levels were logged every 10 s over a 24-h period in at least three different locations simultaneously within each area: at the nurses' station, in the hallway, and in a nearby patient room. The resulting data were analyzed in terms of the hourly A-weighted equivalent sound levels (Leq) as well as various exceedance levels (Ln). Results indicate that the material finishes in a zone are found to affect the ambient sound levels within the nurses' stations and hallways, while peak levels remain similar in most of the areas. Sound levels in the patient rooms, however, appear less correlated to materials in the hallway and nurses' station and more impacted by the peak levels coming from those spaces as well as patient equipment.

1:45

2pNS2. Mitigating existing mechanical noise in a new state-of-the-art Indoor Training Center for University of Washington Golf. Julie A. Wiebusch (The Greenbusch Group, Inc., 1900 West Nickerson, Ste. 201, Seattle, WA 98119)

For several years, excessive mechanical noise levels had rendered the "multi-purpose room" in the UW Hec Edmundson Pavilion virtually unusable by every athletic department. Adjacent to a mechanical room and separated only by a pair of poorly sealed doors, the room served as the main pathway for all ductwork from the mechanical room to other areas in the building. In addition, a large duct branched off of the main duct and ran vertically along one wall to exhaust the locker rooms below. The vertical portion of this duct had been partially enclosed in a gypsum board chase, which was making contact with the ductwork. The vibrational energy was re-radiated as airborne sound, emitting a low-frequency rumble throughout the space. Background noise levels were measured in excess of 70 dBA and NC 67. The space was offered to the Men's and Women's Golf Teams for their new Indoor Training Facility. This case study presents the mitigation techniques used to reduce noise levels to allow the room to be transformed into a cutting edge training facility.

2:00

2pNS3. Controlling ball mill noise with acoustic blankets: A case study. David R Zuchelli (Mine Safety and Health Administration, P.O. Box 18233, Pittsburgh, PA 15236, zuchelli.david@dol.gov)

The purpose of wrapping ball mills with thermal/acoustic blankets is twofold: (1) to maintain the desired temperature inside the mill and (2) to reduce the sound emanating from the mill. This field case study focuses solely on the latter, evaluating and documenting the sound level reduction attributed to the installation of the blankets and its potential to reduce the noise dose experienced by the process attendant. One of the three ball mills had already been treated (baseline) before the evaluation started. However, treating the remaining two mills still proved to be very effective, further reducing ball mill noise in the near field by as much as 10.0 dBA and producing far field noise reductions of 3.0 dBA or greater, extending as far as 35 ft from the mills and to the floor above. The model developed from the survey data shows that a process attendant spending a substantial portion of a shift in the mill building will experience a significant reduction in overall nose dose.

2:15

2pNS4. Evaluation of hydraulic silencers. Nicholas E. Earnhart, Kenneth A. Marek, and Kenneth A. Cunefare (Woodruff School of Mech. Eng., Georgia Inst. of Technol., 771 Ferst Dr., Grad Box 261, Atlanta, GA 30332, nick.earnhart@gmail.com)

Hydraulic silencers are devices used to mitigate fluid-borne noise in fluid power systems. Fluid-borne noise may be produced by positive-displacement pumps and interacts with system components to generate vibration and air-borne noise. It is of interest to quantify the performance of silencers with regard to the input reflection coefficient and transmission loss. A test rig has been constructed that uses a six-sensor method to determine these quantities. Construction and methodology will be presented along with results for a commercially available silencer.

2:30

2pNS5. Numerical model for a hydraulic in-line silencer. Kenneth A. Marek, Nicholas E. Earnhart, and Kenneth A. Cunefare (Dept. of Mech. Engr., Georgia Inst. of Technol., 771 Ferst Dr., Grad Box 334, Atlanta, GA 30332, ken.marek@gatech.edu)

It is proposed that an in-line hydraulic silencer with a particular engineered lining can provide an effective alternative to other current silencing technologies. As a first step in validating this hypothesis, and to help with the design process, a numerical model is developed for a hydraulic silencer. The silencer consists of an annular dispersive medium with a rigid outer boundary, where fluid flow is directed through the annulus. Silencer characteristics, including input port impedance and acoustic transmission loss, are predicted by this model. Effects of silencer geometry and material properties, including nonlinearities, are discussed.

2:45—3:00 Break

3:00

2pNS6. Improvement of vowel formant discrimination: Effects of global and local enhancement. Ashley Woodall and Chang Liu (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, Austin, TX 78712)

Vowel formant discrimination measures the smallest change in vowel formant frequency that is detectable. Thresholds of vowel formant discrimination were examined for F2 frequency of three American English vowels on normal-hearing and hearing-impaired listeners. Vowel stimuli were manipulated with a global enhancement by increasing speech level from 70- to 90-dB SPL and a local enhancement by increasing F2 amplitudes of 3, 6, and 9 dB. For each listener, vowel formant discrimination was tested for six sets of stimuli (3 local F2 enhancement $\times$ 2 global enhancement). Results showed that both normal-hearing and hearing-impaired listeners significantly benefited from local enhancement and local plus global enhancement, but not from global enhancement only. In addition, the improvement in vowel formant discrimination for hearing-impaired listeners was substantially higher than for normal-hearing listeners. However, given the same formant structure (original or enhanced F2), normal-hearing listeners did not receive any benefit from increasing the global level from 70- to 90-dB SPL. On the other hand, hearing-impaired listeners improved their thresholds from 70- to 90-dB SPL, but only when enhanced F2 was presented. There results indicate that, to improve vowel formant discrimination, local enhancement of target formant amplitude is more important than global enhancement.

3:15

2pNS7. Testing the limits: Quantifying the degradation of automatic speech recognition in reverberant environments. Stephen Secules (Arup Acoust., 13 Fitzroy St., London W1T 4BQ, United Kingdom, stephen.secules@arup.com) and Jonas Braasch (Rennselaer Polytechnic Inst., Troy, NY 12180)

Surprisingly little is known about the specific character of the depreciation of automatic speech recognition in reverberation—its primary acoustics causes (e.g., room geometry and reverberation strength) or speech effects (blurring of syllables, plosives, and consonants). The focus of this study is to precisely quantify the depreciation of speech recognition accuracy for reverberant signals using a black box experiment to vary reverberation characteristics and observe speech recognition accuracy. The methodology tests two speech recognition platforms on a recognition task of similar sounding word lists. A range of reverberant settings was simulated by convolution with an impulse response. The recognizers had the least reverberant recognition accuracy for words which only differed by their ending consonants. The depreciation of recognition accuracy from early reflections alone was lower than the overall room effect; however, the overall depreciation with respect to the absorption coefficient was well predicted by the strength of the reverberant tail. The results were compared to the results of prior research.

3:30

2pNS8. Angle dependent effects for impulse noise reduction for hearing protectors. William J. Murphy, Amir Khan, and Edward L. Zechmann (Hearing Loss Prevention Team, Nat. Inst. Occup. Safety and Heatlh, 4676 Columbia Parkway, MS C-27, Cincinnati, OH 45226-1998, wjm4@cdc.gov)

The proposed U.S. Environmental Protection Agency regulation for labeling hearing protection devices (HPDs) includes an impulsive noise reduction rating. In 2009, the American National Standards Institute Subcommittee for noise approved a revised standard for measuring the impulsive insertion loss of HPDs, ANSI/ASA S12.42-2009. The exposure at the ear in response to a forward-propagating wave depends strongly on the orientation of the head with respect to the direction of propagation. Furthermore, the insertion loss varies with the peak sound pressure level. This paper reports

the results of tests performed using an acoustic shock tube to produce peak impulses of approximately 160-dB peak sound pressure level. Two manikins were evaluated: the GRAS KEMAR manikin equipped with 1/2- and 1/4-in. microphone in a GRAS 711 IEC coupler and the Institute de Saint Louis manikin equipped with a Bruel & Kjaer IEC 711 coupler equipped with a 1/4 in. microphone. The manikin heads were rotated through ±90 deg relative to the direction of the oncoming wavefront and impulsive peak insertion loss was measured according to S12.42-2009. [Portions of the research were supported by U.S. EPA Interagency Agreement No. 75921973-01-0.]

3:45

2pNS9. *In-situ* measures of user's preferred listening levels with a portable digital music device. Edward L. Goshorn, Kathryn J. White, and Brett E. Kemker (Speech and Hearing Sci. Dept., 118 College Dr. 5092, Univ. of Southern MS, Hattiesburg, MS 39406-0001)

The recent development and popular use of personally worn digital music devices (PDMDs) have led to concerns about the intensity levels they produce. There is concern that users will set a PDMD to levels that are known to be hazardous to human hearing. The concern stems from the private nature of a PDMD in that insert earphones are worn that restrict the output to the user's ears, thus prohibiting monitoring of such devices by parents, teachers, employers, etc. This project measured the *in-situ* dB SPL produced by PDMDs worn by 31 young adults (18–23 years) who had been wearing PDMDs for 3 years or less. Subjects set their PDMD to their preferred listening level (PLL) for a popular tune. Once set, *in-situ* SPL measures of the tune at a reference point and then for a white noise were obtained with a probe microphone located near the eardrum. Hearing thresholds were also obtained. Results showed a range of 52–107 dB with a mean PLL of 83 dB for music and a range of 53–99 dB with a mean PLL of 79 dB for white noise. None of the subjects had significant hearing loss.

4:00

2pNS10. Ultrawideband filter for audible sound. M. S. Kushwaha (Dept. of Phys., Univ. Sci. Tech. Lille1, Citye Scientifique, 59650 Villeneuve D'Ascq, France)

Extensive band structure computation has been performed for two-dimensional periodic arrays of rigid stainless steel cylinders in air, with Bloch vector being perpendicular to the cylinders. It is proposed that the fabrication of a multiperiodic system in tandem could create a huge hole in sound within the human audible range of frequencies.

TUESDAY AFTERNOON, 27 OCTOBER 2009 REGENCY EAST 3, 1:55 TO 4:05 P.M.

Session 2pPA

Physical Acoustics and Engineering Acoustics: A Man For All Seasons: Tribute to Robert T. Beyer

Timothy K. Stanton, Cochair
Woods Hole Oceanographic Inst., Dept. of Applied Ocean Physics and Engineering, Woods Hole, MA 02543

Murray S. Korman, Cochair
U. S. Naval Academy, Physics Dept., 572 C Holloway Rd., Annapolis, MD 21402

Chair's Introduction—1:55

Invited Papers

2:00

2pPA1. Robert T. Beyer: Thoughts and reflections from his family. Rick Beyer (34 Outlook Dr., Lexington, MA 02421, rick @plateofpeas.com)

Thoughts and reflections on what made Robert T. Beyer such an inspiring and unforgettable person, through family stories, 50 years of correspondence, and memories that friends and colleagues have shared with his children.

2:20

2pPA2. Contributions of Robert T. Beyer to acoustical science: A review. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Stephen V. Letcher (Dept. of Phys., Univ. of Rhode Island, Kingston, RI 02881), and Mark B. Moffett (731 Annaquatucket Rd., North Kingstown, RI 02852)

From his appointment as Instructor at Brown University in 1945 until well after his retirement, Beyer made significant contributions to the discovery and dissemination of knowledge about physical acoustics. Beginning with water and electrolytic solutions, Beyer and his students used ultrasonics to study relaxation processes in liquified gases, organic liquids, and single crystals, and in liquid metals. The 1951 review article with Markham and Lindsay [Markham, Beyer, and Lindsay, Rev. Mod. Phys. **23**, 353411 (1951)] became a starting point for all ultrasonicists. Nonlinear acoustic phenomena also received special attention by Beyer and his students, who measured absorption and distortion of finite-amplitude waves; interaction of sound with sound, noise, and turbulence, including the first empirical verification of the parametric acoustic array in both water and air; and nonlinearity parameter of fluids as diverse as water, organic liquids, and liquid metals. New apparatus and methods were developed to aid the measurements, as of radiation pressure for the determination of absorption, a radio-frequency pulse system to measure distortion, and a complex wattmeter to measure intensity in a reactive acoustic field.

2:40

2pPA3. Robert T. Beyer: A sustained record of distinguished service to the profession. Allan D. Pierce (Acoust. Society of America, Ste. 1NO1, 2 Huntington Quadrangle, Melville, NY 11747, adp@bu.edu)

A historical and anecdotal account is given of Robert T. Beyer's professional service. Highlights include his service to the Acoustical Society as Member of the Executive Council (1956–1959), Vice-President (1961–1962), Chair of the Physical Acoustics Technical Committee (1963–1964), President-Elect (1967–1968), President (1968–1969), and Treasurer (1974–1994). He was Chair of the Investments Committee (1990–1992) and served as an Associate Editor of the Journal for three distinct terms: 1952–1955 (References), 1965–1969 (Acoustical News from Abroad), and 1985–1991 (Book Reviews). He also served as interim Editor-in-Chief after the death of Robert B. Lindsay in 1985. Beyer was a voracious reader, an indefatigable translator, and a prolific writer. He wrote several textbooks, numerous book reviews, and frequent reports in JASA on the finances of the Society. His activities as a translator and editor of translated journals are legendary. Beyer's connections with the American Institute of Physics include Board of Governors (1969), Member of Executive Committee of Board (1974), Member of Translation Advisory Board (1955–1978) and its Chairman (1957–1977), and Board of Translations Editors (1978). The ASA awarded Beyer its Distinguished Service Citation in 1978, but he undoubtedly did enough more between 1978 and 2008 to justify being awarded several additional Service Citations.

Contributed Papers

3:00

2pPA4. Robert Beyer's laboratory in nonlinear acoustics: 1975–1982 and earlier perspectives. Timothy K. Stanton (Dept. of Appl. Ocean Phys. and Eng., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, tstanton@whoi.edu) and Murray S. Korman (U.S. Naval Acad., Annapolis, MD 21402)

Professor Robert T. Beyer's laboratory at Brown University transitioned in the mid-1970s to conduct experiments (under Navy funding) on the nonlinear interaction of sound with noise (by T.K.S.) and, shortly after, on the nonlinear interaction of sound with turbulence (by M.S.K.). In this presentation, T.K.S. and M.S.K. recall their experiences in Beyer's laboratory during their tenure as graduate students. Under Beyer's guidance and support both graduate students ("unofficially" numbers 20 and 21) built up their experiments—mostly from scratch—but were able to use some remnants left over from an earlier era of graduate students, which included some intricately made ultrasonic transducers and precise mechanical staging. Beyer's earlier graduate students' dissertations also proved to be an invaluable source of knowledge. T.K.S. and M.S.K. developed considerable apparatus for their research. Prototype electronic equipment in that era involved mostly analog electronics because of the significant challenges associated with building the digital version at that time. The development of equipment including the complex acoustic watt meter (T.K.S.) and the hot film anemometer (M.S.K.) will be presented along with descriptions of their measurements of nonlinear scattering. Professor Beyer's mentoring—the teacher, the historian, and storyteller—along with his fatherly guidance are heartfelt today.

3:15

2pPA5. Beyer's legacy in determination of nonlinear acoustic parameters. E. Carr Everbach (Eng. Dept., Swarthmore College, 500 College Ave., Swarthmore, PA 19081 ceverba1@swarthmore.edu)

Beyer's classic 1974 text Nonlinear Acoustics, which was first published in 1976 and was later reprinted in 1994, provided the pedagogical foundation for generations of productive acoustics work. This talk will focus on Beyer's contributions to the theory and measurement of acoustic nonlinearity parameters, and on the later research that grew out of it, especially efforts to quantify and employ the acoustic nonlinear parameter B/A in innovative imaging strategies.

3:30

2pPA6. Robert T. Beyer: A portrait in nonlinear acoustics. Murray S. Korman (Dept. of Phys., U.S. Naval Acad., Annapolis, MD 21402) and Thomas G. Muir (Natl. Ctr. for Physical Acoust., University, MS 38677)

Robert T. Beyer's early work in nonlinear acoustics is explored through his publications, his book, *Nonlinear Acoustics*, and his work with graduate students and colleagues. Beyer's work was at the crossroads of the fundamental physics of nonlinear acoustics and his dogged pursuit of the B/A constant provided the essential information necessary for the field to develop.

3:45—4:05 Panel Discussion

Session 2pSC

Speech Communication: Speech Dynamics, Methods, and Models (Poster Session)

Augustine Agwuele, Chair
Texas State Univ., Dept. of Anthropology, San Marcos, TX 78666

Contributed Papers

All posters will be on display from 1:00 p.m. to 4:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 2:30 p.m. and contributors of event-numbered papers will be at their posters from 2:30 p.m. to 4:00 p.m.

2pSC1. Kinematic modeling and acoustic measures of breathy voice. Robin A. Samlan, Brad H. Story, and Kate Bunton (Dept. Speech, Lang., Hrg. Sci., Univ. of Arizona, 1131 E. 2nd St., Tucson, AZ 85721-0071, rsamlan@email.arizona.edu)

A variety of disordered vocal fold structural and vibratory parameters have been identified in patients with breathy dysphonia, although the contribution of each parameter to voice quality is unknown. This study is part of an effort to determine the relation between production and voice quality. A kinematic vocal fold model [based on I. R. Titze, Speech Commun., **8**, 191–201 (1989)] that emulates the medial surface of the vocal folds was used to ascertain factors that lead to breathy voice. Based on the ratings of four expert listeners, five parameters were selected for further study. Combinations of multiple levels of adduction, bulging, nodal point, epilaryngeal area, and phase were used to generate 180 samples of sustained /ɑ/ and /ɪ/. Seven acoustic analyses were performed for each sample: maximum flow declination rate, maximum area declination rate, two measures of H1-H2, two measures of cepstral peak prominence, and noise to harmonics ratio. All parameters led to systematic changes in acoustic measures reflecting increasing breathiness, with several inconsistencies highlighting the uniqueness of each measure. The H1-H2 measures were inconsistent with the other data. Results will be described in light of how each measure can inform researchers about production. [Work supported in part by NIH R01-DC04789.]

2pSC2. VOICESAUCE: A program for voice analysis. Yen-Liang Shue (Dept. of Elec. Eng., Univ. of California, Los Angeles, 405 Hilgard Ave., Los Angeles, CA 90095, yshue@ee.ucla.edu), Patricia Keating, and Chad Vicenik (UCLA, Los Angeles, CA 90095-1543)

VOICESAUCE is a new application, implemented in MATLAB, which provides automated voice measurements over time from audio recordings. The measures currently computed are F0, H1(*), H2(*), H4(*), H1(*)-H2(*), H2(*)-H4(*), H1(*)-A1, H1(*)-A2, H1(*)-A3, energy, Cepstral Peak Prominence, F1–F4, and B1–B4, where (*) indicates that harmonic amplitudes are reported with and without corrections for formant frequencies and bandwidths [Iseli *et al.* (2006)]. Formant values are calculated using the Snack Sound Toolkit, while F0 is calculated using the STRAIGHT algorithm; harmonic spectra magnitudes are computed pitch-synchronously. VOICESAUCE takes as input a folder of wav files, and for each input wav file produces a MATLAB file with values every millsecond for all measures. It can operate over the whole input file or over segments delimited by a PRAAT textgrid file. VOICESAUCE then takes these MATLAB outputs, optionally along with electroglottographic measurements obtained separately from PCQUIRERX, and provides condensed outputs in text format; alternatively it can write the MATLAB outputs to the format used by the Emu Speech Database system. We compare results of VOICESAUCE analysis with manual measurements from FFT spectra and with measurements from a PRAAT script. [Work supported by NSF.]

2pSC3. Voice quality variation with fundamental frequency in English and Mandarin. Patricia Keating (Phonet. Lab., Dept. of Linguist., UCLA, Los Angeles, CA 90095-1543, keating@humnet.ucla.edu) and Yen-Liang Shue (Univ. of California, Los Angeles, Los Angeles, CA 90095)

Previous research has shown that F0 is positively related to H1*-H2* across male speakers of English [Iseli *et al.* (2006)] and to H1-H2 (after inverse filtering) within individual male speakers of Dutch [Swerts and Veldhuis (2001)]. That is, males who have overall higher-pitched voices generally have overall higher values of H1*-H2* (cross-speaker relation), and as an individual male's F0 goes up, H1-H2 generally also goes up (within-speaker relation). The present study investigates both of these relations, cross-speaker and within-speaker, for male and female speakers of English and Mandarin, and extends them to a large set of voice quality measures. The speech samples consist of repeated rising and falling tone sweeps, in which speakers began at a self-selected comfortable pitch, and then swept either up or down in pitch to their highest or lowest comfortable pitch. The beginnings of the sweeps are tested for cross-speaker relations, while the entire sweeps are tested for within-speaker relations. VOICESAUCE, a new program for voice analysis, is used to extract F0, energy, cepstral peak prominence, formants and bandwidths, and a variety of harmonic amplitude measures. Many measures are shown to be strongly related to F0. [Work supported by NSF.]

2pSC4. Voice quality and emotion classification on the valence scale. Ka Won Choi and Jeung-Yoon Choi (School of Elec. and Electron. Eng., Yonsei Univ., Seoul 120-749, Republic of Korea, toktok@dsp.yonsei.ac.kr)

While discriminating between several basic human emotions, such as neutral, joy, sadness, and anger, it has been observed that the most difficult emotions to tell apart automatically are between the two states of joy and anger. Since these two emotions have similarities on the arousal scale, it is difficult to distinguish them by simply using pitch and energy related feature measurements. Therefore, in this study, other additional feature parameters, related to voice quality, that are useful for discriminating between the two emotions of joy and anger are focused. For voice quality related features, global statistics of normalized spectral band energy, spectral tilt, open quotient, and first formant bandwidth values, along with their respective slopes and convexities, are measured from the happy and angry emotional speech from a Korean emotional database. From ANOVAtests, parameters of normalized spectral band energy, spectral tilt, and open quotient appear to be useful. Also, slopes and convexities of voice quality measurements appear to be more important than the values itself. These results are meaningful for classifying emotional states distributed on the valence scale and are expected to contribute in improving an overall emotion recognition system.

2pSC5. Automated extraction of prosodic features and analysis for emotional states. Suk-Myung Lee and Jeung-Yoon Choi (Yonsei Univ., 134 Sinchondong, Seodaemun-gu, 120-749 Seoul, Republic of Korea, pooh390@dsp.yonsei.ac.kr)

This study investigated the relationship between emotional states and prosody. A prosody detection algorithm [Choi *et al.*, J. Acous. Soc. Am. **188**, 2579–2587] was applied to extract accents and intonational boundaries automatically. The measurements used are derived from duration, pitch, harmonic structure, spectral tilt, and amplitude. Detection experiments on the Boston University Radio Speech Corpus show equal error detection rates around 70% for accent and intonational boundary detection. This algorithm was applied to a Korean emotional database subset, in which five sentences

were spoken by 15 speakers over four emotions: neutral, joy, sadness, and anger. By comparing the ratio of events that were detected as accent and intonational boundaries between neutral speech and emotional speech, our experiments find different distributions of these events for each emotion. In preliminary experiments, joy and anger tended to have fewer events classified as boundaries compared to other emotions. Also, joy and sadness have more events corresponding to accents. These results indicate that prosody detection can be useful for classification of emotion.

2pSC6. Modeling segmentation precision and inter-segmenter variability. Diana Stojanovic (Dept. of Linguist., Univ. of Hawaii at Manoa, 1890 East-West Rd., Honolulu, HI 96822, stojanov@hawaii.edu)

Inter-segmenter differences have been reported in the literature to affect the measurements of segment durations [Allen (1975)] and more specifically rhythm metrics [Mairano and Romano (2007)]. Segmenters may differ in criteria used for segmentation as well as resolution of uncertainty due to gradual change between segments (such as between nasalized vowel and coda nasal). In addition, small variations in the location of segmentation border (one period of the waveform, or 5–10 ms for adult speakers) are likely both between two segmenters and between different segmentation sessions of the same segmenter. This study investigates the effect of segmentation errors on the classification power of rhythm metrics. Errors are modeled to represent (1) small random variation and (2) resolution of larger interval uncertainties (20–30 ms). Simulations using the two scenarios were performed on samples of manually segmented speech. Results show that errors in scenario-1 do not affect the separation of languages in (%V, stdC), (rPVI-C, nPVI-V), and (%V, VarcoV) space, which implies that precision of less than 5 ms is not required in studies that use these measures. Errors in scenario-2 may be large enough to shift a language closer to the group it does not belong, depending on the frequency of such errors.

2pSC7. Aerodynamic modeling for concatenative speech synthesis. Kevin B. McGowan (Dept. of Linguist., Univ. of Michigan, 611 Tappan St., Ann Arbor, MI 48109, clunis@umich.edu)

Listeners can perceive and use a wide array of fine-grained phonetic details, including the detailed coarticulatory influences of adjacent sounds, when perceiving speech. Details like anticipatory nasalization in *can*, for example, potentially provide the listener with a rich network of informative cues and are a key to understanding listeners' ability to disambiguate speech sounds from seemingly ambiguous input. Unfortunately, these coarticulatory cues are generally missing or contradictory in the output of speech synthesis systems. These systems work by concatenating variable-length sound units chosen from a large database of recorded speech. Units are chosen to minimize two functions: the cost of aligning a particular unit with the desired speech output (target cost) and the cost of adjoining the next sound to the most recently selected unit (join cost). Generally, these costs are calculated using features which can be automatically extracted from the acoustic speech signal. A unit selection database is created, automatically segmented and automatically labeled with nasal and oral airflow feature vectors. These aerodynamic features are used as a proxy for articulatory information in the calculation of join and cost functions. Listeners' mean opinion scores are obtained on output from this system and a baseline acoustic system for comparison.

2pSC8. Normalization for vocal tract differences using long term average spectrum. Antonia D. Vitela, Andrew J. Lotto, and Brad H. Story (Speech, Lang. and Hearing Sci., Univ. of Arizona, 1131 E. 2nd St., Tucson, AZ 85721)

Differences in vocal tract size and anatomy lead to substantial differences in acoustic realization of vowels. Much of the variability in vocal tract area functions is captured in the area function of the "neutral" (non-articulating) vocal tract of each talker. A listener could perceptually normalize talker variability substantially by extracting the neutral vocal tract and using it as a referent. Alternatively, the listener could extract the auditory representation of the output of the neutral tract and use it as a referent for the auditory representation of the vowel space. We will present such a model of perceptual talker normalization for vowels and demonstrate its effectiveness for a variety of vocal tract shapes and lengths. Additionally, we will provide evidence that this normalization may be implemented in an auditory process that (i) extracts the long term average spectrum from running speech (providing an estimate of the neutral vowel), (ii) represents new input relative to this spectral average (providing normalization), and (iii) is not specific to speech. Additional analyses and perceptual results will demonstrate how this model can extend to normalization of consonant perception. [Work supported by NIH/NIDCD to A.J.L. and B.H.S. and Diversity Supplement to A.D.V.]

2pSC9. Analysis of the human phonation with different vocal fold geometries by applying a two-dimensional fully coupled fluid-solid-acoustic finite element scheme. Stefan Zoerner, Manfred Kaltenbacher (Dept. of Appl. Mechatronics, Univ. of Klagenfurt, Universit Str. 65-67, Klagenfurt, 9020, Austria, stefan.zoerner@uni-klu.ac.at), Reinhard Lerch, and Michael Dllinger (Univ. Erlangen-Nürnberg, Erlangen 91054, Germany)

A numerical two-dimensional model is presented simulating the human laryngeal voice production. Air flowing through the larynx interacts with the structural mechanics of the vocal folds forcing them to vibrate and induce sound. In turn, the vibrations of the vocal folds cause the airflow to pulsate which acts as a further sound source inside the larynx. This coupled field problem is modeled by utilizing an enhanced finite element method. The presented approach takes all three involved physical fields consisting of fluid mechanics, solid mechanics, acoustics, and their interactions fully into account. The results of these simulations clearly demonstrate the impact of different vocal fold geometries on the fluid field, e.g., changing the occurrence of the Coanda effect. In addition, influences on the produced sound signal resulting in modified main acoustic frequencies are observed. An additional eigenfrequency analysis on the vocal fold model was performed showing that the first eigenmode and the vibrational frequency in the transient case are identical. The first eigenmode was in the range of about 100 Hz depending on the vocal fold geometry.

2pSC10. Articulatory characteristics of coronal consonants in Argentine Spanish: An electropalatographic study. Alexei Kochetov (Dept. of Linguist., Univ. of Toronto, 130 St. George St., Toronto M5S 3H1, Canada, al.kochetov@utoronto.ca) and Laura Colantoni (Univ. of Toronto, Toronto M5S 1K7, Canada)

Previous articulatory investigations of Spanish have been largely limited to its Peninsular varieties. This study uses electropalatography (EPG) to investigate articulatory characteristics of coronal consonant contrasts in Argentine Spanish as part of a larger project examining phonetic variation across Spanish dialects. Simultaneous EPG and acoustic data were collected from four female speakers of Buenos Aires Spanish reading sentences with various intervocalic coronal consonants. Results revealed consistent differences in terms of anterior/posterior tongue placement and the amount of linguopalatal contact with the primary distinction between (denti-)alveolar and post-alveolar articulations within the classes of stops/affricates, fricatives, nasals, and laterals/rhotics. Inter-speaker variation was observed, however, in the articulation of some consonants, namely, in the apical or laminal realization of the post-alveolar fricative and in the palatal or palatalized alveolar realization of the post-alveolar nasal. These findings diverge from those reported for Peninsular Spanish [A. M. Fernández Planas, Estudios de fonética experimental, **16**, 11–80 (2007)], revealing some dialectal differences in the degree of fronting of denti-alveolars. Additionally, results suggest a sound change in progress in Argentine Spanish, such as the depalatalization of the palatal nasal. [Work supported by Connaught, University of Toronto.]

2pSC11. An acoustic and electroglottographic study of breathy phonation in Gujarati. Sameer ud Dowla Khan (Dept. of Linguist., Cornell Univ., 203 Morrill Hall, Ithaca, NY 14853, sameeruddowlakhan@gmail.com)

While it has long been established that breathy and modal vowels in Gujarati can be reliably distinguished based on the H1-H2 measure [Fischer-Jorgensen (1967); Bickley (1982)], and that Gujarati listeners attend solely to H1-H2 when distinguishing phonation types in Gujarati and in other languages [Bickley (1982); Esposito (2006)], new research [Khan and Thatte (2009)] suggests that in more casual speech styles, H1-H2 may play a smaller role in distinguishing breathy-modal minimal pairs. In such speech styles, additional acoustic measures (e.g., H1-A3, CPP, and rms energy) are more effective in distinguishing breathy and modal vowels for some

speakers. To more closely examine the phonetics of breathy vowels in casual speech, the current study examines both acoustic and electroglottographic data collected from naturalistic productions of breathy-modal minimal pairs. Preliminary data suggest that while all speakers distinguish modal and breathy vowels, the strategies used by each speaker to produce the contrast vary considerably. [Work supported by NSF.]

2pSC12. An acoustic and electroglottographic study of White Hmong phonation. Christina M. Esposito, Joseph Ptacek (Dept. of Linguist., Macalester College, 1600 Grand Ave., St. Paul, MN 55105, esposito@macalester.edu), and Sherrie Yang (UCLA, Los Angeles, CA 90095)

This study examines the phonation of White Hmong, a language with seven tones (traditionally described as high, mid, low, high-falling, mid-rising, low-falling, and mid-low), five of which are associated with modal phonation, and two of which are associated with non-modal phonation; the low-falling tone is creaky and mid-low tone is breathy. Thirty-three speakers were recorded producing words with all seven tones; 12 also made electroglottographic (EGG) recordings. Acoustic measures were cepstral peak prominence (CPP) and harmonic amplitudes $H1^*$ and $H2^*$, $H1^*$-$H2^*$, $H1^*$-$A1^*$, $H1^*$-$A2^*$, $H1^*$-$A3^*$, and $H2^*$-$H4^*$. EGG measures were closed quotient (CQ) and peak-closing velocity (PCV). Measures were made automatically using VOICESAUCE and PCQUIRERX. Results showed that none of the measures tested distinguished all three phonation types. However, several measures distinguished two categories: H1 distinguished creaky versus non-creaky, H1-H2 distinguished breathy from creaky, and the EGG measures CQ and PCV both distinguished breathy from non-breathy. $H1^*$-$A1^*$, $H1^*$-$A2^*$, $H1^*$-$A3^*$, and $H2^*$-$H4^*$ did not distinguish any of the phonation types. This suggests that phonation contrasts are realized across several phonetic dimensions in White Hmong. In addition, there was a gender difference in the production of phonation, with females having significantly higher CPP than males, suggesting that female phonations are noisier. [Work supported by NSF.]

2pSC13. Effects of altered sensorimotor function of the tongue on spectral characteristics of sibilants /s, z/. Juha-Pertti Laaksonen, Jana Rieger (Dept. of Speech Pathol. and Audiol., Univ. of Alberta, 2-70 Corbett Hall, Edmonton, AB T6G 2G4, Canada laaksone@ualberta.ca), Jeffrey Harris, and Hadi Seikaly (Univ. of Alberta, Edmonton, AB T6G 2B7, Canada)

Acoustic characteristics of sibilant sounds produced by 17 Canadian English-speaking female (6 patients) and male (11 patients) tongue cancer patients were studied. The patients had undergone a tongue resection of the anterior 2/3 of the tongue, and tongue reconstruction with a radial forearm free flap (RFFF). The acoustic data included measurements of the spectral moments (mean, standard deviation, skewness, and kurtosis) of sibilants /s, z/ analyzed from speech samples (stimulus sentences and Zoo Passage) which were recorded before the tongue resection, and 1 month, 6 months, and 12 months after the resection with RFFF reconstruction. Primarily, the tongue reconstruction with RFFF was found to result in temporary changes on the spectral characteristics such that speech output was found to approach the pre-operative speech quality over the 1-year period. In addition, RFFF reconstruction of the tongue was found to have some gender-specific effects in the acoustic characteristics analyzed. However, there was variability between patients, indicating individual compensatory and adaptive mechanisms for the altered sensorimotor function of the tongue following tongue reconstruction with RFFF. [The protocol for this study has been approved by the Health Research Ethics Board (HREB) of the University of Alberta (Edmonton, AB, Canada).]

2pSC14. Finding prosodic events using voice source measurements over multiple syllables. Ran Han and Jeung-Yoon Choi (Dept. of Elec. and Electron. Eng., Yonsei Univ., 132 Shinchon-dong, Seodaemun-gu, Seoul, 120-749, Korea, magnolia@dsp.yonsei.ac.kr)

Voice source measurements such as slope and convexity of pitch, harmonic structure, spectral tilt, and amplitude have been shown to be closely correlated with prosodic events, and using these measurements over multiple syllables has been shown to produce better performance than using measurements taken over a single syllable. In this study, these measurements for specific prosodic events, namely, for nine types of accents are more closed examined, observed in the Boston University Radio Speech Corpus. Examination of distributions of these measurements at various accent levels shows that the presence of a high accent is correlated with less negative slopes of these measurements and a low accent is correlated with more negative slopes. From ANOVA tests, analyses show that several measurements yield high significance in distinguishing among accent levels. This study also includes an examination of the distributions of these measurements across boundary levels, and results show that voice source measurements are useful in distinguishing among the various accent levels in prosody.

2pSC15. Aeroacoustics of voiceless fricatives: An *in-vitro* investigation on simplified vocal tract geometries. Olivier Estienne, Annemie Van Hirtum (Grenoble Universities), Hélène Bailliet (LEA, Université de Poitiers, ENSMA, CNRS), and Xavier Pelorson (Grenoble Universities, Grenoble 38100 France)

Production of fricatives consonants is a complex aeroacoustical problem involving the interaction of turbulent airflow with the particular geometry of the vocal tract. The current study introduces a simplified mechanical static vocal tract including a tongue replica and a rectangular thin obstacle downstream of the constriction formed by the tongue and the upper tract wall. First, pressure of the vocal tract flow and far-field noise are measured for different entrance volume velocities, longitudinal tongue positions, and obstacle heights. Changing the geometrical parameters induces spectral noticeable changes and, particulary, when the ratio obstacle height/tract height is close to 1, spectra are shaped like those of sibilant fricatives. Next, velocity-turbulence measurements in the same conditions by means of hot film anemometry are realized, which permits to find relations between aerodynamics and previous acoustical observations. It will also complete pressure data for the validation of future vocal tract flow models, which could serve for fricatives synthesis applications.

2pSC16. Auditory feedback and articulatory timing. Takashi Mitsuya (Dept. of Psych., Queen's Univ., 62 Arch St., Kingston, ON K7L3N6, Canada, takashi.mitsuya@queensu.ca), Ewen N. MacDonald, and Kevin G. Munhall (Queen's Univ., Kingston, ON, Canada)

Talkers listen to their own voice while they speak and use that feedback to monitor and control fine details of speech production. When auditory feedback is perturbed in real time, talkers spontaneously alter their speech production to compensate for the perturbation. Most research using real-time altered auditory feedback has focused on spectral manipulations of vowels with little attention devoted to temporal manipulations of consonants. In the present study, we examine the role of acoustic feedback in control of voice onset time (VOT). Utterances of the words "tip" and "dip" were recorded from native English speakers, and several representative productions were selected for each speaker. After this, talkers were asked to repeatedly produce either tip or dip. During these productions a real-time processing system was used to provide modified feedback through headphones. When talkers said one word, they simultaneously heard their own voice saying the other word. Results showed that the speakers compensated for the VOT perturbation such that they lengthened their VOT for /t/ when the VOT of the feedback was shorter (/d/). Based on these results, a comparison of the role of auditory feedback in controlling temporal and spectral aspects of speech production will be discussed.

2pSC17. Relationships among vowel formant discrimination, discrimination of intensity and frequency, and frequency selectivity. Leah Guempel, Kathleen O'Brien, and Chang Liu (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, Austin, TX 78712)

The goal of the present study was to investigate if vowel formant discrimination could be accounted for by intensity discrimination and frequency discrimination as well as by frequency selectivity. Intensity discrimination and frequency discrimination were measured for tonal signals at 250, 500, 1000, and 2000 Hz presented at 70-dB SPL for 21 young normal-hearing listeners with American English as their native language. Auditory filter was measured using notched noise method at the four frequencies above. In addition, thresholds of vowel formant discrimination were examined for F1 and F2 of three English vowels for the same group of listeners. Overall, thresholds of vowel formant discrimination had significantly modest correlation with intensity discrimination, but no significant correlations

with frequency discrimination and frequency selectivity. These results indicate that vowel formant discrimination may be associated with intensity discrimination of harmonics in speech sounds, but not determined by frequency discrimination and frequency selectivity. Other factors accounting for vowel formant discrimination will be discussed.

2pSC18. Segregation of co-channel speech signals using a least-squares approach. Srikanth Vishnubhotla and Carol Espy-Wilson (Inst. for Systems Res. and Elec. & Comput. Engg., Univ. of Maryland, College Park, MD 20742, srikanth@umd.edu)

In this work, an approach to segregate overlapping speech signals from a single-channel recording is presented, and its performance demonstrated for the two-speaker case. The approach relies on a previous algorithm to estimate the pitch frequencies of the participating speakers and then identify their voiced regions. Depending on the voiced-unvoiced combination of the two speakers, an appropriate over-determined system of equations is set up which is then solved to unravel the contributions of the two speakers to the speech mixture. In the voiced-voiced case, the harmonics of the two pitch frequencies are used to set up the basis set. For the voiced-unvoiced and unvoiced-unvoiced cases, a wavelet-based approach is used to construct the basis set. The segregated contributions are then assigned to the appropriate speakers by imposing the temporal continuity of the mel-frequency cepstral coefficients of these separated speech segments. The perceptual quality of the segregated speech is evaluated using the PESQ measure, which shows a significant improvement compared to the mixture signal. On evaluation on a speech recognition task, the algorithm shows comparable or better performance than the state-of-the-art algorithms, even when the target speaker of interest is significantly weaker than the masking speaker.

TUESDAY AFTERNOON, 27 OCTOBER 2009 RIO GRANDE CENTER, 1:00 TO 3:30 P.M.

Session 2pUW

Underwater Acoustics: Reverberation Measurements and Modeling II

Eric I. Thorsos, Cochair
Univ. of Washington, Applied Research Lab., 1013 NE 40th St., Seattle, WA 98105

John S. Perkins, Cochair
Naval Research Lab., Code 7140, Washington, D.C. 20375

Jason D. Summers, Cochair
SAIC, 10401 Fernwood Rd., Bethesda, MD 20817

Contributed Papers

1:00

2pUW1. Improving the efficiency of the comprehensive acoustic simulation system Navy standard reverberation model. Henry Weinberg (23 Colonial Dr., Waterford, CT 06385, chic@ct.metrocast) and Ruth E. Keenan (SAIC, Box 658, Mashpee, MA 02649)

Research projects that use supercomputers and real-time applications for at-sea exercises would benefit from more efficient code. This paper describes various approaches that significantly reduced the run time of the comprehensive acoustic simulation system (CASS) Navy standard reverberation model. These include computer upgrades, parallel processing, and reconfiguring code. For example, typical CASS run times of two 3-GHz computers differed by a factor of 10 because the faster computer had more efficient memory. Three types of parallelization were also tried. The easiest to implement used a platform that contained numerous blades with eight processors per blade. Reverberation runs were split along receiver bearing angles with several bearing angles per processor. The second parallelization attempted fine grain segmentation. This proved inefficient exchanging data between processors until the GRAB eigenray model was restructured. The third approach, OPENMP, has the advantage of using the same code on computers without multiple processors. Of course, there will be no improvement in efficiency. A disadvantage is that OPENMP is difficult to master. Improved hardware and multiple processors alone were insufficient to provide adequate results for many applications. The approach described here requires reverberation and clutter predictions along numerous receiver beams in seconds.

1:15

2pUW2. Modeling and observations of coherent effects in reverberation from the Clutter09 cruise. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

Reverberation is often assumed to be a process which decorrelates multipath, leading to the use of intensity models which ignore coherent interactions. In this talk experimental evidence of coherent effects in reverberation are demonstrated from data obtained during the Clutter09 cruise in the Straits of Sicily conducted by NURC, and the ephemeral nature of these effects is discussed. The theoretical basis for coherence between reverberation multipaths is described for both rough surface and volume scattering, including the theoretical basis and conditions for the appearance and characteristics of striations in the reverberation spectrograms.

1:30

2pUW3. Spectral roughness inferred from measured reverberation time series. Steven A. Stotts and Robert A. Koch (Appl. Res. Labs., The Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758, stotts@arlut.utexas.edu)

Reverberation data are often interpreted in terms of either Lambert's law or a scattering kernel representation with a specified roughness spectrum. The parameters in these descriptions are adjusted and combined with a propagation model to produce simulated reverberation time series that fit the data. An alternative is to retain the scattering description with two separable factors: a well-defined physics-based vertical angle factor that can be justified analytically and a roughness spectrum factor that is adjusted to reproduce the measured data. For this purpose the scattering description is based on a two-way coupled-mode formalism that was shown to accurately produce reverberation time series for a two-dimensional, rough-bottom Pekeris waveguide [Knobles, Reverberation Workshop I, Austin, TX (November 2006)]. For water-sediment interface scattering, the off-diagonal components of the vertical angle factor in the Born approximation were shown to be equivalent to the kernel adopted by the Reverberation Workshop [Stotts

and Koch, JASA Exp. Lett., EL242–EL248 (2008)]. Examples from the Reverberation Workshop problem sets will be used to demonstrate the fitting process for specific roughness spectra.

1:45

2pUW4. A practical approach towards a timely seafloor interaction database. Juan I. Arvelo, Jr. (Appl. Phys. Lab., The Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099), David Zeddies (Marine Acoust. Inc., Arlington, VA 22203), and William T. Ellison (Marine Acoust. Inc., Middletown, RI 02842)

A seafloor database of physics-based parameters may be considered an ideal resource. However, geoacoustic inversion to a model accurately describing the sediment structure, sub-bottom inhomogeneities, interface roughness, and anisotropy is still a very challenging research topic. While such geoacoustic inversion challenges are resolved, a more practical approach toward the generation of a timely seafloor interaction database is sorely needed. This practical approach must offer the flexibility to accommodate any physics-based and empirical model, direct measurement, and geoacoustic inversion from all research and operational sonars. It must also offer consistency with other seafloor-related databases, sufficiency to be useful to any sonar model, and completeness to model sonar systems across a wide frequency range. Empirical seafloor scattering models have a limited number of parameters to fit all measurements across grazing angle and frequency, while the most complete physics-based models have too many parameters for high-confidence geoacoustic inversion. Therefore, a more practical worldwide seafloor database must also offer enough flexibility to accurately reconstruct the seafloor reflection and scattering functions with a minimum number of parameters. Finally, this approach must quickly arrive at an unambiguous unique solution. Such an approach will be described and demonstrated. [This effort was supported by SPAWAR.]

2:00

2pUW5. Prediction of clutter on the Malta Plateau. P. L. Nielsen (NURC, V. S. Bartolomeo 400, 19126, La Spezia, Italy), M. Prior (CTBTO, Vienna Int. Ctr., P.O. Box 1200,1400 Vienna, Austria), and C. Harrison (NURC, V. S. Bartolomeo 400, 19126, La Spezia, Italy)

NURC has conducted a series of experiments on the Malta Plateau, Mediterranean Sea, with the focus on collecting environmental and acoustic clutter data. Strong correlation between the environmental and processed acoustic data has been established for certain regions which provides important information to enhance prediction capabilities and signal processing algorithms and to eventually reduce the clutter returns in the sonar processing chain. However, strong clutter-like returns are persistently observed from a confined location on the Plateau with no clear indication of significant scattering features, i.e., no evidence of bathymetric features, variations in sub-bottom stratification, or changes in bottom geoacoustic properties. Only the sub-bottom profiling shows inclusions in the upper 5 m of the sediment, and these inclusions are hypothesized to have slightly different geoacoustic properties than the surrounding. This hypothesis is presented and verified by backscattered time-series computations using a two-way prediction model applied in an environment similar to that of the acquired acoustic data. [Work supported by the NATO Undersea Research Centre, the CLUTTER JRP partners, and the Office of Naval Research OA321.]

2:15

2pUW6. Target echo and clutter calculations, including time spreading using normal-mode group velocities. Dale D. Ellis (DRDC Atlantic, P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, dale.ellis@drdc-rddc.gc.ca)

The shallow-water normal-mode reverberation approach [D. D. Ellis, J. Acoust. Soc. Am. **97**, 2804–2814 (1995)] can be extended to handle the echo from a target or discrete clutter objects. Normal modes are used to handle the propagation, and ray-mode analogies are used to handle the echo and scattering. The time spreading of the echo can be approximated using the modal group velocities. This description for the time spreading is approximate, but the computations are quite efficient, since the normal modes need only be calculated at one frequency. The target can be point-like and mode functions can be used, or the target can be considered more extended and the mode amplitude envelopes can be used. The formulation will be presented, along with a number of calculations based on problems from the ONR Reverberation Modeling Workshops. [Work supported in part by ONR, Code No. 321OA.]

2:30

2pUW7. Reverberation and scattering measurements from seabed clutter features using an autonomous undersea vehicle. Charles W. Holland (Appl. Res. Lab., The Penn State Univ., State College, PA, cwh10@psu.edu), Peter L. Nielsen, and Reginald Hollett (NATO Undersea Res. Ctr., La Spezia, Italy)

Acoustic reverberation in littoral ocean waveguides is often controlled by small-scale seabed heterogeneities. Thus, validation of reverberation models and development of realistic simulations require measurements of those scales. Clutter features for mid-frequency active sonars span scales from order 1–1000 m. The smaller scale seabed features O(1–10)m, however, are not generally represented in bathymetric data, even from modern hull-mounted multibeam systems. One promising tool for isolation and quantitative measurement of small-scale seabed features is an AUV with a source (800–3500 Hz) and towed array. Recent measurements from the Clutter09 experiment in the Straits of Sicily show that O(1–10)m scale seabed features (e.g., carbonate chimneys on mud volcanoes) are resolved and can be quantified. [Work supported by the Office of Naval Research OA321 and the NATO Undersea Research Centre.]

2:45

2pUW8. Scattering from superspheroidal acoustic objects. Christopher Feuillade (Laboratorio de Acústica, Universidad Tecnológica de Chile-INACAP, Sede Pérez Rosales, Brown Norte 290, Ñuñoa 779-0569, Santiago, Chile, chris.feuillade@gmail.com)

The extended boundary condition technique of Waterman [J. Acoust. Soc. Am. **45**, 1417–1429 (1969)] has been used to study scattering from extended axisymmetric acoustic objects. These objects are formed using the mathematical function for a "super-ellipse" [i.e., $(x/a)^s + (z/b)^s = 1$, where $s = 2n$, $n = 1, 2, 3, \ldots$] and revolving around the z-axis. For $s = 2$, the object is a spheroid with aspect ratio $\alpha = b/a$. As s increases, the shape of the object approaches a right circular cylinder of radius a and length $2b$. The method allows the scattered field to be accurately determined for all azimuthal angles as a function of frequency. The method is applied to the case of air-filled objects in water, which has importance for the interpretation of acoustic scattering from oceanic objects such as air-bubbles, the swim bladders of some fish, and zooplankton. It is found that the frequency increases with α, exactly as predicted using a geometrical method by Weston, and increases in a relatively minor way with \$s\$. In addition, the method shows that the monopole resonance, which leads to a spherically symmetric scattering distribution, continues to dominate low-frequency scattering even for cylindrically shaped, air-filled objects with an aspect ratio up to $\alpha = 40$ and $s = 32$.

3:00

2pUW9. A numerical study of rough interface reverberation suppression, clutter elimination and target resonance excitation in heterogeneous media with single channel iterative time reversal. Yingzi Ying, Li Ma, and Bingwen Sun (Inst. of Acoust., Chinese Acad. of Sci., 21 Beisihuanxi Rd., Beijing 100190, China, yingyz05@mails.gucas.ac.cn)

The presence of echo returns from the rough interface of two layered medium, and the clutter of volume scattering hinders the detection of target buried in heterogeneous media. The situations are often raised in ultrasonic breast tumor diagnosis and underwater acoustic buried mine detection. This work investigates the application of monostatic single channel iterative time reversal in mitigating the difficulties with a numerical study. Simulations based on pseudospectral finite-difference time-domain method are performed with a sphere buried in the heterogeneous media of lower layer, a transmitter is situated at the upper homogeneous domain, and the interface position is normally distributed. A wideband signal is launched to initiate the process, and the time-reversed echo received at same position is emanated as renewed interrogation pulse for next iteration. Some field snapshots are taken and the echo is recorded in each iteration. The results illustrate as the number of iteration increases, small volume scattering is eliminated, and rough interface reverberation is suppressed relatively. The echoes will con-

verge to a narrowband waveform corresponding to an object's dominant resonance mode. The detection of target is achieved by exploiting this important acoustic signature. [Work supported by the CAS Innovation Fund.]

3:15

2pUW10. Three-dimensional reverberation modeling. Edmund J. Sullivan (Prometheus Inc., Portsmouth, RI 02871, paddypriest@aol.com) and Carlos Godoy (Naval Undersea Warfare Ctr., Newport, RI, 02841)

Previously, reverberation modeling at sonar and torpedo frequencies was based on spectral factorization, which does not allow correct broadband, non-zero Doppler simulation in real time for a multi-channel receiver. This paper presents a new approach which avoids these limitations. Instead of transmitting the pulse itself, the method transmits a superposition of many copies of the pulse, each with a random amplitude and phase. This "noiselet" mimics the point-scatterer method without requiring a prohibitively large number of scattering points. We refer to it as the early randomization method (ERM). The propagation model used, the GRAB ray model, contains the bathymetry information, enabling the ERM to produce a statistically realistic reverberation field with a minimal number of rays. The computational load increases linearly with the number of receiver elements instead of with its cube, since there is no power spectral density matrix to factor. Since the randomness is introduced directly into the noiselet, the ERM is not limited to a particular statistical model, allowing non-Gaussian cases to be treated. Previously, this approach has shown excellent results for the two-dimensional, range-dependent case. Here, the method is generalized to also allow cross-range dependence. Comparisons with the CASS/GRAB model are shown.

TUESDAY EVENING, 28 OCTOBER 2009 7:30 TO 9:30 P.M.

Opening Meetings of Technical Committees

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held starting immediately after the Social Hours at 7:30 p.m. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings including proposals for special sessions, workshops and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussion

Committees meeting on Tuesday are as follows:

Acoustical Oceanography	Rio Grande East
Architectural Acoustics	Rio Grande Center
Engineering Acoustics	Pecan
Musical Acoustics	Live Oak
Physical Acoustics	Rio Grande West
Psychological & Physiological Acoustics	Blanco/Llano

Session 3aAA

Architectural Acoustics and Education in Acoustics: Architectural Acoustics Program Serving Consultants and Professionals

David S. Woolworth, Chair
Oxford Acoustics, Inc., 356 CR 102, Oxford, MS 38655

Chair's Introduction—8:00

Invited Papers

8:05

3aAA1. Undergraduate acoustical engineering programs at the University of Hartford. Robert D. Celmer and Michelle C. Vigeant (Acoust. Program and Lab., Mech. Eng. Dept., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117)

The University of Hartford has two undergraduate engineering programs in the area of acoustics: (1) the Bachelor of Science in Mechanical Engineering (BSME) with Acoustics Concentration and (2) the Bachelor of Science in Engineering with a major in Acoustical Engineering & Music. The first curriculum is part of a BSME degree program that has required courses in vibration as well as engineering acoustics since the 1960s. The Acoustical Engineering & Music degree is a unique program that was instituted in 1976, where applicants must have the equivalent math and science background required of all engineering students and also must successfully pass the entrance requirements of the Hartt School (University of Hartford's music conservatory)—including the audition. Both programs encompass the same engineering vibrations and acoustics courses, as well as the same acoustics projects sequence. Alumni of both undergraduate programs have successfully obtained positions in consulting (architectural and environmental) audio product and A/V design, musical instrument design, hearing- and psychoacoustic-related design, noise control, as well as pursued graduate degrees. The use of real world industry-sponsored acoustic projects for engineering design courses throughout the curriculum will be described, as well as the programs' continuous improvement plan and accreditation status.

8:25

3aAA2. Education in acoustics and noise control: A liberal arts perspective. Dominique J. Cheenne (Dept. of Audio Arts & Acoust., Columbia College Chicago, 33 E Congress, Chicago, IL 60605, dcheenne@colum.edu)

An education in acoustics and noise control is traditionally provided by engineering and/or graduate schools; however, since its inception in 1998 the Acoustics Program of Columbia College Chicago has graduated 99 students with a liberal arts background; 71 were hired by consulting firms in acoustics & noise control and by manufacturers, 65 are still employed in the field, 8 went on to graduate school, and 1 was recently voted into the "40 under 40" list of outstanding professionals by Consulting Specifying Engineer Magazine. Our program currently offers 14 undergraduate courses in acoustics and it draws upon the college's resources for calculus, physics, technical writing, management, accounting, architecture, and programming courses. We are in the second year of a scholarship program with the Institute of Acoustics at Tongji University, and we recently hosted the Highway Traffic Noise seminar from the National Highway Institute. Besides well-equipped test laboratories and a very wide array of computer modeling tools, a vinyl album cutting lathe was recently added to the program's inventory and starting in 2010, our students will be required to demonstrate that they have the skills to prepare for, schedule, and execute a very complex task that leaves no room for errors.

8:45

3aAA3. The University of Kansas architecture and architectural engineering programs offer focused coursework in architectural acoustics, electro-acoustics, and noise control. Robert C. Coffeen (School of Architecture, Design, and Planning, Univ. of Kansas, 1465 Jayhawk Blvd., Lawrence, KS 66045)

Courses relating to architectural acoustics, architectural noise control, mechanical and electrical noise control, and fixed installation electro-acoustic systems are presented by the architecture program of School of Architecture, Design, and Planning and by the architectural engineering program of the Department of Civil, Environmental, and Architectural Engineering, School of Engineering. This presentation will describe the courses taught by the author which have guided approximately 30 architecture and architectural engineering students plus a few electrical engineering students into professional acoustical consulting and sound system contracting. Courses taught in the 500–700 series include architectural acoustics, electro-acoustic systems, listening to architecture, senior design II acoustics emphasis, architectural acoustics for architects, and acoustic and theatrical design considerations for performance spaces. In addition, the flexible master of arts in architecture program works well for students entering the architecture program with an undergraduate degree in another field and desiring to have a career in acoustical consulting. To generate interest in acoustics, portions of several architecture and architectural engineering courses introduce building acoustics to first and second year students.

9:05

3aAA4. Acoustical education in architecture. Gary, W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611)

Sound can provide valuable contributions to the education and practice of architecture. First, architecture students should develop a theory of how sound and other environmental systems can become part of a larger theory of the making of architecture. Second, architects are trained to have highly developed sensibilities in the aesthetic design of space. Expanding the aesthetic component of buildings and environments to include the aural environment has the potential to enrich the fabric of communities and buildings. Third, architecture students must become acquainted with the materials, analytic techniques, and design approaches that acoustical consultants use. Furthermore, for acoustical information to become an integral part of architectural design and education, it must be organized in a way that allows expression or poesis to occur because these aesthetic principles are the building blocks of architectural design. Therefore, the science and engineering principles of noise control and architectural acoustics must be transformed into material that can become spatial, visual, and manipulable in creative ways by a design team. This requires that the basics of acoustics be reformulated in an architectural manner. This paper will address several of the transformations that have been attempted in the architectural acoustics curriculum at the University of Florida.

9:25

3aAA5. Graduate studies in acoustics at Purdue University. Patricia Davies, J. Stuart Bolton, Kai Ming Li (Ray W. Herrick Labs., Purdue Univ., 140 S. Martin Jischke Dr., West Lafayette, IN 47907-2031), Elizabeth Strickland, and George Wodicka (Purdue Univ., West Lafayette, IN 47907-2032)

The acoustics research and academic programs at Purdue can be divided into five broad categories: biomedical acoustics, engineering acoustics and noise control, physical acoustics and wave propagation, speech and hearing sciences, and acoustics in arts. There are 50 faculty from across 16 departments involved in various aspects of acoustics [http:/www.purdue.edu/research/phase]. The architectural engineering program has only very recently been started at Purdue, within the Civil Engineering, but there is a long history of Purdue graduates, particularly those in engineering, working in fields related to architectural engineering. One of the strengths of Purdue acoustics is this rich environment where students can study a broad range of topics in various disciplines each focused on different aspects of acoustics. Collaborations between engineering, liberal arts, psychology, and life science are a common theme in this community. An overview of the program is given but aspects of the program most closely related to architectural engineering will be described in more detail.

9:45

3aAA6. A broad education in acoustics still makes sense. Victor W. Sparrow and Karen P. Brooks (Grad. Program in Acoust., Penn State, 201 Appl. Sci. Bldg., Univ. Park, PA 16802, vws1@psu.edu)

For students preparing for a career in acoustical consulting, or for professionals thrust into job duties involving acoustics but without formal training, there is much to be gained by obtaining a broad education in acoustics, vibration, noise control, transducers, and signal processing. For the graduating student or the professional changing jobs, a broad education helps ensure that they will be adequately prepared for whatever they will face. Equipment and computer programs will come and go over time, but the ability to think clearly, solve problems, and understand important fundamental physical phenomena never goes out of style. At the Graduate Program in Acoustics at Penn State, our focus relies on a core curriculum that forms a solid foundation. Residence classes are currently blended (co-presented) for participation by students at a distance anywhere in the world using ADOBE CONNECT software and Smartboard presentation hardware. Distance students and guest presenters, e.g., practicing consultants, can also link in, making live presentations to everyone in the class. There is no reason why anyone with an internet connection and sufficient math and physics background cannot gain a broad education in acoustics to last a lifetime.

10:05—10:15 Break

10:15

3aAA7. Graduate Program in Architectural Acoustics at the School of Architecture, Rensselaer Polytechnic Institute. Ning Xiang and Christopher Jaffe (Graduate Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

Established in 1998, the Graduate Program in Architectural Acoustics at Rensselaer's School of Architecture has grown rapidly. The program offers both masters and doctoral degrees in a stimulating academic and research atmosphere. Applicants come from a variety of disciplines, including those with bachelor-level degrees (B. Arch., B.S., or B.A.) in Architecture, Music, Engineering, Audio /Recording Engineering, Physics, Mathematics, Computer Science, Acoustics, Electronic Media, Theater Technology, or related fields. The core curricula include Architectural Acoustics, Sonics Research Laboratory, Engineering Acoustics, Applied Psychoacoustics, and Communication Acoustics. Since its inception, the program has awarded more than 65 masters degrees and four doctorates. A majority of the program's graduates have developed careers in acoustics consulting. The emphasis of the program's curricula, its preparation for serving architectural acoustics consultants and professionals, and the research strategy at both the MS and Ph.D. degree levels will be presented.

10:35

3aAA8. Studying architectural acoustics through the University of Nebraska's Architectural Engineering Program. Lily M. Wang and Siu-Kit Lau (Architectural Engr. Prog., Peter Kiewit Inst., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, lwang4@unl.edu)

Architectural engineering (AE) is a discipline of study that is only currently available at 17 schools across the United States, and even fewer of these AE programs include acoustics as a main option in their curricula. This presentation will review the Nebraska Acoustics Group, housed within the AE Program at the University of Nebraska which began in 1998. Students may study acoustics within any of our five engineering degree programs (BSAE, MAE, MEng, MS and Ph.D.). Currently there are two AE faculties out of 11 who focus in acoustics at Nebraska, and the program regularly offers at least six recurring acoustics courses. Descriptions of the acoustics courses, the research interests of the Nebraska Acoustics Group, and where our graduates are to date will be given. Specifically highlighted will be the theme of our acoustics group: to promote the advancement and science of architectural acoustics by closely tying our coursework and research to practice in the "real-world."

10:55—12:00 Panel Discussion

WEDNESDAY MORNING, 28 OCTOBER 2009 REGENCY EAST 1, 8:45 TO 11:05 A.M.

Session 3aAB

Animal Bioacoustics: Acoustical Surveys, Sampling, and Population Assessment of Animals

David K. Mellinger, Chair
Oregon State Univ., Cooperative Inst. for Marine Resources Studies, 2030 SE Marine Science Dr., Newport, OR 97365

Chair's Introduction—8:45

Contributed Papers

8:50

3aAB1. Taming the Jez monster: Estimating fin whale spatial density using acoustic propagation modeling. David K. Mellinger, Elizabeth T. Küsel (Cooperative Inst. for Marine Resources Studies, Oregon State Univ., 2030 S. Marine Sci. Dr., Newport, OR 97365), Len Thomas, and Tiago A. Marques (Univ. of St. Andrews, Fife KY16 9AJ, Scotland)

The spatial density of calling animals may be estimated acoustically using methods presented by Buckland *et al.* [*Advanced Distance Sampling* (Oxford University Press; Oxford, 2004)]; information on the call (or other cue) rate, the call detection probability as a function of range, and the probability of a false detection to obtain an estimate of spatial density. Here we use similar methods to estimate the density of calling fin whales (*Balaenoptera physalus*) from the level of received sound near 20 Hz—the so-called "Jez monster." Using published source levels and call rates of fin whales, a Monte Carlo method is developed that simulates a given spatial density of whales randomly situated around a hydrophone and uses acoustic propagation modeling (specifically, a parabolic equation model) to estimate the resulting level of received sound. Using this technique at several deep-water sites in the North Atlantic, we derive a function that maps loudness to spatial density, and then use this function to estimate spatial density at these locations.

9:05

3aAB2. Detecting the clicks of beaked whales with wavelets. Mark Fischer (AguaSonic Acoust., P.O. Box 1073, Rio Vista, CA 94571-3073, aguasonic@gmail.com)

Most of the algorithms used for classifying cetaceans by call operate by analyzing acoustic data displayed by spectrographic analysis. Because spectrograms are based on Fourier transforms, this standardized technique works well at classifying the frequency modulated calls that Fourier transforms are most capable of illuminating. This includes the whistles and rhythmic pulses produced by many whale species. By contrast, the foraging clicks of Ziphidae consist almost entirely of exceedingly short duration, wide frequency, and fast-repeating click trains. Fourier-based algorithms provide a clumsy tool at best to interpret such transient data. Using the results of wavelet transforms with certain types of data, especially transient data, can vastly improve the task of classifying Ziphidae clicks. By calibrating the scales of individual transforms, wavelets can also be fine-tuned to provide a better match with particular kinds of clicks. The result of this optimization process is called a "wavelet profile," the combination of a wavelet transform with a range of scales matched to the clicks of a particular species. This paper explains a process for tuning and programming the selection of such profiles. The objective of this work is to suggest a better method for automatically classifying the calls of the cetacean species.

9:20

3aAB3. Classifying bowhead whale call types using linear discrimination function analysis and neural networks. Delphine Mathias, Aaron M. Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., San Diego, CA 92093-0238), Susanna B. Blackwell, Alice Green, Katherine H. Kim, Kristin Otte, Sara Tennant, Charles R. Greene, Jr. (Greeneridge Sci., Inc., Santa Barbara, CA 93117), and Michael Macrander (Shell Exploration and Production Co., Anchorage, AK 99503)

Bowhead whale calls are typically divided into five categories: upsweeps, downsweeps, U and N shaped undulated calls, and complex. The frequency range, duration, and details of the frequency modulation of these calls vary considerably, even within the same call type, creating difficulties when using matched-filtering or spectrogram correlation methods. A manually reviewed test dataset has been assembled that contains examples organized by call category and signal-to-noise ratio. Using a combination of contour tracing techniques, various features were extracted, including slope, curvature, number of inflection points, and the time of the first inflection point normalized by duration. Both linear discriminant functions and a feed-forward neural network were applied to discern appropriate decision boundaries for optimum statistical classifiers. The use of linear discriminant functions resulted in a 70% good classification for all classes except for the n-shaped calls. The neural networks improved the performance to 81% for upsweeps and downsweeps but had trouble classifying calls with significant curvature. A partial explanation for the classification difficulties stems from subjective definitions of what constitutes an inflection. Results of applying

the trained classifiers to bulk data collected in 2008 will be presented. [Work supported by Shell Exploration and Production Company.]

9:35

3aAB4. Large-scale evaluation of automated detection and localization of bowhead whale calls in the Beaufort Sea. Aaron M. Thode, Delphine Mathias (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., San Diego, CA 92109-0238), Christopher S. Nations, Trent L. McDonald (WEST, Inc., Cheyenne, WY 82001), and Michael Macrander (Shell Exploration and Production Co., Anchorage, AK 99503)

An automated procedure has been developed for automated detection and localization of bowhead whale sounds with arbitrary frequency-modulated tones and applied to 2008 data collected from 41 directional autonomous seafloor acoustic recorders deployed over a 280-km swath in the Beaufort Sea. The procedure has seven sequential stages: an incoherent spectral band detector, an interval estimator, a feature extractor, a feed-forward neural network classifier, a stage to flag common calls among recorders, bearing extraction, and final localization. The three-layer neural network has ten hidden-layer units and uses 20 features extracted from the spectrogram classifying signals as either whale calls or other sounds. Manual analysis was conducted on recordings from six non-consecutive days chosen to span a variety of seismic activity levels. The first 12 h of each day yielded 141 796 true calls and 1.15×10^6 other signals that were used to train the network, with 60% of that data used to adjust the weights and 40% used to evaluate convergence. The performance of both the neural network stage and the complete tracking algorithm on the remaining 12 h of manually-analyzed data from each day will be presented. [Work supported by Shell Exploration and Production Company.]

9:50

3aAB5. Automatic detection of marine mammals using spectral entropy. Christine Erbe (JASCO Appl. Sci., P.O. Box 4037, Brisbane Technol. Park, Eight Mile Plains, Queensland 4113, Australia)

An automatic detector for tonal and narrow-band signals of a variety of co-habiting marine mammal species is presented. The detector was developed to automatically identify signals in years of autonomous underwater recordings made in the Arctic. These recordings contained calls of odontocetes, mysticetes, and pinnipeds. Three types of automatic detectors were tested; two were based on peak-energy detection, and one was based on spectral entropy computation. A set of test data was assembled by active listening and visual discrimination of spectrograms. Performances of the automatic detectors were compared by means of receiver operating characteristics. The entropy detector outperformed the other two. It ran considerably faster than real time and can be used as a first step in automatic signal analysis, quickly scanning recordings for potential signals. It should be followed by automatic classification, recognition, and identification algorithms to group and identify signals.

10:05—10:20 Break

10:20

3aAB6. Effects of seismic exploration in the Beaufort Sea on bowhead whale call distributions. Christopher S. Nations (WEST, Inc., 2003 Central Ave., Cheyenne, WY 82001, cnations@west-inc.com), Susanna B. Blackwell, Katherine H. Kim (Greeneridge Sci., Inc., Goleta, CA 93117), Aaron M. Thode (Scripps Inst. of Oceanogr., San Diego, CA 92109-0238), Charles R. Greene, Jr. (Greeneridge Sci., Inc., Goleta, CA 93117), A. Michael Macrander (Shell Exploration and Production Co., Anchorage, AK 99503), and Trent L. McDonald (WEST, Inc., 2003 Central Ave., Cheyenne, WY 82001)

We analyzed locations of bowhead whale (*Balaena mysticetus*) calls in the Beaufort Sea during the fall migrations of 2007 and 2008, including periods of seismic exploration. Analysis examined relationships between received sound levels from airgun pulses and both spatial and temporal distributions of whale call locations. Locations were estimated by triangulation of calls received on Directional Autonomous Seafloor Acoustic Recorders (DASARs) arranged in five arrays along 280 km of Alaska's coast. Received levels either were measured directly by DASARs or estimated from transmission loss equations. For analysis, a grid was overlaid on the study region, and the 7-week study period was divided into short, regular intervals. Whale calls were counted and received levels estimated within each grid cell and time interval. The relationship between call count and received level was assessed using statistical models that accounted for anthropogenic sounds, date, time of day, bathymetry, and east-west position along the coast. Furthermore, the models incorporated adjustments for detectability (DASARs are less likely to detect more distant calls) and spatial and temporal covariances. Preliminary results indicate declines in calling rate and temporary onshore and offshore shifts in location as received levels increase. [Work supported by Shell Exploration and Production Company.]

10:35

3aAB7. Quantifying masking effects of seismic survey reverberation off the Alaskan North Slope. Melania Guerra, Aaron M. Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., San Diego, CA 92093-0238), Susanna B. Blackwell, Charles R. Greene, Jr. (Greeneridge Sci., Inc., 1411 Firestone Rd., Santa Barbara, CA 93117), and Michael Macrander (Shell Exploration and Production Co., 3601 C St., Ste. 1000, Anchorage, AK 99503)

Quantifying the potential impacts of anthropogenic acoustic noise on marine life faces tremendous hurdles, including defining appropriate metrics to be extracted from the raw acoustic data. In shallow Arctic environments, reverberation from seismic arrays can persist over periods longer than the duration of the direct signal, suggesting that reverberation might play a role in masking communication between endemic animals like the bowhead whale (*Balaena mysticetus*). Presented here is a quantitative metric related to the masking properties of impulsive anthropogenic signals in shallow water environments. The approach taken is to measure the minimum level of the acoustic noise field, where the level is typically defined in terms of rms pressure or sound exposure. Three time scales need to be defined in order to specify the minimum level: the biologically-relevant energy-integration timescale of a species's hearing mechanism, the time-scale over which the noise signal can be considered wide-sense stationary, and the time-scale over which significant secular changes take place in the source, receiver, or propagation characteristics of a cyclical (multi-pulse) scenario. Examples of this metric are provided using data collected in 2008 from seismic surveys in the Beaufort Sea in 25–50-m depth water. [Work supported by Shell Exploration and Production Company.]

10:50

3aAB8. Developing acoustical criteria for rodent housing facilities. Kimberly Lefkowitz, Robert Bruce, Jessica Arellano (CSTI Acoust., 16155 Park Row, Houston, TX 77084, kim@cstiacoustics.com), Travis Burgess, Stacy LeBlanc, Arnold J. Granger, Andrew T. Pesek, Jr., Dzung X. Pham, and David Bammerlin (M.D. Anderson, Houston, TX 77030)

Medical researchers often use mice and rats for numerous procedures. In typical rat housing rooms, there can be as many as 200 rats for an investment of $800,000 and in typical mouse rooms, as many as 2570 mice for an investment of $10 million. Do the sound and ultrasound exposures of these animals interfere with the medical research objectives of the scientists? Research has shown that high-noise levels have an effect on the physical and psychological responses of rodents. In order to eliminate this factor from influencing the medical experiments, criteria limiting noise and ultrasound in animal-housing facilities are needed. Mice have a hearing range of 1000–91 000 Hz, rats from 200 to 76 000 Hz, and other rodents can hear well above the human threshold. This makes creating a standard even more important because the majority of the frequencies perceived by rodents, including frequencies in which they communicate, are not perceptible to humans and therefore cannot be easily assessed. This study considers typical environments where these animals are housed, the levels at which mice are effected by noise and ultrasound, and the existing guidelines in non-US countries. Preliminary noise and ultrasound criteria are presented.

Session 3aAO

Acoustical Oceanography: Acoustic Measurements of Water Column Scatterers Using Multibeam Sonar

Thomas C. Weber, Cochair
Univ. of New Hampshire, Ctr. for Coastal and Ocean Mapping, 24 Colovos Rd., Durham. NH 03824

Ralph A. Stephen, Cochair
Woods Hole Oceanographic Inst., 360 Woods Hole Rd., Woods Hole, MA 02543-1542

Chair's Introduction—8:15

Invited Papers

8:20

3aAO1. Seismic oceanography: Low-frequency acoustical imaging of oceanic fine-structure. W. Steven Holbrook (Dept. of Geology and Geophys., Univ. of Wyoming, Laramie, WY 82071) and Raymond W. Schmitt (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Seismic oceanography, the processing of low-frequency (10–200 Hz) marine reflection data with an emphasis on water-column reflections, has opened up new ways of visualizing thermohaline finestructure. Studies to date have produced stunning images of fronts, eddies, water mass boundaries, and internal waves. As it approaches its sixth birthday as a discipline, seismic oceanography is entering a critical stage of development. An early rush to document the physical basis for the reflections and to catalog images in different oceanic environments is yielding to an effort to extract useful, trustable, and quantitative information on physical oceanographic processes from the images. Here we review progress to date and point to key areas of current and future works. Promising areas of research include emerging techniques to quantify internal wave energy and turbulence dissipation from seismic images, the acquisition of industry SO data, and the production of three-dimensional and time-lapse images of finestructure. A principal challenge for the future is the merging of synthetic and field seismic data with realistic physical models of oceanic temperature/density structure, calculated at the dense horizontal and vertical spacings needed to simulate the seismic data.

8:40

3aAO2. Population density imagery of migrating herring shoals by ocean acoustic waveguide remote sensing. Zheng Gong, Mark Andrews, Duong Tran, Daniel Cocuzzo, Saumitro Dasgupta (Dept. of Elec. and Com. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02176, zgong@ece.neu.edu), Srinivasan Jagannathan, Ioannis Bertsatos, Deanelle Symonds, Tianrun Chen (Massachusetts Inst. of Tech., Cambridge, MA 02139), Hector Pena, Ruben Patel, Olav Rune Godoe (Inst. of Marine Res., Nordnes, N-5817 Bergen, Norway), Redwood W. Nero (Southeast Fisheries Sci. Ctr., Pascagoula, MS 39568), J. Michael Jech (Northeast Fisheries Sci. Ctr., Woods Hole, MA 02543), Nicholas C. Makris (Massachusetts Inst. of Tech., Cambridge, MA 02139), and Purnima Ratilal (Northeastern Univ., Boston, MA 02115)

Ocean acoustic waveguide remote sensing (OAWRS) is a bistatic multibeam wide area sonar system that enables unaliased monitoring of fish populations over ecosystem scales. This allows us to quantify the formation processes of vast shoals of Atlantic herring (*Clupea harengus*) containing hundreds of millions of individuals, imaged during the Autumn 2006 spawning season. Areal population density estimation requires calibration of the low-frequency target strength of an individual shoaling herring. This is estimated from experimental data acquired simultaneously at multiple frequencies in the 300–1200-Hz range using (1) the OAWRS system, (2) areal population density calibration with several conventional fish finding sonar systems (CFFS), (3) fish length distributions obtained from trawl samples, and (4) local low-frequency transmission loss measurements. High spatial-temporal co-registration was found between shoals imaged by OAWRS and concurrent CFFS line transects, which also provided fish depth distributions. The mean scattering cross-section of an individual shoaling herring is found to exhibit a strong dependence on frequency in the OAWRS range, which is consistent with resonant scattering from an air-filled swimbladder. We show that a rapid transition from disordered to highly synchronized behavior occurs as the herring population density reaches a critical value, after which an organized group migration occurs.

9:00

3aAO3. Efficacy of using multiple acoustic systems to quantify Atlantic herring (*Clupea harengus*) aggregation metrics before and after fishing events. Jason D. Stockwell (Gulf of Maine Res. Inst., 350 Commercial St., Portland, ME 04101, jstockwell@gmri.org), Thomas C. Weber (Univ. of New Hampshire, Durham, NH 03824), J. Michael Jech (Natl. Marine Fisheries Service, Woods Hole, MA 02543), Adam J. Baukus, and Daniel J. Salerno (Gulf of Maine Res. Inst., Portland, ME 04101)

The use of midwater trawls to harvest Atlantic herring (*Clupea harengus*) in the Gulf of Maine has been a controversial topic for many years. A majority of the concern revolves around the potential physical disruption of Atlantic herring aggregations and the potential impact on ecological processes and other industries (e.g., other fisheries and eco-tourism) that rely on these aggregations. Preliminary acoustic data collection in fall 2008 using a combination of Simrad ES60 echosounders (38 and 120 kHz) and a Simrad SP90 omni-directional sonar (20–30 kHz) showed promise for quantifying herring aggregation metrics. In this presentation, we report on a pilot study conducted in the summer of 2009 that uses these acoustic systems to quantify and describe herring aggregations before and

after midwater trawling with a pair of fishing vessels. We used a systematic grid design to survey an area continuously for 24 h both before and after pair trawl fishing over multiple fishing trips. We report on the efficacy of using acoustic systems to quantify the potential impacts of fishing on herring aggregations and discuss the next steps to fully evaluate these potential impacts at spatial and temporal scales relevant to the fishery.

9:20

3aAO4. Characterization of pelagic scatterers using multibeam echosounder data: Echo amplitude and phase, and their variabilities and frequency spectra. David A. Demer, George R. Cutter (Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037, david.demer@noaa.gov), Thomas C. Weber (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, Durham, NH 03824), and Christopher D. Wilson (U.S. Natl. Marine Fisheries Service, NOAA, NE, Seattle, WA 98115)

A recently-developed statistical-spectral approach to acoustic-target identification (SSID) incorporates information contained in the frequency-dependent signal amplitudes and their variances. In addition to identifying biological targets, the SSID has demonstrated utility for estimating fish aggregation densities, abundances, and behaviors, and detecting and classifying the seabed (e.g., accurate depth, within-beam slope, hardness and roughness, and the height of the unresolved boundary region, and the so-called dead zone). Further, the related multi-frequency, bi-planar, interferometric technique (MBI) allows much higher-resolution measurements of the seabed and sub-beam measures of slope, hardness, and roughness. Here we extend the utilities of the SSID and MBI methods, with explicit consideration of the single-frequency signal phases and their variances, and frequency-dependent split-aperture phases and their variances, for improved estimations of pelagic-target sizes, and their aggregation densities, behaviors, and shapes (e.g., surface rugosity). Data from the Simrad ME70, split-aperture, multibeam echosounder allow these methods to be applied to data collected over a broad bandwidth and range of incidence angles for more classification and observation possibilities.

9:40

3aAO5. Variable-aperture processing of multibeam echosounder data to better resolve fish locations and improve abundance estimates. George R. Cutter, Jr. and David A. Demer (NOAA Southwest Fisheries Sci. Ctr., 8604 La Jolla Shores Dr., La Jolla, CA 92037, george.cutter@noaa.gov)

Acoustic classifications of fish and estimations of their orientation distributions are possible using multi-frequency or broad bandwidth measurements of their acoustic target strengths (TSs) with knowledge of their scattering directivity pattern and size distributions. To measure TS of *in situ* fish, single-frequency interferometric methods provide information to detect resolvable single targets and estimate their location within the acoustic beam. This technique is compromised by multiple targets that are unresolvable because of their spacing, but this situation can be mitigated with the use of multi-frequency interferometry. The ambiguity caused by coincident echoes can also be substantially reduced using single-frequency, multiple-aperture interferometry. This method uses phase differences from multiple sub-arrays of a single-frequency (200-kHz) multibeam echosounder to estimate robustly when echoes originate from a resolvable single target and its position within the acoustic beams. Results provide accurate measures of beam-compensated TS and, in cases of low-density scatterers, estimates of their volume density. Multiple-aperture interferometry can significantly improve the utility of single-frequency multibeam echosounders for quantitative measures of fish and zooplankton, and seabed-range detections.

10:00—10:15 Break

10:15

3aAO6. Measuring suspended sediment concentrations and flow velocities using multibeam sonar. Stephen M. Simmons (Aquatec Group Ltd., High St., Hartley Wintney, Hampshire RG27 8NY, United Kingdom, ssimmons@aquatecgroup.com), Jim L. Best (Univ. of Illinois, Urbana-Champaign, IL), Dan R. Parsons, Gareth M. Keevil (School of Earth and Environment, Univ. of Leeds, United Kingdom), Kevin A. Oberg (USGS Office of Surface Water, Urbana, IL), Chris Malzone (Myriax, Inc., San Diego, CA), Kevin K. Johnson, and Jonathan Czuba (USGS Illinois Water Sci. Ctr., Urbana, IL)

Modern data handling and storage technologies facilitate the logging of the large quantity of water-column backscatter information received by multibeam sonars. Methods of using these data to derive estimates of the mass concentration and flow velocities of suspended sediment flow structures have been developed. The results obtained by the application of these methodologies to data collected at the confluence of the Parana and Paraguay rivers in Argentina and the confluence of the Mississippi and Missouri rivers in the United States will be presented. An analysis of those data in conjunction with a set of experimental data collected in a large-scale test facility will be also given. The applicability and limitations of the use of multibeam sonar for deriving suspended sediment concentrations will be discussed. By enabling the simultaneous measurements of suspended sediment concentration, flow velocities, and bathymetric data, multibeam echo-sounders are demonstrated to be a versatile tool for the surveying of the various components relating to the transport of sediment in fluvial systems.

10:35

3aAO7. Phase and amplitude monopulse techniques to increase the accuracy of within-beam bearing estimates of volume scatterers. Daniel S. Brogan and Kent A. Chamberlin (Dept. of Elec. and Comput. Eng., Univ. of New Hampshire, 33 Academic Way, Durham, NH 03824-2619, daniel .brogan@unh.edu)

Traditionally, designers of multibeam sonars have limited the accuracy of echo angle-of-arrival (bearing) estimates by mapping each measured echo to its beam's maximum response axis. Although this bearing accuracy can be improved for large targets, e.g., the ocean floor, by applying filtering techniques to the time series of bearing estimates, the echoes from water column scatterers are often only a few time samples in duration. Therefore, it is beneficial to both increase the quantity and improve the accuracy of these few echo bearing estimates. Phase and amplitude monopulse techniques use linear mappings between the sine of the echo bearing and (1) the differential signal phase and (2) the natural logarithm of the ratio of the signal magnitudes, respectively, of neighboring beams to provide within-beam echo bearing estimates. Using either of these techniques, the accuracy of the bearing estimates of the echo's acoustic center is improved compared to maximum-response-axis mapping. By using both techniques simultaneously, the effective number of echo bearing estimates is doubled. These techniques are applicable to both planar and cylindrical multibeam sonar systems.

10:50

3aAO8. Use of a multibeam sonar to characterize fish aggregations in the northern Pacific. Sarah C. Stienessen (NOAA-AFSC, Bldg. 4, 7600 Sand Point Way, NE, Seattle, WA 98115), Thomas C. Weber (Univ. of New Hampshire, Durham, NH 03824), and Christopher D. Wilson (NOAA-AFSC, Seattle, WA 98116)

Acoustic observations to describe small-scale spatial patterns or aggregations of fish have traditionally been made in two dimensions using vertical echosounders. Recently, multibeam sonars have been utilized as a tool to study the three-dimensional structure of fish aggregations. This study describes small-scale spatial patterns of juvenile walleye pollock, Theragra chalcogramma, with data collected using a calibrated ME70 multibeam sonar. The ME70 has reduced side-lobes and a very high dynamic range, which allow better characterization of the water column backscatter than typical multibeam systems. For this study, the sonar was configured with a 120 deg swath composed of 31 narrow bandwidth beams, ranging from 70–120 kHz. Data were collected during winter 2009 in the Gulf of Alaska and summer 2009 in the Bering Sea to evaluate whether seasonal and geographical differences existed in the juvenile fish aggregation patterns. Size- and shape-related aggregation descriptors (e.g., height, length, volume, and surface area) are described, and the variability in these metrics as a function of time and space is explored. Particular challenges for using the ME70 as a tool to describe the small-scale spatial patterns of marine animals are highlighted.

11:05

3aAO9. Single- and multi-beam echosounder assessment of Antarctic krill aggregations. Joseph D. Warren (School of Marine and Atmospheric Sci., Stony Brook Univ., 239 Montauk Hwy., Southampton, NY 11968), Martin J. Cox (Univ. of St. Andrews, Fife KY16 8LB, United Kingdom), David A. Demer, George R. Cutter (Southwest Fisheries Sci. Ctr., La Jolla, CA 92037), and Andrew S. Brierley (Univ. of St. Andrews, Fife KY16 8LB, United Kingdom)

Acoustic backscatter measurements from aggregations of Antarctic krill were collected over two field seasons (2006 and 2007) in the waters north of Livingston Island using two 5.5-m inflatable vessels. One was equipped with a dual-frequency (38- and 200-kHz) single-beam echosounder, while the other used a 200-kHz multi-beam system. In addition to independent survey operations, both vessels participated in joint survey operations where the two boats followed the same trackline separated by a small distance (10–40 m). These data allow us to compare how the two echosounder systems "see" the same (or very similar) krill aggregations. During 2006, some tracklines were also occupied by a large (300-ft) vessel although vessel separation distances were larger during these operations. The combination of measurements from single- and multi-beam echosounders (and nearby net tows) provides the opportunity to characterize krill aggregation size, shape, numerical density, and overall biomass. Observations of krill predators were made from both small vessels during these surveys. Results discussing the interactions between krill predators and krill aggregations will be presented.

Session 3aEA

Engineering Acoustics and Signal Processing in Acoustics: Geophysical Acoustics

Fernando Garciá-Osuna, Chair
Schlumberger, Technology Innovation, 110 Sclumberger Dr., Sugarland, TX 77478

Chair's Introduction—8:05

Invited Paper

8:10

3aEA1. A low-frequency underwater sound source for seismic exploration. Elmer L. Hixson (Dept. of Elec. and Comput. Eng., Univ. of Texas, 1 Univ. Station 0803, Austin, TX 7712, ehixson@mail.utexas.edu)

A very low-frequency (1–10 Hz) sound source is needed for greater penetration in the earth below a water layer. A neutral buoyant acoustic dipole is proposed for this purpose. With the dipole positioned vertically near the pressure release surface, the upward wave produced is inverted and adds to the downward wave. A model of such a dipole to operate in the range 20–40 Hz was built. It was driven by an inertial force generator. It was tested in an 11-m deep tank under low signal/noise conditions. Only near field pressure and water loading could be measured. In order to get design parameters for a 1–10 Hz dipole, scaled experiments using dipoles operating in the 3–5 kHz region with adjustable source spacing and depth were built and tested. A trade-off between source spacing for greater sound pressure and depth for summing the two waves were studied. In the deep tank, near- and far-field pressures were measured. [This study was supported by EXXONMOBIL.]

Contributed Papers

8:30

3aEA2. Shear waves and grain-rolling in granular media. Nicholas P. Chotiros and Marcia J. Isakson (Appl. Res. Lab., The Univ. of Texas at Austin, TX 78713-8029, chotiros@arlut.utexas.edu)

In a randomly packed unconsolidated granular medium, there is a random asymmetry in mechanical coupling between grains that has been shown to cause an increase in the effective inertia, hence a reduction in sound and shear wave speeds [N. P. Chotiros and M. J. Isakson, J. Acoust. Soc. Am., **121**, EL70–EL76 (2007)]. The increase in inertia for shear waves is revisited because, in addition to the incoherent motion due to the random asymmetry, there is an additional component due to coherent grain rotation. Shear wave speed is predicted to be dependent on grain diameter, as well as the contact stiffness. [Work supported by ONR, Ocean Acoustics.]

8:45

3aEA3. Geoacoustic inversion using specific acoustic impedance. Steven E. Crocker (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841), James H. Miller (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882), Kevin B. Smith (Naval Postgrad. School, Monterey, CA 93943), Paul C. Hines, and John C. Osler (Defence R&D Canada Atlantic, Dartmouth, NS B2Y 3Z7, Canada)

An inversion method for the estimation of geoacoustic properties of the ocean bottom using data from a small number of acoustic vector sensors was developed. The experimental data used for the inversions were short duration, gated continuous wave transmissions acquired on four acoustic vector sensors that spanned the water-sediment interface during the Sediment Acoustics Experiment 2004 (SAX04) conducted in the Gulf of Mexico. The acoustic signals were analyzed for their specific acoustic impedance over a bandwidth ranging from 600 to 2400 Hz. The specific acoustic impedance was obtained from simultaneous measurements of the acoustic pressure and particle acceleration for arrivals at normal incidence. A differential evolution algorithm was used for the inversion of the depth dependent sediment properties. A wave number integration routine was used to compute the specific acoustic impedance corresponding to the estimated sediment parameters. Experiment details, forward modeling, inversion methods, and results are discussed. [Work supported by ONR.]

9:00

3aEA4. Exploiting geophysical phenomena in processing seismic array data for border subsurface surveillance. Jason D. Holmes, Erin, M. Aylward, Peter A. Krumhansl, Richard J. McNeil, Darren R. Fabbri, Christopher S. Fortin (BBN Technologies, 10 Moulton St., Cambridge, MA 02138), Jason R. McKenna, Michael S. Mattice, and Sarah McComas (U.S. Army Engineer Res. and Development Ctr., Vicksburg, MS 39180)

The ability to detect tunneling activity under borders and perimeters has become an increasingly important problem. Persistent, automated, passive surveillance of expansive areas for tunnel activity is a technical challenge because of the processing and hardware costs. Further, data bandwidth concerns prohibit transmission of large amounts of time-series data to a central processing location but there remains a desire for human review of suspect signals and archiving to develop trends in time. To meet these technical challenges, sensor array and signal processing approaches which exploit geophysical phenomena to distinguish tunnel activity from surface noise were developed to rapidly and reliably detect potential subsurface threat signals with a minimized set of sensors, very low data bandwidth requirements, and a reduced set of centralized processing hardware. Time series review and archiving can then be performed for limited sets of the most suspect data. Analysis and results from recent controlled tunneling experiments on an operational system will be presented. [Work sponsored by U.S. Army Engineer Research and Development Center.]

9:15

3aEA5. Biologically inspired ultrasonic measurements of porous media, with applications in geophysics. C. Hopper, S. Assous, M. A. Lovell (Dept. of Geology, Univ. of Leicester, Univ. Rd., Leicester LE1 7RH, United Kingdom), D. A. Gunn, P. D. Jackson, and J. G. Rees (Br. Geological Survey, Nottingham NG12 5GG, United Kingdom)

Biologically inspired ultrasound has been investigated for measuring properties of materials in an underwater environment. Broadband transducers have been deployed which operate in the 40–200-kHz frequency range, with similar frequencies to those used by some echolocating mammals. Signals have been designed, which optimize the available bandwidth of the transducer, and analysis procedures have been developed to extract the desired information from acquired data. Measurements on single-layer targets,

comprising plastic, metal, and glass, in solid and porous forms have established the performance of the system on well-controlled synthetic samples. The target thicknesses ranged from 0.5 to 40 mm, thus spanning the wavelength range of the signals used. Adequate penetration of the ultrasound into the samples, at the frequencies used, was demonstrated. Material and thickness discrimination was possible using frequency domain results, and modeling of the traces was performed in order to extract velocity and attenuation information. Scanning electron microscopy measurements on the porous targets revealed structural information that informs the interpretation of the ultrasonic results. The work on synthetic targets forms a basis to inform experimentation on natural geological materials exhibiting a wide range of structural characteristics. Initial results for sandstone samples are presented.

9:30

3aEA6. Scattering properties of a time-evolving multiple scattering elastic medium. Alexey Sukhovich, Philippe Roux, and Eric Larose (LGIT, Universit Joseph Fourier, 1381 rue de la Piscine, UMR CNRS 5559, Grenoble, France)

We investigate the propagation of Lamb waves in a thin duralumin plate with randomly distributed strong scatterers. Each scatterer is created by rigidly clamping the plate with a screw to a vertical postlocated under the plate. The evolution of the medium properties is created by the time variation of the plate temperature. The basic effect of the temperature variation is the change of both longitudinal and shear sound speeds in the plate due to the change in the elastic moduli of the plate material. Two types of the change in the medium are considered: the *global* change with slow variation in the overall temperature of the plate and the *local* change introduced by heating the plate with a point-like heat source. In both cases the change in the medium is quantified by measuring magnitude and time-shift of the cross-correlations between successive signals as a function of temperature. The two changes are expected to have different effects on the wave propagation in the plate and were found to manifest themselves differently. Our preliminary results indicate the possibility of future applications in seismology by allowing characterization of the changes in the wave speed propagation in the earth crust.

9:45

3aEA7. Bayesian parameter estimation of porous materials. Eric A. Dieckman (Dept. of Appl. Sci., The College of William and Mary, P.O. Box 8795, Williamsburg, VA 23187, eadieckman@wm.edu) and Ning Xiang (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

This work proposes a new application of Bayesian analysis to determine the physical parameters of a rigid frame porous material from the measurement of the material's acoustic impedance. Bayesian analysis allows for numerical estimation of physical parameters, their uncertainties, and interrelationship of the parameters from the measured complex impedance. Complex impedance data are simulated based on measurements of a sample material to test the Bayesian formulation. Using synthesized data allows for investigations of different frequency resolutions and different levels of noise on the algorithm. Results obtained from testing the parameter estimation algorithm with synthesized data will be utilized in applying the algorithm to measured complex impedance data.

WEDNESDAY MORNING, 28 OCTOBER 2009 LIVE OAK, 8:30 TO 11:40 A.M.

Session 3aMU

Musical Acoustics: Psychoacoustic Response to Musical Instruments

Rolf Bader, Chair
Univ. of Hamburg, Inst. of Musicology, 20354, Hamburg, Germany

Invited Papers

8:30

3aMU1. Recent research in musical timbre perception. James W. Beauchamp (Dept. of Elec. & Comput. Eng. and School of Music, Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801 jwbeauch@illinois.edu), Andrew B. Horner (Hong Kong Univ. of Sci. & Technol., Clear Water Bay, Kowloon, Hong Kong, horner@cse.ust.hk), and Michael D. Hall (Dept. of Psych., James Madison Univ., Harrisonburg, VA 22807, hallmd@jmu.edu)

Various methods have been used in an attempt to reveal salient parameters for musical timbre perception and to estimate perceptual distances between timbres based on their spectra. The multidimensional scaling method compares exemplars of musical sounds using listener estimation of timbral dissimilarity between sound pairs. Data are coalesced into a space whose dimensions are easily interpretable in terms of acoustic parameters such as spectral centroid, spectral flux, spectral irregularity, and attack time. Sounds can be normalized with respect to some parameters in order to reveal others. The direct comparison method attempts to discover how discrimination, recognition, and degree of perceptual dissimilarity depend on the degree of modification of the time-varying spectrum or of particular features of the spectrum. Morphing between spectrum pairs is one method of producing spectral modifications. Dissimilarity can be predicted by computing a "distance" between two time-varying spectra. For musical sounds, spectral distance has been computed by taking the average Cartesian distance between spectral measures treated as vectors. For constant-F0 sounds, the most obvious vector is the set of harmonic amplitudes, either in linear or decibel units. Other vectors are based on critical bands and cepstral coefficients. [Work supported by Research Grants Council Grant No. 613508.]

8:55

3aMU2. Perception strategies for violin sounds. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstr. 13, 20354 Hamburg, Germany, r_bader@t-online.de)

Subjects use different perception strategies when listening to violin tones coming from realistic and from unrealistic violin geometries. The perception of realistic-geometry violin top plate tones is based on pattern recognition using multiple sound features. In a multidimensional scaling method, experiment subjects were able to linearly sort sounds according to linear geometric changes of the violin top plate only as long as the geometry was within realistic limits. Here learned sound features of brightness of different spectral bands were detected and clearly associated with features of the violin geometry. When unrealistic-geometry sounds were presented, the perception strategy changed and only relations between single spectral bands and perceived sound similarities could be observed. In a second experiment, the back plate was investigated likewise where no pattern recognition could be found, because the back plate is radiating much less and so the overall violin timbre known to subjects is the top rather than the back plate one. The sounds were produced by a whole body computer physical model simulation, where the violin top and back plates were changed in thickness within the range 1–3 mm in steps of 0.25 mm playing a virtual open e-string.

9:20

3aMU3. Perception of timbre dimensions: Psychophysics and electrophysiology in humans. Anne Caclin, Marie-Hélène Giard (INSERM U821, 69 675 Bron Cedex, France, anne.caclin@inserm.fr), and Stephen McAdams (McGill Univ., Montreal, PQ, H3A 1E3, Canada)

Timbre is a perceptual attribute that characterizes the identity of a sound source. Psychoacoustical experiments have revealed that this perceptual attribute is multidimensional, with timbre dimensions (such as sharpness or brightness) corresponding to a variety of acoustical parameters: attack time, spectral centroid, spectrum fine structure, spectral flux, etc. The question arises whether timbral dimensions are processed holistically or separately in the auditory system. This issue was addressed in a series of behavioral and event-related potential (ERP) studies in humans [Caclin *et al.*, J. Acoust. Soc. Am. (2005); J. Cogn Neurosci. (2006); J. Cogn Neurosci. (2008); Brain Res. (2007)]. Different timbres were created by manipulating attack time, spectral centroid, and spectrum fine structure. Dissimilarity ratings confirmed that these three parameters were indeed major determinants of timbre. Their representation in auditory sensory memory was explored using the mismatch negativity (MMN) component of the auditory ERP. MMN results indicated mainly separate sensory memory representations of these dimensions in auditory cortex. Using Garner interference paradigm, evidence was found for an interactive processing of timbre dimensions at other levels of analysis than sensory memory, including early perceptual processing stages. These results can be synthesized in a model of timbre perception postulating separate channels of processing for different timbre dimensions, with cross-talk between those channels.

9:45

3aMU4. Development of auditory phase-locked activity for music sounds. Antoine Shahin (UC Davis Ctr. for Mind and Brain, Davis, CA 95618, ajshahin@ucdavis.edu)

The auditory cortex undergoes anatomical and functional development that reflects specialization for learned sounds. Auditory maturation is evident in the transient auditory evoked potentials (AEPs) and auditory phase-locked oscillatory activity. Development of AEPs and the phase-locking strength of oscillatory auditory responses to piano, violin, and pure tones were examined. The hypothesis was that if oscillatory activity in different frequency bands reflects different aspects of sound processing, then the development of phase-locking at these frequencies will have different maturational trajectories. Phase-locking for theta (4–8 Hz), alpha (8–14 Hz), lower-to-mid-beta (14–25 Hz), and upper beta and gamma (25–70 Hz) bands strengthened with age and was stronger for musical than pure tones. The phase-locking increase for gamma and upper beta bands mainly reflected the maturation of the spectral representations for sounds. In contrast, increase in phase-locking for theta, alpha, and lower-to-mid-beta was mainly attributed to sensitivity to sound temporal onset rise time. Frequency-specific phase-locking provides a tool to assess auditory development for spectral and temporal aspects of naturalistic complex sounds. [This research was supported by grants from the Canadian Institutes of Health Research and the National Institutes of Health, National Institute on Deafness and other Communication Disorders.]

10:10—10:25 Break

10:25

3aMU5. Psychoacoustics and organology. John M. Hajda (Dept. of Music, Univ. of CA., Santa Barbara, 1315 Music Bldg., Santa Barbara, CA 93106-6070, jhajda@music.ucsb.edu)

Organology, a branch of ethnomusicology, is the study of musical instruments and their classification. The first Western system–which classified instruments primarily as idiophones, membranophones, chordophones, or aerophones—was proposed by Sachs and von Hornbostel in 1914. This system has been the progenitor and inspiration for numerous other classification systems. While the van Hornbostel–Sachs system has proven to be useful in many contexts, it does not offer a consistent approach in terms of the transfer of energy from driver to generator to resonator. In addition, it does not take into account the common practice in modern popular and art music by which digitized samples of sounds are disembodied and appropriated from musical instruments from around the world. In other words, rather than encountering musical instruments as visual, tactile, and sound-producing, musicians and audiences encounter only pure sound. This paper, therefore, addresses two issues: How would a classification system of musical instrument {\sound} take into account the vibrational characteristics of acoustical musical instruments as well as the results of perceptual research on musical timbre from the past 40 years? The resulting system effectively turns Hornbostel–Sachs "on its ear" by reinterpreting primary classes as subclasses and vice versa.

10:50

3aMU6. The role of formant positions and micro-modulations in blending and partial masking of musical instruments. Christoph Reuter (Musicological Inst., Univ. of Vienna, Spitalgasse 2-4, Campus Hof 9, A-1090 Vienna, Austria, christoph.reuter@univie.ac.at)

In experiments about the distinction of instruments playing in unison, it has been shown that timbres with overlapping formant areas are blending homogeneously whereas timbres with different formant areas are distinguishable very well (Reuter, 1996). On the other hand timbre separation is enhanced by micro-modulations and vibratos [e.g., McAdams (1982); Reuter (2007)]. The aim of this contribution is to find out to what extent timbre separation is influenced by different micro-modulations/vibratos versus different formant positions. 30 subjects listened to scale fragments of simultaneous playing instruments with equivalent formant areas and with different formant areas, with and without starting/ending transients, and with and without vibratos/micro-modulations. They had to answer the following question: Do you hear one sole instrument (or two of the same instruments) or do you hear two different instruments? In almost all cases featuring equivalent formant areas most of the subjects perceived one blending timbre, independent of the presence of vibratos/micro-modulations and transients. In the presence of vibratos/micro-modulations, differing formant areas predominantly led to the perception of two different timbres. Without vibratos/micro-modulations, timbre separation is much worse. These observations are explainable by a psycho-acoustical model of timbre perception.

11:15

3aMU7. Observations of the singing voice during pregnancy. A case study. Filipa Lã (Dept. of Commun. and Arts INET, Campus Universitario de Santiago, 3810-193, Portugal) and Johan Sundberg (KTH, SE-10044 Stockholm, Sweden)

Previous studies showed a strong correlation between cervical and vocal fold smears [Abitbol *et al.*, J. Voice **13**, 424–446 (1999)]. Myers *et al.* [Eur. J. Obstet. Gynecol. Reprod. Biol. **144S**, S82–S89 (2009)] found significant changes in mechanical properties of tissues during pregnancy. This suggests that similar changes may occur also in laryngeal tissues, affecting conditions for vocal fold vibration. Recordings of audio, electrolaryngograph, oral pressure, and air flow signals were made of a professional classically trained soprano once per week, starting on week 28 of pregnancy. The tasks comprised reading a text and singing a song. In addition, the singer repeated the syllable [pae] while performing a diminuendo at various pitches, allowing determination of the lowest pressures producing vocal fold vibration and contact, i.e., the phonation and contact threshold pressures [Enflo and Sundberg (unpublished)]. Oral pressure during the occlusion for the consonant [p] was accepted as an estimate of subglottal pressure. Voice source was analyzed by means of inverse filtering and the relation between this pressure and various voice source parameters was analyzed. Preliminary results of this exploratory study show substantial effects of pregnancy on voice production.

WEDNESDAY MORNING, 28 OCTOBER 2009 REGENCY EAST 2, 8:00 TO 11:55 A.M.

Session 3aPA

Physical Acoustics, Biomedical Ultrasound/Bioresponse to Vibration, and Engineering Acoustics: Light and Sound in Science, Engineering, and Medicine

Joel Mobley, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Ronald A. Roy, Cochair
Boston Univ., Dept. of Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Chair's Introduction—8:00

Invited Papers

8:05

3aPA1. Vibrational modes of nanoparticles studied by ultrafast optical spectroscopy. Gregory V. Hartland (Dept. Chem. and Biochem., Univ. of Notre Dame, Notre Dame, IN 46556)

Excitation of metal nanoparticles with an ultrafast laser pulse causes rapid heating. This can excite the vibrational modes of the particle that correlate with the expansion coordinate. Comparison of the experimentally measured periods to continuum mechanics calculations (usually done through finite element modeling) allows us to assign the observed vibrational modes. In most cases the breathing mode is excited, and its frequency depends on the size, shape, and elastic constants of the particles. For typical particle sizes of tens of nanometers, the frequencies are in the range of 10–100 GHz. A range of different shapes has been studied: spheres, cubes, rods, and even hollow particles. The measured and calculated periods are usually in good agreement, which implies that the elastic constants of small metal particles are similar to the bulk material. Experiments on single metal particles have also been performed. These measurements allow us to determine the damping times of the acoustic vibrational modes, which cannot be done in conventional ensemble experiments because of the different sizes and shapes that are invariably present in the samples. These measurements allow us to study energy relaxation and propagation in nanomaterials, and how nanoparticles interact mechanically with their environment.

8:25

3aPA2. Superheterodyne techniques in laser ultrasonics. Todd W. Murray, Suraj Bramhavar, Ashwin Sampathkumar, Kamil L. Ekinci (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, twmurray@bu.edu), and Bruno Pouet (Bossa Nova Technologies, LLC, Venice, CA 90291)

Laser-based ultrasonics is a powerful tool for the mechanical characterization of thin films and nanostructures. In this paper, superheterodyne approaches to single-point and full-field detection of laser generated ultrasonic signals will be discussed. In these techniques, an amplitude modulated laser source is used to excite a narrow-bandwidth signal. The detection laser, incorporated into an interferometric detection scheme, is also amplitude or phase modulated at frequency that is offset from the generation laser modulation frequency by a fixed amount, serving as the local oscillator for superheterodyne detection. Introduction of the local oscillator allows for the optical down-conversion of the high-frequency intensity modulation associated with sample motion to a low and fixed intermediate frequency given by the difference between excitation and detection laser modulation frequencies. The primary benefit of using this approach is that the upper frequency bound is not dictated by the speed of the detection photodiode and electronic circuitry or, in the case of full-field detection, the frame rate of the CCD. Results are presented demonstrating single-point superheterodyne detection of gigahertz frequency bulk and surface acoustic waves using a low-frequency photodiode and full-field superheterodyne detection of megahertz vibrations of nanostructure arrays using a low frame rate camera.

8:45

3aPA3. Effects of speckles in laser Doppler vibrometry. Vyacheslav Aranchuk and James M. Sabatier (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677, aranchuk@olemiss.edu)

Laser Doppler vibrometry has now started to become a common technique for vibration measurements, offering significant advantages due to its non-contact nature over traditional vibration sensors. When a laser Doppler vibrometer (LDV) is used for measurements of an optically rough surface, the effect of laser speckles can influence the LDV performance and its practical applications. The Doppler signal carrying the information of the vibration of the target results from the coherent addition of the light scattered from the target, which is the speckle field, with the reference beam. As a result, the Doppler signal and the demodulated velocity signal are effected by the speckle field. Changes in the speckle field caused by target motion or a scanning laser beam can generate noise on the LDV output, referred to as speckle noise. The speckle noise is caused by the combined effect of spikes in the demodulated velocity signal and the phase fluctuations of speckles. The origins of LDV speckle noise, its effect on practical vibration measurements, and methods of speckle noise reduction are presented.

9:05

3aPA4. Intravascular ultrasound and photoacoustic imaging. Stanislav Emelianov (Dept. of Biomedical Eng., Univ. of Texas at Austin, Austin, TX 78712)

Catheter-based intravascular ultrasound (IVUS) is one of the imaging tools for the clinical evaluation of atherosclerosis. However, histopathological information obtained with IVUS imaging is limited. We developed a combined intravascular photoacoustic (IVPA) and IVUS imaging to characterize atherosclerotic plaques. The amplitude, spectral, and temporal characteristics of the photoacoustic signal are primarily determined by optical absorption properties of different types of tissues and can be used to differentiate the lipid, blood, fibrous, and fibro-cellular components of an inflammatory lesion. Furthermore, using bioconjugated contrast agents such as metal nanoconstructs, molecular IVPA imaging is possible since different molecules are overexpressed during various stages of atherosclerosis. Imaging experiments were performed using clinical IVUS imaging catheters interfaced with a pulsed laser system. The performance of the IVPA/IVUS imaging was assessed using vessel-mimicking phantoms. To detect the lipids in the plaque and to assess plaque composition, multi-wavelength (680–950 nm) spectroscopic IVPA imaging of a normal and an atherosclerotic rabbit aorta was performed. Molecular and cellular IVPA imaging was demonstrated using intravenously injected plasmonic nanoparticles. Finally, IVPA imaging was used to visualize coronary stents within the vessel wall. Our studies suggest that plaque detection, characterization, and even treatment can be improved using the combined IVPA.

9:25

3aPA5. Optical detection of motion generated in soft tissue by a focused ultrasound beam. Emmanuel Bossy (Institut Langevin, ESPCI ParisTech, CNRS UMR 7587, Inserm ERL U979, 10 rue Vauquelin, 75 231 Paris Cedex 05, France)

This presentation will discuss various approaches combining light and sound for tissue characterization. All the approaches discussed here involve the optical detection of motion generated at depth in tissue by use of a focused ultrasound transducer. In the first part of the presentation, the acousto-optic imaging approach will be described, in which the optical phase modulation created by an ultrasound beam is used to derive optical properties with the resolution of the ultrasound beam. In particular, it will be demonstrated that randomly phase-modulated millisecond ultrasound burst can be used to obtain both transverse and axial resolutions. In this first approach, the optically detected motion consists of the ultrasonic motion of insonified light scatterers. In the second part of the presentation, it will be demonstrated that it is also possible to optically detect the transient shear motion generated by the acoustic radiation force in soft tissue. As for acousto-optic imaging, this approach is based on the interaction between a diffuse coherent light field and a localized motion and therefore provides localized information in the sample. In addition to provide information on optical properties, this approach also provides information on shear stiffness on the medium. For both approaches, experimental results obtained *in vitro* on tissue or tissue-mimicking phantoms will be shown, and the current challenges toward *in vivo* measurements will be discussed.

9:45

3aPA6. *In vivo* photoacoustic tomography and its clinical application. Chulhong Kim and Lihong, V Wang (Optical Imaging Lab., Dept. of Biomedical Eng., Washington Univ. in St. Louis, Campus Box 1097, One Brookings Dr., St. Louis, MO 63130, chkim@biomed.wustl.edu)

High-resolution volumetric optical imaging modalities, such as confocal microscopy, two-photon microscopy, and optical coherence tomography, have become increasingly important in biomedicine. However, due to strong light scattering, the penetration depths of these imaging modalities are limited to the optical transport mean free path (1 mm) in biological tissues. Photoacoustic imaging, an emerging hybrid modality that can provide strong endogenous and exogenous optical absorption contrasts with high ultrasonic spatial resolution, has overcome the fundamental depth limitation while maintaining high spatial resolution. The image resolution, as well as the maximum imaging depth, is scalable with ultrasonic frequency within the reach of diffuse photons. In biological tissues, the imaging depth can be up to a few centimeters deep. In this presentation, the following topics of photoacoustic imaging will be discussed; (1) multi-scale photoacoustic imaging systems, (2) non-invasive sentinel lymph node mapping using photoacoustic imaging and its clinical application, and (3) functional and molecular photoacoustic imaging.

10:05—10:25 Break

Contributed Papers

10:25

3aPA7. Nanoparticle-targeted photoacoustic cavitation for deep tissue optical imaging. James R. McLaughlan, Ronald A. Roy, and Todd W. Murray (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, jmcl@bu.edu)

Photoacoustic tomography is a non-invasive imaging technique based on broadband acoustic emissions from light absorption in tissue. Light-absorbing gold nanoparticles can be introduced and targeted to specific cell populations, thereby promoting both contrast and the ability to delineate tissue types. For sufficiently high laser fluence, a transient vapor cavity is formed and undergoes inertial collapse, generating enhanced emission and additional contrast. However, the fluence required to achieve this effect usually exceeds the maximum permissible exposure for tissue. By combining ultrasonic and optical pulses, the light and sound thresholds required to repeatedly generate inertial cavitation were reduced to 5 mJ/cm^2 and 1 MPa, respectively. Experiments employed a transparent arylimide gel possessing a small ($<$600 μm) region doped with 80nm diameter gold nanoparticles and simultaneously exposed to pulsed laser light (532 nm) and pulsed ultrasound (1.1 MHz). The amplitude of broadband emissions induced by both light and sound exceeded that produced by light alone by almost two orders of magnitude, thereby facilitating imaging a deeper depth within tissue. Two-dimensional images of doped regions generated from conventional photoacoustic and ultrasound-enhanced emissions are presented and compared. Implications for imaging and HIFU therapy are discussed. [Work is supported by a Boston University Dean's Catalyst Award.]

10:40

3aPA8. Quantitative sensing of optical properties of diffusive media by pressure contrast acousto-optic imaging. Puxiang Lai, Ronald Roy, and Todd Murray (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Acousto-optic imaging is a dual-wave modality that combines ultrasound with diffuse light to achieve deep-tissue imaging of optical contrast with the spatial resolution of ultrasound. While substantial progress has been made in recent years in detecting and imaging optical inhomogeneities in highly scattering media based on the acousto-optic response, quantitative measurement of subsurface optical properties remains difficult. A technique for the local and quantitative measure of the optical properties of turbid media, referred to as pressure contrast acousto-optic imaging, is presented. In this approach, the acousto-optic signals elicited by ultrasound pulses at two different peak pressures are measured, and the ratio of the amplitudes of these two signals determined. The resulting pressure ratio, once calibrated for a given set of pressure pulses, is found to give a direct measure of the optical transport mean free path of the interaction region between the light and sound. Measurements of the transport mean free path of inclusions buried in diffuse tissue phantoms, with no *a priori* knowledge of the optical properties of the phantom in which the inclusion is embedded or sample dimensions, are demonstrated. [Work supported by the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

10:55

3aPA9. The use of the photoacoustic effect for non-invasive temperature monitoring during high intensity focused ultrasound exposures. James R. McLaughlan (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, jmcl@bu.edu), Parag V. Chitnis, Jonathan Mamou (Riverside Res. Inst., New York, NY 10038), Todd W. Murray, and Ronald A. Roy (Boston Univ., Boston, MA 02215)

High intensity focused ultrasound (HIFU) is a non-invasive technique for treating soft tissue tumors. A feasibility study for using the photoacoustic (PA) effect to monitor *in-situ* temperature changes during HIFU exposures is presented. A PA wave is generated from the thermoelastic expansion of a light-absorbing medium with its amplitude related to the Grüneisen parameter of the medium, which, for water, depends linearly on temperature. A 2-MHz HIFU transducer heated a cylindrical graphite inclusion in which a wire thermocouple was embedded. A 532-nm pulsed laser illuminated the inclusion at a rate of 10 Hz, and a 15-MHz passive transducer monitored the PA response throughout the 30-s exposure. Singular-value-decomposition analysis of the PA response was performed to extract the contribution from temperature change. Thermocouple measurements indicated a temperature increase from 22–60 °C, depending on the HIFU intensity employed, with a concomitant 20%–30% increase in PA amplitude. PA-based temperature curves correlated with the thermocouple measurements (rms deviation $<$5 °C). This technique tracked both the heating and cooling phases of the HIFU exposure and potentially could be used to noninvasively monitor *in-situ* temperature in real-time during HIFU. [Work supported by a Boston University Dean's Catalyst Award and the RRI's Biomedical Engineering Research Fund.]

11:10

3aPA10. Acousto-optic detection of high-intensity focused ultrasound lesions in real time. Andrew B. Draudt, Puxiang Lai, Todd W. Murray, Robin O. Cleveland, and Ronald A. Roy (Dept. of Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Tissue ablation by high-intensity focused ultrasound (HIFU) leads to changes in the optical absorption and scattering of the tissue. Here acousto-optic (AO) sensing was used to detect HIFU induced changes in the optical contrast of chicken breast. The tissue was insonified by 1-MHz HIFU and illuminated by infrared light. The diffuse light that exited the tissue was collected and processed using photo-refractive crystal based interferometer configured to detect phase modulated light that had interacted with the HIFU. Here the HIFU served both to ablate the tissue and elicit the AO response, and as a lesion forms, the increase in optical absorption and scattering should result in a reduction in modulated light being detected by the interferometer. The HIFU was amplitude modulated with a 50% duty cycle at 50 Hz and the interferometer output was processed through a lock-in amplifier tuned to the modulation frequency to improve the signal-to-noise ratio. Using a focal pressure of 6 MPa (derated from water), the AO signal remained constant for ~10 s of HIFU, and a "*post-mortem*" in this timeframe found no detectable lesion. Once the AO signal decreased a lesion

was always detected and the lesion size correlated with the amount of decay of the AO signal. [Work supported by NSF CenSSIS EEC-9986821]

11:25

3aPA11. Modeling the response of a laser Doppler vibrometer to photoacoustic signal from solids. Logan Marcus, Richard Raspet, James Sabatier, and Vyacheslav Aranchuk (NCPA, Univ. of Mississippi, 1 Coliseum Dr., University, MS 38677)

Different techniques for photoacoustic signal detection have been developed since its inception as a spectroscopic method. Detection tools such as microphones, piezoelectric sensors, laser beam deflection systems, and more have been used to study the response of different materials to photoacoustic excitation. The authors have proposed a method of detection of the photoacoustic signal from solids using a laser Doppler vibrometer (LDV). An analytical model of the phenomena underlying the change in phase of a LDV probe beam resulting from the photoacoustic effect from a condensed sample has been constructed. The model takes into account two phenomena responsible for changing phase of the LDV probe beam. The two phenomena are the surface vibrations caused by the periodic heating of the solid sample, and the fluctuations of index of refraction in air caused by the temperature variations and acoustic waves. Experimental verification of the model is discussed.

11:40

3aPA12. Plasma conditions during single bubble sonoluminescence. Kenneth S. Suslick and David J. Flannigan (Dept. of Chemistry, California Inst. of Technol., Pasadena, CA 91125, flanniga@caltech.edu)

There is a remarkable lack of experimental data on the conditions created during cavitation bubble collapse. Indeed, only recently has strong evidence of plasma formation been obtained during single bubble cavitation. Here we have determined for the first time the plasma electron density and the ion broadening parameter during single-bubble sonoluminescence and examined them as a function of acoustic driving pressure. We find that the electron density spans four orders of magnitude and can exceed 10×10^{21}/cc (which is comparable to the densities produced by intense laser-induced inertial confinement fusion experiments, e.g., the NOVA ICF laser at Livermore) with effective plasma temperatures ranging from 7000 to more than 16 000 K. At the highest acoustic driving force, neutral Ar lines can no longer be used as spectroscopic reporters due to the extent of ionization and to leveling of the population of states. Accounting for the temporal profile of the sonoluminescence pulse suggests that the ultimate conditions generated inside the collapsing bubble may far exceed those determined from emission from the outer transparent region of the plasma core.

WEDNESDAY MORNING, 28 OCTOBER 2009 BOWIE (LOSOYA CENTER), 8:00 A.M. TO 12:00 NOON

Session 3aPP

Psychological and Physiological Acoustics: Physical Aspects and Models (Poster Session)

Chang Liu, Chair

Univ. of Texas at Austin, Dept. of Communication Sciences and Disorders, 1 University Station, Austin, TX 78712

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of event-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

3aPP1. Mechanical model of bone conduction. David Chhan and Charles Thompson (Dept. of Elec. and Comput. Eng., Univ. of Massachusetts Lowell, 1 Univerisity Ave., Lowell, MA 01854, david_chhan@student.uml.edu)

A mechanical model of the human auditory system will be presented. Of particular interest is bone conduction as a mean for energy transfer. A model is developed based on the lumped parametric model of the human ear proposed by Feng and Gang. Stapes displacement is examined for both air conduction and bone conduction (BC) stimulations. Results indicate that with the adjustment of the skull parameters, the model produced similar responses observed by various studies and experiments on BC.

3aPP2. Infrared position tracking. Megha Sunny (Elec. Eng., Univ. of Massachusetts Lowell, 1 University Ave., Lowell, MA 01854 US)

In this work we examine methods for position determination using infrared-sensors. We are using a Wii remote and infrared LED glasses to find the distance between the head and the Wii remote sensors. The relative distance is calculated by using triangulation method. The spatial orientation of the head with respect to the Wii remote sensor was also measured using the infrared LED and the infrared sensors of Wii remote. The spatial data obtained will control the acoustic characteristics of a head related transfer function that will be used in source localization in virtual environments.

3aPP3. A method for analyzing the biodiversity in bat pinnae. Rolf Müller (Dept. of Mech. Eng., Virginia Tech. & Inst. for Adv. Learning and Res., 150 Slayton Ave., Danville, VA 24540, rolf.mueller@vt.edu) and Jianguo Ma (Shandong Univ., 250100 Jinan, China)

The pinnae of bats show conspicuous interspecific variability in global as well as in local shape features. Evidence from case studies suggests that at least some of these features may serve acoustic functions. Consequently, the variability in pinna shapes may represent evolutionary adaptation patterns to a range of different acoustic sensing tasks. An automated nonparametric method has been developed to extract these patterns from the biodiversity in pinna shapes. It allows the computation of three-dimensional principal components, *eigenears*, from high-resolution digital representations of the entire pinna surface. To achieve this, the inner and outer pinna surfaces that diffract incoming sound and thereby define the pinna's acoustic properties have been subjected to a cylindrical transformation. The radius values of discrete points on the pinna's surfaces have been taken to define members of a continuous valued vector space suitable for a principal component analysis. The results of the analysis show that the biodiversity in pinna shapes can be approximated well by a reduced number of dimensions that could be used to capture design trends. The natural global-to-local ordering of the eigenears can be used as a recipe to describe the sequential "design" of individual bat ears from a basic shape.

3aPP4. Echolocation of objects by using the frequency modulated sound. Ikuo Matsuo (Dept. of Information Sci., Tohoku Gakuin Univ., 2-1-1 Tenjinzawa, Sendai 981-3193, Japan and Bio-Navigation Ctr., Doshisha Univ., 1-3 Miyakodani, Kyotanabe, Kyoto 610-0321, Japan, matsuo@cs.tohoku-gakuin.ac.jp)

Using the echolocation, bats can capture moving objects in real three dimensional space. Bats emit the frequency modulation sound and can identify

objects with an accuracy of less than a millimeter. To determine delay times of multiple objects requires estimating the sequence of delay separations by extracting temporal changes in the interference pattern of the echoes. The models have been previously proposed to determine delay times of multiple objects by using the frequency modulation sound. In order to extract the temporal changes, Gaussian chirplets with a carrier frequency compatible with emission sweep rates were used, and each delay time was estimated by using the temporal changes. It was examined that this model could estimate each delay time within the accuracy less than 1 mm in the case of measuring echoes from multiple objects with the same size. In the real environment, it is necessary to estimate the multiple objects with the same and different size. In this paper, the echoes were measured from multiple objects, and it was examined that this model could estimate each delay time of these objects.

3aPP5. Possible role of steady streaming flows in fish hearing. Charlotte W. Kotas, Peter H. Rogers, and Minami Yoda (Dept. of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, charlotte.kotas@gatech.edu)

Over the last five years, the authors have studied the time-independent portion, or the steady streaming, of flows near model otoliths and determined that these flows encode both the direction and frequency of incident sounds. This talk summarizes these findings and discusses whether the fish ear can detect the displacements due to these flows. Although the results for a model cod otolith indicate that information relating to the direction of the local sound field is indeed contained within the flow patterns in the vicinity of the sensory macula, the threshold for the detectability of these directional cues appears to be much higher than published thresholds for directionalization in fish. There is some evidence, however, that the observed thresholds may have been underestimated. While these studies indicate that steady streaming flows may be insufficient to explain directionalization in fish, the results suggest that steady streaming may provide a mechanism for fish to hear ultrasound. [Work supported by ONR.]

3aPP6. Visualizing unsteady fluid flows inside the fish ear. Charlotte W. Kotas, Peter H. Rogers, and Minami Yoda (Dept. of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332, charlotte.kotas@gatech.edu)

Fish sense their acoustic environment using ears that are typically able to detect sounds of $O(10–10^3$ Hz) at 80–130 dB *re* 1 μPa (for the Atlantic cod *Gadus morhua*, for example). Each ear contains three stone-like masses (the "otoliths") embedded in a fluid filled membrane. The sulcus, a groove on one side of the otolith, is lined with $O(10^4)$ sensory hair cells which presumably hear by detecting the acoustically induced motion of the fluid and tissue relative to the otolith. The "auditory retina hypothesis" suggests that the fish ear uses the fluid flow patterns, sensed by the array of hair cells, to characterize incident sounds. Although the time-independent steady streaming flow patterns have been shown to encode the frequency and direction of the oscillation (and hence incident sounds), the hair cell cilia displacements due to such flows are probably too small to be sensible. This experimental study focuses instead on the unsteady flows near a 350% model cod otolith oscillating in an otherwise still fluid. Results are presented for an otolith oscillating at different orientations (corresponding to the excitation from sounds from different directions) at frequencies of 3–25 Hz and amplitudes of 0.1–0.5 mm. [Work supported by ONR.]

3aPP7. Comparison of measured head-related transfer functions using spatio-temporal frequency analysis. Yasuko Morimoto (Dept. of Media Sci., Grad. School of Information Sci., Nagoya Univ., Furo-cho, Chikusa-ku Nagoya, Aichi 464-8603, Japan, y-morimoto@sp.m.is.nagoya-ac.jp), Takanori Nishino (Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan), and Kazuya Takeda (Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8603, Japan)

We make a comparison of measured head-related transfer functions (HRTFs) in spatio-temporal frequency domain. An HRTF is an acoustic transfer function between a sound source and the ear canal entrance, and defined as a function on time and space. The spatio-temporal frequency characteristics can be visualized and analyzed by showing the spectrum computed by multidimensional Fourier transform on time and space. In our experiments, we investigated the basic property of the spatio-temporal frequency characteristic and the difference between all data for the HRTFs obtained by numerical analysis and actual measurements. The reverberation and HRTF individuality were also examined. From the results, the characteristics were mostly concentrated in a specific band frequency, and there were also differences among databases.

3aPP8. The effect of context priming on the auditory potential evoked by semantic analysis (N400). Mary Kathryn Reagor, James Jerger, Tara Davis, and Jeffrey Martin (School of Behavioral and Brain Sci., Univ. of Texas at Dallas, Dallas, TX 75080)

Context is an important aspect of the perception of ongoing speech. Understanding a speech message in the presence of background noise is aided by knowledge of the context of the message even though the audibility of specific phonemes may be compromised. In this experiment, we used a context priming paradigm to investigate the importance of context to word recognition. We studied the auditory ERP, specifically the N400 response, to presented words which had been previously heard in the context of a story. Results showed that words preceded by a word from the target story elicited a smaller N400 negativity than words preceded by a word from either a different story or from neither story. Thus, words that came from the same story (i.e., had congruent context) were processed with less effort than words from a different story. This "context effect" is discussed in relation to the perception of ongoing speech under difficult listening conditions.

3aPP9. Continuous reconstruction of head related transfer function using optimal-degree spherical harmonics expansion. Dan Rao (Acoust. Lab., Phys. Dept., School of Sci., South China Univ. of Technol., Guangzhou 510641, China, phdrao@scut.edu.cn)

Head related transfer function (HRTF) is a continuous function of spatial direction. In practice, only HRTFs at discrete directions can be measured. In order to obtain HRTFs at unmeasured directions, a continuous reconstruction (or interpolation) procedure is needed. In this paper, a method for continuous reconstruction of HRTF based on optimal-degree spherical harmonics expansion is developed. Expansion coefficients were derived by using the pseudo-inverse method. To address the ill-posed problem in inverse calculation, the expansion degree for each frequency bin was optimized by a heuristic method. This method searches the minimum-norm errors of the reconstructed HRTFs with respect to measured ones which were not used to derive the coefficients. The KEMAR HRTF at 710 directions and a human HRTF at 493 directions were employed in reconstruction. Measured HRTFs from KEMAR or human subject were dived into three groups according to directions, in which the first group of 2/5, the second group of 2/5, and the third group of 1/5 data were used to calculate the expansion coefficients, determine the optimal expansion degree, and evaluate the reconstruction performance, respectively. Compared with the expansion methods using a constant degree for all frequency, the reconstruction performance of the current method is prior.

3aPP10. Errors in the measurements of individual head-related transfer function. Xiao-li Zhong and Bo-sun Xie (Acoust. Lab., Phys. Dept., School of Sci., South China Univ. of Technol., Guangzhou 510641, China, xlzhong@scut.edu.cn)

Head-related transfer function (HRTF) describes the transmission process from a point sound source to ears in free field and thus includes the effect of sound diffraction imposed by human anatomical structures. Because each person owns a unique configuration of anatomical structures, HRTF is individual-dependent. Individual HRTF plays an important role in guaranteeing authenticity in virtual auditory display. Measurement is a common way to obtain individual HRTF. In practice, measurement error is unavoidable. Moreover, measurement error and individual characteristic are often blended together, especially at high frequencies where both the measurement error and the individual characteristic are obvious. The object of this work is to evaluate the errors in the measurements of individual HRTF, and extract individual characteristic of HRTF from measured results. HRTFs from six human subjects and KEMAR were measured at 24 spatial positions with six repetitions. Significance of the individual characteristic was examined using analysis of variance. In addition, psychoacoustic experiments were conducted to investigate the audibility of the measurement error and

individual characteristic, respectively. Results show that the individual characteristic of HRTF dominates the measured results. [Work supported by Nature Science Fund of Guangdong province (07300617).]

3aPP11. Spatial principal components analysis on head-related transfer functions and individualized customization from fewer measurements. Bo-sun Xie (Acoust. Lab., Phys. Dept., School of Sci., South China Univ. of Technol., Guangzhou 510641, China, phbsxie@scut.edu.cn)

Head-related transfer function (HRTF) is an individual-dependent and continuous function of sound source position and frequency. Measurement is a common way to obtain individual HRTFs. It is difficult or even impractical to measure HRTFs with high directional resolution for individual, however. To solve the problem, spatial principal components analysis (SPCA) of HRTFs and resulting customization of individual HRTF from fewer measurements were proposed in present paper. In SPCA, HRTF is decomposed as a weighted combination of common spatial basis functions (CSBFs) which are only direction-dependent, while the weights are frequency- and individual-independent. The CSBFs were derived from the measured HRTFs of 20 subjects with high directional resolution. Based on the calculated CSBFs, for any new subject, the weight coefficients can be estimated from fewer measured HRTFs data, and subsequently the individual HRTFs with high directional resolution can be customized. SPCA were applied to HRTF magnitudes and head-related impulse responses, respectively, and the individual data were customized for six subjects. Psychoacoustic experiments were also conducted to evaluate the customization performance. Theoretical and experimental results validate the performance of proposed method. [Work supported by of National Nature Science Fund of China Grant No. 10774049.]

3aPP12. Effect of the fundamental frequency and vocal register on the voice pitch compensation. Shuntaro Okazaki, Koichi Mori, and Chang Cai (Res. Inst. of NRCD, 4-1 Namiki, Tokorozawa, Saitama 359-8555 Japan, okazaki@rehab.go.jp)

Fundamental frequency (F0) of our vocalization is regulated by auditory feedback. Manipulation of F0 in the auditory feedback invokes compensatory F0 response in vocalization, whose amplitude is known to depend on the task (singing or speaking; instruction to follow/ignore the heard pitch change) and the averaged F0. In order to exclude the voluntary factor and to test if the vocal register affected the response amplitude, we used a parametric modulation of F0 according to the maximal sequence [TAF, Kawahara (1992)] and recovered an impulse response with deconvolution during vowel phonation at various averaged F0 (low, at speech fundamental frequency, high) and in two vocal registers (modal and falsetto). The normalized response amplitude increased with F0, which replicates the previous finding, while the vocal register change did not affect the response. Our findings clarify some of the characteristics of tasks dependency of the pitch control.

3aPP13. Pitch strength, tone segregation, and frequency difference limen. Hans Hansen and Reinhard Weber (Acoust. Group, Inst. of Phys., Carl von Ossietzky Univ., Carl-von-Ossietzky-Strasse 9-11, 26111 Oldenburg, Germany, hans.hansen@uni-oldenburg.de)

For pure tones in narrowband noise, two percepts are considered. At a sufficiently high tone-to-noise ratio, the tone can be clearly distinguished from the noise: the tone and the noise segregate. At a low signal-to-noise (S/N) ratio, slightly above the masking threshold, only a tonal noise is perceived: the tone is fused with the narrowband noise. With respect to these phenomena, two hypotheses are tested relating pitch strength, tone segregation, and frequency difference limen. The first hypothesis is that with the aforementioned stimuli, pitch strength alone is capable to serve as a segregation cue. The second hypothesis tested is that frequency difference limen is the psychophysical correlate of this sensation. Three experiments were conducted. The first experiment was an identification experiment. Listeners had to identify a segregated tone at five center-frequencies (250–4000 Hz, octave-wise) in the presence of narrowband noise. The second experiment measured the detection threshold of the tone. The third experiment determined the frequency difference limen at various S/N ratios. The results suggest that not only pitch strength can serve as segregation cue but is tied to the frequency resolvability.

3aPP14. Low order all-pole and all-zero models of human and cat directional transfer functions. Bahaa Al-Sheikh, Mohammad Matin (Dept. of Elec. & Comput. Eng, Univ. of Denver, 2390 S. York St., Denver, CO 80208), and Daniel, J. Tollin (Univ. of Colorado Med. School, Aurora, CO 80045, daniel.tollin@ucdenver.edu)

Directional transfer functions (DTFs), the directional components of head related transfer functions, are generally measured at finite locations in azimuth and elevation. Thus models are needed to synthesize DTFs at finer spatial resolution than the measured data to create complete virtual auditory displays. Here, minimum-phase all-pole and all-zero models were used for modeling both human and cat DTFs. For the human DTFs, model orders were chosen to achieve specific objective error criteria published in previous studies that were based on subjective listening tests. Because subjective listening tests are not always feasible in animals, objective methods must be used to assess the quality of the DTF reconstructions. Here, the same error criteria reported in subjective tests of human DTF reconstructions were used to constrain models of cat DTFs based on the assumption that if humans cannot discriminate reconstructed versus empirical DTFs for a given objective reconstruction error, then cats will not be able to either provided the reconstruction error is held below that criteria. All-pole and all-zero models of orders as low as 25 were able to model DTFs with errors comparable to previous research findings and preserve the main spectral features in both human and cat DTFs. [Work supported by NIH Grant No. R01-DC6865.]

3aPP15. The role of spectral- and temporal-envelope room-acoustic cues in auditory selective attention. Simon Makin, Anthony Watkins, and Andrew Raimond (School of Psych. and Clinical Lang. Sci., Univ. of Reading, Earley Gate, Reading RG6 6AL, United Kingdom)

Listeners can attend to one of several simultaneous messages by tracking one speaker's voice characteristics. Using differences in the location of sounds in a room, we ask how well cues arising from spatial position compete with these characteristics. Listeners decided which of two simultaneous target words belonged in an attended "context" phrase when it was played simultaneously with a different "distracter" context. Talker difference was in competition with position difference, so the response indicates which cue-type the listener was tracking. Spatial position was found to override talker difference in dichotic conditions when the talkers are similar (male). The salience of cues associated with differences in sounds, bearings decreased with distance between listener and sources. These cues are more effective binaurally. However, there appear to be other cues that increase in salience with distance between sounds. This increase is more prominent in diotic conditions, indicating that these cues are largely monaural. Distances between spectra calculated using a gammatone filterbank (with ERB-spaced CFs) of the room's impulse responses at different locations were computed, and comparison with listeners' responses suggested some slight monaural loudness cues, but also monaural "timbre" cues arising from the temporal- and spectral-envelope differences in the speech from different locations.

3aPP16. The influence of different earphones on measurement values of simultaneous masking threshold. Juncong Shen and Zhiwen Xie (Acoust. Lab, Phys. Dept., School of Sci. South China Univ. of Technol., Guangzhou, China)

In many psychoacoustic experiments, earphones or headphones are used to replay sound signals. In this paper, the influence of earphones on measurement values of simultaneous masking threshold is investigated. Three kinds of earphones (DT770 pro HD250 and ER-2) are used to measure the simultaneous masking threshold in four critical bands for ten normal-hearing subjects by using a 90% MLE procedures. These earphones had been calibrated with B&K 4153 artificial ear before used. We conclude that the characters of headphones have important impact on the measurement values of simultaneous masking threshold.

3aPP17. Factors affecting a loudness asymmetry in real-room reverberation. A. Raimond and A. J. Watkins (Dept. of Psych., Univ. of Reading, Reading RG6 6AL, United Kingdom)

Narrowband stimuli with envelopes that have a slow-attack and a fast-decay (S-F) are judged to be louder than temporally-reversed, equal-energy versions of the stimuli (i.e., fast-attack and slow-decay, or F-S). We have found that this perceived loudness asymmetry is also apparent with real-room reflection patterns using dichotic presentation of sounds processed with binaural room impulse responses. These findings are consistent with the idea that the energy in the decaying tails of stimuli is discounted from loudness judgments because they are perceptually attributed to room reverberation. These results might therefore arise from the same perceptual mechanism that compensates for reverberation in speech which seems to operate within but not between frequency bands. Here, we ask whether the loudness asymmetry is similarly a "band-by-band" mechanism using conditions where loudness comparisons are made between S-F and F-S stimuli that are in the same or in different frequency bands.

3aPP18. Modeling the sensory distributions and decision space of frequency-discrimination tasks with deviant stimuli. Blas Espinoza-Varas (Commun. Sci. & Disord., Univ. Oklahoma Health Sci. Ctr., 1200 North Stonewall, Oklahoma City, OK 73117, blas-espinoza-varas@ouhsc.edu)

Discrimination tasks with deviant stimuli are used often in electrophysiological studies of selective auditory attention [e.g., Rinne *et al.*, Brain Res. **1077**, 135–143 (2006)]; for example, participants may be asked to discriminate between complex tones differing in fundamental frequency (186 versus 196 Hz). Presentation probability is high (e.g., 0.82) for fixed level (e.g., 60-dB SL) standard stimuli and low (e.g., 0.18) for deviant stimuli that are 3–9 dB higher or 3–9 dB lower than the standard. Since the level differences are irrelevant to the frequency-discrimination task, the longer response time and higher error rate of deviant versus standard stimuli may be interpreted in terms of distraction caused by deviance. In this study, the sensory effects of standard and deviant stimuli were modeled in terms of two-dimension (frequency by level) probability distributions with densities proportional to presentation probability. It was assumed that the perceptual dimensions (pitch and loudness) are partly correlated, and that perceptual decisions are based on a decision criterion controlled mainly by the standard frequency difference. The model suggests that, in addition to distraction, other factors could also account for deviance effects.

3aPP19. Frequency regions responsible for activating the efferent response in human ears. Kyle P. Walsh, Edward G. Pasanen, and Dennis McFadden (Dept. of Psych., Ctr. for Perceptual Systems, Univ. of Texas at Austin, 1 University Station A8000, Austin, TX 78712)

The nSFOAE is a nonlinear version of the stimulus-frequency otoacoustic emission (SFOAE). In past research, when a wideband noise was presented simultaneously with an ipsilateral 4.0-kHz tone, the nSFOAE to the tone showed a brief hesitation and then a gradual rise that asymptoted after about 100 ms. A similar dynamic response was also seen when the noise was low-pass filtered at 3.8 kHz. However, no dynamic response was seen when the noise was band-passed (400 Hz wide, centered at 4.0 kHz) or high-passed above 4.2 kHz. The interpretation is that the medial olivocochlear efferent system is activated by the wideband and low-pass noises. To identify the frequency components of the noise that are most relevant for activating the rising dynamic response, the nSFOAE was measured for the 4.0-kHz tone in the presence of 400-Hz bands of noise having a spectrum level of about 30-dB SPL and centered at 3.6, 3.2, 2.8, 2.4, and 2.0 kHz. Further exploration was done with noise bands 200 Hz in width and of 33-dB SPL spectrum level, as well as with additional frequencies of the tone. There were individual differences in the frequency regions yielding the maximum dynamic nSFOAE response. [Work supported by NIDCD.]

WEDNESDAY MORNING, 28 OCTOBER 2009 PECAN, 10:30 A.M. TO 12:00 NOON

Session 3aSA

Structural Acoustics and Vibration: Applications of Structural Acoustics and Vibrations I

Wen L. Li, Chair
Wayne State Univ., Dept. of Mechanical Engineering, 5050 Anthony Wayne Dr., Detroit, MI 39759-8270

Contributed Papers

10:30

3aSA1. Predicting the dynamic responses of and energy flows in complex systems in the presence of model uncertainty. Hongan Xu, Jingtao Du, and Wen L. Li (Dept. of ME, Wayne State Univ., 5050 Anthony Wayne, Detroit, MI 48202)

As frequency increases, the dynamic behaviors of mechanical systems tend to become unpredictable due to inherent uncertainties/errors with some input variables. In statistical energy analysis the uncertainties and variances of the system responses are typically alleviated via spatial and temporal (or frequency) averaging processes which prove to be very useful for high-frequency analysis. Such processes, however, do not explicitly deal with suspicious variables and are not entirely equivalent to the statistical processing of the input variables and system responses. In this study, a robust and sophisticated method is presented for the dynamic analysis of complex systems. By combining with Monte-Carlo simulations (or other statistical methods), this model can be effectively used to address the probabilistic impact of any model variable(s) and predict the statistics of the system responses. This approach is demonstrated by examining how the vibrational responses and energy flows in a two-plate system are affected by the randomness of the model variables like thickness, Young's modulus, coupling angle, etc. It is found that such a statistical process is particularly important and necessary in the medium frequency range where the ensemble means can still exhibit strong resonance-like behavior which may be of design interest and concerns.

10:45

3aSA2. Ultrasonic sound transmission through common building materials. Jayrin Farley (Dept. of Phys. & Astron., Brigham Young Univ., Provo, UT 84602, phantom_farley@hotmail.com) and Brian E. Anderson (Los Alamos Nat. Lab., Los Alamos, NM 87545)

Transmission loss measurements of building materials at audible frequencies are commonly made using various techniques such as plane wave tubes. These measurements provide vital information for control of noise isolation in architectural acoustics. To our knowledge, however, not much has been done to explore airborne ultrasonic sound transmission through common building materials. This paper will present various measurements of the ultrasonic transmission loss for various building materials over a frequency range of 20–200 kHz. The materials include fir plywood, particle board, medium density fiberboard, styrofoam, galvanized steel, and acrylic and polycarbonate plastics. This presentation will also discuss the challenges involved in making such measurements.

11:00

3aSA3. Spectral shaping of system response using an array of subordinate oscilator. Kalin Petersen (Mech. Eng., Catholic Univ., 620 Michigan Ave., Washington, DC 20064, 82petersen@cardinalmail.cua.edu) and John Judge (Catholic Univ., Washington, DC 20064)

This work demonstrates that the spectral response of an oscillating structure can be tailored by attaching one or more subordinate oscillators with designed properties. A closed form governing equation of motion is derived for the coupled system composed of the primary system and the subordinate oscillator array. Examples show that the frequency response can be tailored to desired shapes by prescribing the distributions of the dynamic properties in a relatively small array of such subordinate oscillators. In these examples, the number of attached oscillators is less 100 and having mass that is less than 1% of the mass of the primary structure, it is possible to tailor frequency response functions of the primary oscillator to have a nearly linear phase or constant amplitude over a frequency band of interest. The frequency range over which spectral shaping is achieved is determined by the band of the attached oscillators in isolation. We show that the common analytic design methodology for determining a dynamic vibration absorber represents the limiting case of a single oscillator in the subordinate set, but that increasing the number of subordinate oscillators (without increasing the total added mass) offers a number of advantages in reshaping the dominant system spectral response.

11:15

3aSA4. Vibro-acoustic analysis of an arbitrarily shaped vibrating structure. Logesh Natarajan, Xue-feng Zhang, Sean Wu, and Wen L. Li (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

Understanding the interrelationships between sound and vibrations of a structure is a crucial step toward developing the most cost-effective noise abatement strategy. In this paper, such interrelationships are established using Helmholtz equation least squares based nearfield acoustic holography. Specifically, the radiated acoustic pressure is linked to the normal surface velocity through a transfer function based on the spherical wave functions and spherical harmonics. The forced vibro-acoustic components (F-VACS) of a structure are identified by projecting the transfer function onto a set of mutually orthogonal basis functions. Finally, the radiated acoustic power is calculated, and contributions from individual F-VAC are indicated by the eigenvalues of the matrix that correlates the acoustic power to the normal surface velocity. The dominant F-VAC responsible for sound radiation is then linked to the dominant natural modes of the structure, which enables one to employ the most cost-effective measures to abate resultant sound by suppressing the identified natural modes. Such an approach to vibro-acoustic control is illustrated by a baffled square plate subject to a point force excitation under free-free boundary conditions. The effectiveness of this methodology on reducing sound radiation is demonstrated through numerical simulations. Similar methodology can also be used to abate structural vibrations.

11:30

3aSA5. Pressure and displacement fields inside an absorbing fluid sphere ensonified by a point acoustic source: comparison of analytic and numerical solutions. Kenneth G. Foote (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), David T. I. Francis (Univ. of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom), and Mario Zampolli (TNO Defense, Security and Safety, 2509 JG The Hague, Netherlands)

The interaction of an external point acoustic source in a lossy immersion medium with an absorbing fluid sphere is modeled. The analytic boundary-element-method and finite-element-method solutions for the internal pressure and displacement fields are evaluated numerically for a 50-mm-diameter fluid sphere in an immersion medium of mass density of 1000 kg/m^3 and sound speed of 1500 m/s. The mass density and sound speed of the fluid sphere are treated as parameters varying from 0.5 to 2 times the value of the corresponding medium property. Absorption is neglected in the immersion medium but is assumed to be 0 or 10 dB per wavelength in the fluid sphere. Each of two frequencies is considered, 10 and 100 kHz, for which the approximate wavenumber-radius products are 1 and 10, respectively. Results for the three solution methods are compared. [Work partly supported by NOPP through ONR Award No. N000140710992.]

11:45

3aSA6. Spatial mapping of modal damping in vibrating plates. Hande Öztürk and J. Gregory McDaniel (Dept. of Mech. Engrg., Boston Univ., 110 Cummington St., Boston, MA 02215)

Effective design and placement of damping treatments for complex structures benefit greatly from a knowledge of where the vibrational energy is being dissipated. Previously, the authors presented a method for spatially mapping contributions to the modal loss factor of a viscously damped structure [McDaniel *et al.*, J. Acoust. Soc. Am. **125**, 2603 (2009)]. The method is based on an inverse approach that uses experimental data to construct a damping matrix as a summation over elemental dashpots that connect measurement points to each other and to ground. This is followed by an assumption that ignores coupling in the modally transformed damping matrix, resulting in an expression for the modal loss factor as a weighted sum of elemental dashpots. In the present work, the method is applied to a vibrating plate with spatially distributed damping. An experiment is numerically simulated in which vibrational responses are measured at a small number of points. From these measurements, contributions to the modal loss factor are computed and used to construct a spatial map of damping effectiveness. Issues such as measurement error and modal coupling are used to further understand the method's potential. [Work supported by ONR under Grant No. N000140810531.]

Session 3aSC

Speech Communication: Fluid–Structure Interaction in Voice Production: Experiments and Modeling I

Zhaoyan Zhang, Chair
UCLA School of Medicine, 31-24 Rehabilitation Ctr., 1000 Veteran Ave., Los Angeles, CA 90095-1794

Chair's Introduction—8:00

Invited Papers

8:05

3aSC1. Toward high-fidelity computational fluid dynamics based tools for phonosurgery. Rajat Mittal, Xudong Zheng, Qian Xue (Dept. of Mech. Eng., Johns Hopkins Univ., Baltimore, MD 21218), and Steven Beilamowicz (George Washington Univ., Washington, DC 20052)

Vocal fold paralysis/paresis resulting from stroke accounts for nearly a third of all voice disorders diagnosed in the US. Stoke can damage the nerves that drive vocal fold adduction, leading to incomplete glottal closure and hence an inability to produce sustained vocal-fold vibrations. One common treatment for this pathology is medialization laryngoplasty wherein an implant is inserted into the paretic vocal fold in such a way as to force it to the glottal midline. If done properly, this procedure can lead to full voice recovery. However, the implant shape/location has to be determined intra-operatively through a trial-and-error process, and this can lead to a fairly high revision rate. The long term objective of the current project is to develop a computational fluid dynamics based tools that can help with the preoperative planning of this surgical procedure and thereby reduce the revision rate of this surgery. The tool is based on a coupled immersed-boundary-finite-element solver that can solve for the glottal flow as well as flow induced vibration of the vocal folds. We are currently using the solver to study the fundamental biophysics of phonation, and results from these studies will be presented. [Work supported by NIH.]

8:25

3aSC2. Synthetic and computational vocal fold modeling: Advances and issues. Scott L. Thomson (Dept. of Mech. Eng., Brigham Young Univ., Provo, UT 84604 thomson@byu.edu)

The use of computational and synthetic models to simulate vocal fold vibration is discussed. First, issues pertaining to computational modeling are addressed. The results of simulations using compressible and incompressible flow models coupled with identical finite element vocal fold models are compared. It is demonstrated that because incompressible flow models lack acoustic coupling, non-physical vocal fold motion may be predicted. Also discussed is the common use of contact segments to restrict medial vocal fold motion in order to prevent complete fluid mesh collapse and subsequent solver failure. The sensitivity of vocal fold vibration and intraglottal flow response to contact segment medial placement is quantified. Two advances in synthetic modeling are then presented. Using high-speed imaging, it is shown that models based on realistic vocal fold geometry (e.g., from MRI data) produce more life-like motion than models with similar sizes and material properties but more idealized geometries. Fabrication methods for modeling the thin vocal fold epithelial layer and the flexible superficial lamina propria layer are described, and corresponding improvements to vocal fold motion are demonstrated. [Work supported by NIH.]

8:45

3aSC3. Experimental study of fluid-structure interaction which mimics phonation. Michael Krane, Zachary Cates, and Shane Lani (ARL Penn State, State College, 16804-0030 PA, mhk5@only.arl.psu.edu)

This presentation will detail some measurements of a fluid-structure interaction in a physical model of the human vocal folds. The goal of the study is to empirically determine the partition of incident airstream energy into vocal fold kinetic and potential energy, acoustic field energy, and dissipation in the glottal jet and the vocal tract resonator. The vocal system model is life-scale, and the working fluid is air. The model is made of acrylic and aluminum, with a compliant coating on the model vocal fold to simulate the mucosal layer. Airflow through the model is provided by a compressed air supply. Experiments are performed in an anechoic chamber. Measurements presented include subglottal pressure, optical characterization of model vocal fold wall motion, and sound pressure outside the "mouth" of the model. The measurements are then used as empirical input for a theoretical model of the energy flow. Trends showing the relationship between the various mechanisms of energy transfer are presented, and the implications for vocal efficiency are discussed.

9:05

3aSC4. An inhomogeneous mechanical replica of the vocal folds. Xavier Pelorson and Annemie Van Hirtum (Gipsa Lab., Dept. Parole et Cognition, 961 rue de la Houille Blanche, Domaine Universitaire, BP 46, 38402 Saint Martin d'Hres, France, xavier.pelorson @gipsa-lab.inpg.fr)

In parallel with theoretical modeling, *in-vitro* mechanical replicas of the human larynx have become very popular in the speech science community. These experimental devices are indeed extremely useful in order to evaluate the relevance and the accuracy of

physical models of speech production. In this paper, a new mechanical model of the vocal folds is presented. It consists of a self-oscillating structure made of various layers with different mechanical properties (silicone, rubber, and/or liquid). The elasticity of the deeper layer can be controlled by stress/strain system. Although very simplified, such mechanical model is intended to replicate an inhomogeneous structure comparable to the human vocal folds one. The mechanical replica is then mounted on a air reservoir whose pressure can be controlled. The pressure upstream and downstream of the replica are measured together with the mechanical displacement of the artificial folds, thanks to a laser device. Typical examples of measurements will be presented for various alimentation pressures and elasticity conditions. The relevance of the observed effects with respect to human phonation and the possibility to explain them using simple mass-spring mechanical models will be discussed.

9:25

3aSC5. Experimental observations on the influence of supraglottal flow structures on phonation. Zhaoyan Zhang and Juergen Neubauer (UCLA School of Medicine, 31-24 Rehab. Ctr., 1000 Veteran Ave., Los Angeles, CA 90095)

The supraglottal flow has been shown to exhibit many complex phenomena such as vortex shedding, jet oscillation, and recirculation. However, the influence of such airflow features on phonation is still unknown. The objective of this study is to evaluate the influence of the supraglottal flow structures on phonation characteristics. In a mechanical model of phonation, the interaction of the supraglottal flow with vocal folds was studied by using vocal tract tubes of different lengths mounted at different off-center positions and by disturbing the flow with cylinders of different diameters. Preliminary results show that, at normal phonation conditions, changes in the supraglottal flow structures did not lead to noticeable changes in phonation threshold and sound quality. However, for vocal fold models simulating pathological conditions, disturbing the supraglottal flow led to significant changes in phonation frequency, amplitude, and perceived sound quality. Possible mechanisms that could explain the observed difference will be discussed. [Work supported by NIH R01 DC009229.]

9:45

3aSC6. The suction force on the vocal folds as an energy source for vibration and as an acoustic source. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420) and Michael S. Howe (College of Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

It has been shown that single-degree-of-freedom vibration of sprung shutters in a duct with a source of constant pressure is enabled when there is a change in separation point position on the shutters [M. S. Howe and R. S. McGowan, Fluid Dyn. Res. in press]. In this scenario, the flow is supposed to separate on the upstream side of the shutters as they open and on the downstream side as they close. Energy is inputted from the flow into the shutters as the shutters close because of a suction force necessary to turn a potential flow around the edge of each shutter. This suction force is also called the Coanda force in the aerodynamics literature. This time variable suction force is a dipole source of sound that is distinct from drag force source that has already been discussed [M. S. Howe and R. S. McGowan, J. Fluid Mech. **592**, 367–392 (2007)]. With these theoretical considerations, an equation analogous to Fant's equation for the glottal jet can be derived to explore glottal jet kinematics and the effect of the time-varying glottal end correction on the jet just prior to closure. [Work supported by DC-009229 to UCLA under subcontract].

10:05—10:20 Break

10:20

3aSC7. Convective water transport within the human larynx during phonation. Luc Mongeau, Shuangdong Wang, and Karine Terzibachi (Dept. of Mech. Eng., McGill Univ., 817 Sherbrooke St. West, Montreal, PQ H3A 2K6, Canada, luc.mongeau@mcgill.ca)

Proper hydration is known to be important in human phonation. Vocal fold tissue includes a significant amount of water. Sustained inhalation of partly dry air may result in convective water mass transport ultimately contributing to changes in tissue mechanical behavior, voice fatigue, and possibly trauma. In this study, numerical models were used to estimate the net amount of water mass transported from the vocal folds during various types of phonation. The commercially available code FLUENT was used for the mass transport analysis. Experiments were performed in a simple water channel to validate the approach. The convective transport coefficient for one type of hydrogel material was measured and used to approximately characterize that of real tissue. Efforts were made to estimate mass transport from numerical analysis and a number of *ad hoc* assumptions relative to respiratory patterns and ambient air humidity. The possible consequences of convective dehydration for voice production will be discussed. [Work supported by NIH.]

10:40

3aSC8. Empirical results using laryngeal models M5 and M6. Ronald C. Scherer (Dept. Commun. Sci. and Disord., 200 Heath Ctr., Bowling Green State Univ., Bowling Green, OH 43403, ronalds@bgnet.bgsu.edu), Saeed Torkaman, Bogdan R. Kucinschi (Univ. of Toledo, Toledo, OH 43606), Jun Li, Guangnian Zhai (Bowling Green State Univ., Bowling Green, OH 43403), Abdollah A. Afjeh (Univ. of Toledo, Toledo, OH 43606), and Lewis P. Fulcher (Bowling Green State Univ., Bowling Green, OH 43403)

Laryngeal models M5 two dimensional and M6 [three dimensional and eccentric] contribute empirical data on intraglottal and transglottal pressures and flows for numerous glottal configurations. M5 provides results for both symmetric and oblique glottal shapes. M6 has a 3-D glottis (the membranous glottis width is greatest at the A-P midpoint) and is eccentric (with modeled arytenoid cartilages blocking the posterior glottis so that the glottis acts as an eccentric orifice). Results from these models typically yield asymmetric pressures which will be presented and discussed. M5 obliquity results will be used in a multimass model of phonation to indicate the potential effects of asymmetric pressures on glottal flow and stability. [Research supported by NIH Grant #2R56DC003577].

11:00

3aSC9. Parametric excitation of the acoustic eigenmodes of the vocal tract by motion of the vocal folds: Non-aeroacoustic mechanisms of phonation. Gordon Ramsay (Haskins Labs., 300 George St., New Haven, CT 06511, ramsay@haskins.yale.edu)

Studies of voice production have focused on aeroacoustic mechanisms of phonation, describing how energy in the background flow is converted into sources of sound. This paper describes a non-aeroacoustic mechanism that may be important in explaining how sound is generated by vocal fold vibration. According to dynamical systems theory, temporal variations in the parameters of a system may result in parametric excitation of the system. Under appropriate conditions, modulation of the parameters of the acoustic wave operator governing sound propagation in the vocal tract by changes in glottal geometry should be sufficient to induce parametric excitation of the vocal tract eigenmodes, creating sources of sound additional to those predicted by classical aeroacoustics. Rapid changes in the shape of the glottis during phonation are already known to create modulations of the formants within each glottal cycle, but the role of these modulations in exciting the formants has not been established. To test this hypothesis, we use a time-domain finite-volume simulation of acoustic wave propagation in a time-varying vocal tract to compare aeroacoustic and parametric excitation mechanisms. Results show that parametric excitation terms that transfer energy between formants, mainly at the instant of glottal closure, contribute significantly to the overall glottal source.

WEDNESDAY MORNING, 28 OCTOBER 2009 RIO GRANDE WEST, 7:55 A.M. TO 12:00 NOON

Session 3aSP

Signal Processing in Acoustics: Kalman and Particle Filters in Acoustics

Peter Gerstoft, Cochair
Univ. of California, San Diego, Scripps Inst. of Oceanography, 9500 Gilman Dr., La Jolla, CA 92093-0238

Zoi-Heleni Michalopoulou, Cochair
New Jersey Inst.Technology, Dept. of Mathematics, Newark, NJ 07102-1982

Chair's Introduction—7:55

Invited Papers

8:00

3aSP1. Overview of Kalman and particle filters for acoustic applications. Zoi-Heleni Michalopoulou (Dept. of Math. Sci., New Jersey Inst. of Technol., Newark, NJ 07102, michalop@njit.edu), Peter Gerstoft, and Caglar Yardim (Marine Phys. Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, La Jolla, CA 92093-0238)

In this talk, the basic principles behind Kalman and particle filtering are presented. These filters are recursive Bayesian estimators for the state of a dynamic system from a series of noisy measurements. Starting from the simple Kalman filter, we focus on computationally efficient particle filters or sequential Monte Carlo methods. These are often used to formulate and solve problems in a Bayesian framework and are the sequential analog of Markov chain Monte Carlo approaches. Applications where complex problems in acoustics are solved with particle filters will be shown. We will examine source localization and tracking, geoacoustic inversion, and tracking of acoustic field features and discuss the advantages of specific filtering techniques. [Work supported by ONR.]

8:20

3aSP2. Experimental demonstration of geoacoustic tracking using particle filters. Caglar Yardim (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0238, cyardim@ucsd.edu)

This paper discusses practical aspects and performance of particle filters in geoacoustic inversion using data from the MAPEX experiment north of Elba Island, Italy in 2000. The particle filter enables spatial and temporal tracking of environmental parameters and their underlying probability densities, making geoacoustic tracking a natural extension to geoacoustic inversion techniques. The setup includes a ship-towed horizontal line array and a moored, broadband acoustic source in a shallow water environment. The ship track includes two sharp turns that cause bending of the array. The particle filter is used to track the ship parameters (range, bearing, and speed), array parameters (tilt and bow), and environmental parameters (water depth and sediment sound speed). The filter performance is discussed for both the fast turns and the straight, constant velocity sections of the ship track. The results are compared with the inversion results obtained using genetic algorithms. [Work supported by ONR.]

8:40

3aSP3. Extended Kalman filters for tracking time-frequency striations. Lisa M. Zurk, Tobias Oesterlein, Chensong He, and Jorge Quijano (ECE Dept., Portland State Univ., Portland, OR 97201, zurkl@pdx.edu)

The time-frequency structure of the received intensity from a moving source in a shallow water channel is typically observed as striations in the target spectrogram. The relationship between the slope of the striations and the source location can be described with the waveguide invariant, and this relation can be incorporated in the sonar signal processing for improved detection or tracking. However, incorporating this structure requires a robust method of estimating the presence and location of the target striations. In this paper a multi-target extended Kalman filter is presented to track the striations. The filter identifies each striation as a potential target, and then uses the invariance relationship to execute the Kalman update stage. The physics of the channel are further incorporated through the imposition of multi-track consistency and bounds on striation migration. In addition to the striation tracker, a second algorithm is presented which is appropriate for spectrograms with insufficient resolution (either in time or frequency) to visualize the striations. This second algorithm provides a confidence test for a given track hypothesis by accumulating likelihood information along a predicted striation. Results of both algorithms are presented for simulated and real active sonar data.

9:00

3aSP4. Bayesian source tracking in an uncertain ocean environment. Stan E. Dosso and Michael J. Wilmut (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, V8W 3P6 Canada, sdosso@uvic.ca)

This paper considers matched-field tracking of a moving acoustic source when properties of the ocean environment (water column and seabed) are poorly known. The goal is not simply to estimate source locations but to determine track uncertainty distributions, thereby quantifying the information content of the tracking process. To localize and track low-level sources, acoustic data collected for multiple time samples (corresponding to multiple source positions) are inverted simultaneously, with constraints included on the maximum allowable motion between samples. This increases the information content over sequential tracking approaches such as particle filtering, but also increases the dimensionality and difficulty of the inversion. A Bayesian formulation is applied in which the posterior probability density (PPD) is integrated over unknown environmental parameters to obtain a time-ordered sequence of joint marginal probability surfaces over source range and depth. Marginal PPDs are computed numerically using Metropolis–Hastings sampling over environmental parameters (rotated into principal components and applying linearized proposal distributions) and two-dimensional Gibbs sampling over source locations. The approach is illustrated using Mediterranean Sea data, and tracking information content is considered as a function of data quantity (number of time samples and frequencies processed), data quality (signal-to-noise ratio), and level of prior information on environmental parameters.

9:20

3aSP5. Efficient use of particle filters for multi-source data fusion in decentralized sensor networks. Philip J. Haney and Geoffrey S. Edelson (BAE Systems, Electron. Solutions, MER15-2651, P.O. Box 868, Nashua, NH 03061-0868, philip.j.haney@baesystems.com)

A highly-scalable Bayesian approach to the problem of performing multi-source data fusion and target tracking in decentralized sensor networks is presented. Previous applications of decentralized data fusion have generally been restricted to uni-modal/uni-source sensor networks using Gaussian based approaches, such as the Kalman or information filter. However, with recent interest to employ complex, multi-modal/multi-source sensors which potentially exhibit observation and/or process non-linearities along with non-Gaussian distributions, the need to develop a more generalized and scalable method of decentralized data fusion using particle filters is required. The probabilistic approach featured in this work provides the ability to seamlessly integrate and efficiently fuse multi-source sensor data in the absence of any linearity and/or normality constraints. The architecture is fully decentralized and provides a methodology that scales extremely well to any growth in the number of targets or region of coverage. This multi-source data fusion architecture is capable of providing high-precision tracking performance in complex, non-linear/non-Gaussian operating environments. In addition, the architecture provides an unprecedented scaling capability for decentralized sensor networks as compared to similar architectures which communicate information using particle data, Gaussian mixture models, or Parzen density estimators.

9:40

3aSP6. Bayesian tracking predictions in an uncertain ocean environment. Stan E. Dosso and Michael J. Wilmut (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, V8W 3P6 Canada)

This paper considers probabilistic prediction of the future locations of a moving acoustic source in the ocean based on past locations as determined by Bayesian source tracking in an uncertain environment. The Bayesian tracking approach considers both source and environmental parameters as unknown random variables constrained by noisy acoustic data and prior information and numerically integrates the posterior probability density (PPD) over the environmental parameters to obtain a time-ordered sequence of joint marginal probability surfaces over source range and depth. The integration is carried out using Markov-chain Monte Carlo sampling methods which provide a large collection of track realizations drawn from the PPD. Applying a probabilistic model for source motion to each of these realizations produces a sequence of source range-depth probability distributions for future times. These predictions account for both the uncertainty of the source-motion model and the uncertainty in the state of knowledge of past source locations, which is itself dependent on environmental uncertainty. Several source-motion models, which differ in the degree of confidence assigned to future predictions based on past locations, are considered for range-depth tracking and prediction using a vertical sensor array.

10:00—10:15 Break

10:15

3aSP7. Comparison and analysis of equalization techniques for the time-varying underwater acoustic channel. Ballard J. S. Blair (MIT, 77 Massachusetts Ave., Cambridge, MA 02139 and Woods hole Oceanograph. Inst. Joint Program, 266 Woods Hole Rd., Woods Hole, MA 02139, bjblair@mit.edu) and James C. Preisig (Woods Hole Oceanograph. Inst., Woods Hole, MA 02139)

It has long been accepted that adapting a decision feedback equalizer in a time-varying environment by first estimating the channel impulse response and using this estimate to calculate equalizer tap coefficients results in better performance than is achieved by directly adjusting the equalizer taps. There are several empirical demonstrations in the literature, [Shukla and Turner, IEE Proc.-Commun. **138**, 525–543 (1991)], but the explanations for the performance difference are varied and none uses analytical arguments. Analysis shows that at high signal-to-noise ratio (SNR) for time varying channels, the taps of the feed-forward equalizer fluctuate more rapidly than do the taps of the channel impulse response and using channel estimate based adaptation is more effective. The largest changes in the equalizer coefficients occur when the input signal correlation matrix has a large condition number. At low SNR the matrix is better conditioned and the performance gap disappears. These results assume that the delay spread of the channel impulse response can be correctly estimated. When this is not the case, directly adapting the equalizer coefficients results in better performance than is achieved using the channel estimate based approach. [Work supported by ONR Grant Nos. N00014-05-10085 and N00014-07-10184.]

10:30

3aSP8. Comparison of particle filter and histogram filter performance for passive sonar localization. Colin W. Jemmott (Penn State Graduate Program in Acosutics, P.O. Box 30, State College, PA 16804, cwj112@psu.edu), R. Lee Culver, and Jack W. Langelaan (Penn State Univ., State College, PA 16804)

A number of extensions of the famous Kalman filter exist for recursive Bayesian estimation of state vectors with nonlinear update equations and non-Gaussian prior probability density functions. Two of the most powerful are the particle filter, which uses a direct Monte Carlo approach, and the histogram filter, which uses a grid-based approach. Both filters are numerical implementations of recursive Bayesian estimation. The histogram filter uses a grid based approach that is analogous to midpoint rectangular integration, while the particle filter uses a Monte Carlo approach. Their performance is compared for a passive, model-based localization problem in which the target is broadcasting a low-frequency tonal signal while moving through a shallow water waveguide. The state vector being estimated includes range, depth, and source level. Target dynamic models allow for slow changes in velocity, depth, or source level. The example uses RAM realizations with additive noise as a synthetic received signal. In situations where computational power is restricted, the particle filter consistently outperforms the histogram filter in localizing the target. However, with sufficient computational power the performance is equivalent. [This work was sponsored by ONR Undersea Signal Processing.]

10:45

3aSP9. Hybrid grid-particle Bayesian tracking with intelligent likelihood sampling. Jason M. Aughenbaugh and Brian R. La Cour (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029, jason@arlut.utexas.edu)

Characteristic of anti-submarine warfare is the goal of detecting and localizing a single target in a noisy environment in which subtle clues in the signals are essential for differentiating clutter and the target of interest. Bayesian inference is a general framework for performing target tracking in these applications, but both particle filter and grid-based models of the distributions have limitations. Particle filters involve fast motion updates but slow measurement updates. Grid-based models allow for faster measurement updates at the cost of slower motion updates. A particle filter provides adaptive modeling of the prior and posterior distributions, but both methods can fail to provide detailed sampling of the likelihood function in crucial regions, and these small details are often the key to mitigating clutter in active systems. In this paper, a hybrid grid-particle model with intelligent likelihood sampling is introduced. The density is approximated by a fixed grid, and the motion update is implemented using a particle-based sampling of the grid. The integral of the likelihood over each grid cell is calculated in an intelligent manner that adapts to the received measurements and the known sensor model properties. [This work was supported by the Office of Naval Research Contract No. N00014-06-G-0218-01.]

11:00

3aSP10. A particle filtering approach for multipath arrival time estimation from acoustic time series. Rashi Jain and Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102, rj45@njit.edu)

Accurately estimating arrival times from acoustic time series in the ocean leads to a wealth of information on the geometry of the sound propagation environment and environmental parameters such as sound speed in sediments. In this work, a sequential Monte Carlo method is developed that treats the spatial variation of signal paths as moving targets, dynamically modeling their temporal location at spatially separated receiving phones. The spatial rather than the temporal sequential variability of the problem permits the application of smoothing techniques to the arrival time estimates, significantly reducing uncertainty. The results are compared to maximum likelihood estimates for the same test cases; the comparison demonstrates an advantage in using the proposed approach, which can be employed for reduction in uncertainty in arrival time estimation and, consequently, in geometric and geoacoustic inversion. It is further demonstrated how this method accurately extracts path amplitudes, which can then be employed for attenuation inversion. This method, suitable for environments with an unknown number of arriving paths, is also tested successfully on Haro Strait Primer Experiment and Shallow Water 06 data. [Work supported by ONR.]

11:15

3aSP11. Use of within-beam mapping in conjunction with Kalman filtering to improve angle of arrival estimation accuracy in multibeam echo-sounding. Daniel S. Brogan and Kent A. Chamberlin (Dept. of Elec. and Comput. Eng., Univ. of New Hampshire, 33 Academic Way, Durham, NH 03824-2619, daniel.brogan@unh.edu)

The positions of both water column scatterers and the ocean floor can be estimated using data from multibeam sonars. In a previous presentation [D. S. Brogan and C. P. de Moustier, J. Acoust. Soc. Am. **119**, 3352 (2006)] it was shown that a Kalman filter could be used to track echo angles of arrival (bearings) versus time for such targets using quadrature-sampled multibeam sonar data. In that research, the inputs to the Kalman filter were the echo magnitude time series of each beam placed on that beam's maximum response axis. While the results using this approach proved to be superior to those of a conventional split-aperture processor, the bearing accuracy can be further improved by applying a pre-processing stage that maps each echo to its acoustic center rather than to its beam's maximum response axis. This mapping is accomplished using phase and/or amplitude monopulse techniques that provide linear mappings between the sine of the echo bearing and (1) the differential signal phase and (2) the natural logarithm of the ratio of the signal magnitudes, respectively, of neighboring beams to provide within-beam echo bearing estimates. Results are compared both with and without the pre-processing stage.

11:30

3aSP12. Frequency domain tracking of passive vessel harmonics. George Ogden, Lisa Zurk, Martin Siderius, Eric Sorensen (Elec. and Comput. Eng. Dept., Portland State Univ., 1900 SW 4th Ave., Ste. 160, Portland OR, 97201, ogdengl@gmail.com), Josh Meyers, Shari Matzner, and Mark Jones (Pacific Northwest Nat. Lab., Sequim, WA 98109)

This paper presents a method for passive acoustic detection and tracking of small vessels in noisy, shallow water marine environments. Passive spectra of boats include broadband noise as well as tones that are harmonics of

the engine speed and shaft/propeller rotation. Past work suggests that the location in frequency and the relative amplitudes of these harmonics can be used to determine specific characteristics of the vessel such as the number of blades on the propeller and engine type/speed. However, the low signal to noise ratio of quiet targets and Doppler shifts incurred because of source and receiver motion complicate the identification of these tones in the lofargram. To address this issue, a combined detection and tracking approach is proposed in which intermittent and wandering harmonic content is tracked with a multi-dimensional Kalman filter. Results from recorded passive signatures from several classes of vessels in marine and freshwater environments in the Pacific Northwest are presented and discussed.

11:45

3aSP13. Spatial and acoustic feature extraction by particle filtering for efficient noise reduction. Mitsunori Mizumachi (Kyushu Inst. of Tech., 1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka 804-8550, Japan, mizumach @ecs.kyutech.ac.jp)

Noise reduction can be achieved relying on something related to spatial and acoustic features: a direction-of-arrival (DOA), a fundamental frequency, harmonicity, and so on. It is natural that a suitable noise reduction approach is different depending on the available features, because no all-purpose noise reduction method has been established under every possible acoustic condition. To achieve noise reduction effectively and efficiently, it is important to explore and acquire well-suited spatial and acoustic features at all times. Observed acoustic signals originate in some dynamic systems such as a speech production system and behavior of continuous sound source movement, which can be represented by the Markov model in the perspective view, but they are usually distorted by noise and reverberation. In this paper, both feature extraction and feature tracking are attempted by using particle filters aiming at achieving efficient noise reduction. Particle filters can also bring a new feature of the reliability of each DOA estimate based on the effective sample size through the state estimation of spatial features [Mizumachi, Proceedings of the 156th ASA meeting]. In this paper, spatial and acoustic feature extraction by particle filtering is evolved into noise reduction. [Work supported by NEDO, Japan.]

WEDNESDAY MORNING, 28 OCTOBER 2009 RIO GRANDE CENTER, 8:30 TO 11:30 A.M.

Session 3aUW

Underwater Acoustics: Propagation in Multipath Environments

George E. Ioup, Chair
Univ. of New Orleans, Dept. of Physics, New Orleans, LA 70148

Contributed Papers

8:30

3aUW1. Blind deconvolution of remote-source signals from acoustic array recordings in multipath environments. Shima Hossein Abadi (Dept. of Mech. Eng., Univ. of Michigan, 2010 W.E. Lay Automotive Lab., 1231 Beal Ave., Ann Arbor, MI 48109, shimah@umich.edu) and David R. Dowling (Univ. of Michigan, Ann Arbor, MI 48109)

Reconstructing the signal from a remote source in an unknown multipath environment is a task commonly known as blind deconvolution. The existence of generic features of underwater sound fields, propagating modes or ray paths, may be exploited for blind deconvolution via artificial time reversal (ATR), [Sabra and Dowling, J. Acoust. Soc. Am. **116**, 262–271 (2004)] from vertical array recordings. ATR uses a weighted sum of array recordings to determine a frequency-dependent phase correction that aligns the various mode or ray-path arrivals at the receiving array to form an estimate of the original source waveform. A description of ATR will be presented along with parameter studies of ATR performance for simulated signals in a 100-m-deep Perkeris waveguide, and for airborne sounds measured in a reverberant laboratory. Here, the correlation coefficient between original and the ATR-reconstructed signals will be presented as a function of the number and spacing of array elements, and signal characteristics (frequency and bandwidth). The limitations imposed by noise will be quantified for the measured signals. Possible applications of ATR will be briefly discussed; these include reducing inter-symbol interference in underwater communication, tracking and identifying remote sources, and monitoring marine mammals that vocalize. [Work supported by ONR Code No. 321OA.]

8:45

3aUW2. Three-dimensional data analysis for acoustic measurements of a seismic airgun array. Arslan M. Tashmukhambetov, George E. Ioup, Juliette W. Ioup (Dept. of Phys., Univ. of New Orleans, New Orleans, LA 70148, atashmuk@uno.edu), Natalia A. Sidorovskaia, Anca Niculescu (Univ. of Louisiana at Lafayette, Lafayette, LA), Joal J. Newcomb (Naval Ocenographic Office, Stennis Space Ctr., MS), James M. Stephens, Grayson H. Rayborn (Univ. of Southern Mississippi, Hattiesburg, MS 39401), and Phil Summerfield (ExxonMobil Corp., Houston, TX)

The Source Characterization Study 2007 measured the three-dimensional acoustic field of a seismic airgun array. The Littoral Acoustic Demonstration Center performed the experiment, collecting acoustic and related data on three moored hydrophone arrays and one ship-deployed hydrophone array which together spanned the full water column. Sensitive and desensitized phones were deployed at each position to extend the dynamic range. An ultra short baseline localization system was deployed with the EARS moorings to provide array shape. With post analysis this results in time-dependent positions for each of the acoustic sensors. Every channel is calibrated. A seismic source vessel shot a series of lines designed to give detailed angle and range information concerning the field of the primary arrival. The experiment was conducted in the western Gulf of Mexico between the East Break and Alamos Canyon regions. Peak pressures, sound exposure levels, total shot energy spectra, and one-third octave band analyses are measures used to characterize the field. Images of these quantities are generated to show dependence on emission and azimuthal angles and range. [Research supported by the Joint Industry Programme through the International Association of Oil and Gas Producers.]

9:00

3aUW3. Use of high performance computing resources for underwater acoustic modeling. Anca M. Niculescu, Natalia A. Sidorovskaia, Peter Achi (Univ. of Louisiana at Lafayette, UL Box 44210, Lafayette, LA 70504-4210, axn1417@louisiana.edu), Arslan M. Tashmukhambetov, George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

The majority of standard underwater propagation models provide a two-dimensional (range and depth) acoustic field for a single frequency point source. Computational resource demand increases considerably when the three-dimensional acoustic field of a broad-band spatially extended source is of interest. An upgrade of the standard parabolic equation model RAM for use in a high-performance computing (HPC) environment is discussed. A benchmarked upgraded version of RAM is used in the Louisiana Optical

Network Initiative HPC-environment to model the three-dimensional acoustic field of a seismic airgun array. Four-dimensional visualization (time and space) of the generated data volume is also addressed. [Research supported by the Louisiana Optical Network Initiative, TeraGrid Fellowship, and the Joint Industry Programme through the International Association of Oil and Gas Producers.]

9:15

3aUW4. A level set method for high-frequency ocean acoustic propagation. Sheri Martinelli and Andrew Fredricks (Torpedo Systems Dept., Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841, sheri.martinelli@navy.mil)

The level set method due to Osher and Sethian [J. Comput. Phys. **79**, 12–49 (1988)] provides a way to obtain fixed grid solutions to the high-frequency wave equation. Instead of tracing several rays from the source, the level set method embeds the entire wavefront implicitly in the phase space and propagates it according to the velocity field determined by the local ray direction, thus avoiding the complications involved in the spatial reconstruction of the wavefronts from diverging and scattered rays. A level set method has been developed and implemented as a fixed-grid alternative to ray tracing to solve the high-frequency equation for the acoustic phase. Furthermore, an efficient ray-based method that takes advantage of the full knowledge of the wavefronts afforded by the level set solution to the Eikonal equation has been developed to solve for the geometric spreading term in the expansion. Preliminary results are presented with comparisons to ray tracing and full wave equation solutions. [Work supported by ONR.]

9:30

3aUW5. Temporal and spatial coherences of sound at 250 Hz and 1659 km in the Pacific ocean: Demonstrating deterministic effects and internal waves explain observations. John Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316)

The hypothesis tested is that internal gravity waves explain temporal and spatial coherences of sound at 1659 km in the Pacific ocean for a signal at 250 Hz and a pulse resolution of 0.02 s. From data collected with a towed array, the measured probability that coherence time is 1.8 min or longer is 0.8. Using a parabolic approximation for the acoustic wave equation with sound speeds fluctuating from internal waves, a Monte-Carlo model yields coherence time of 1.8 min or more with probability of 0.9. For spatial coherence, two subsections of the array are compared, which are separated by 142 and 370 m in directions perpendicular and parallel to the geodesic, respectively. Measured coherence is 0.54. This is statistically consistent with the modeled 95% confidence interval of [0.52,0.76]. The difference of 370 m parallel to the section causes spatial coherence to degrade deterministically by a larger amount than the effect of internal waves acting on the 142-m separation perpendicular to the section. The models are run without any tuning with data.

9:45

3aUW6. Incorporating array processing to improve resolution of ambient noise data cross-correlation along a horizontal array. Stephanie E. Fried and William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238)

Numerous authors have shown that the cross-correlation of ambient noise recordings at two points can produce an approximation of the Green's function between those points. The resolution of the cross-correlation is fundamentally limited by the characteristics of the noise field and the changing environment. In order to make this technique the most effective, we wish to limit the time needed to extract the desired environmental information. To that end, array processing techniques can be adapted to work with the noise cross-correlation function taken along a horizontal array to decrease the time needed to resolve multiple individual returns. For a given noise field, beam forming between horizontal arrays is effective in reducing the time needed to identify both the direct and single-bounce surface reflection paths.

10:00—10:15 Break

10:15

3aUW7. Cross-correlation of ship noise for water traffic monitoring. Laurent Fillinger, Alexander Sutin, and Alexander Sedunov (Stevens Inst. of Technol., Castle Point on Hudson, Hoboken, NJ 07060, lfilling@stevens.edu)

Various aspects of the monitoring of ship traffic using correlation of signals recorded by a pair of hydrophones are considered and demonstrated on real signals recorded in the Hudson River. The underwater acoustic noise generated by ships reaches the various hydrophones with a delay depending on their relative positions. That delay can be extracted by cross-correlation and can serve as a basis for determination of the direction of the ship. This method allowed finding directions for several ships in heavy traffic of Hudson River. The vessel triangulation can be done using information from two or more appropriately located hydrophone pairs. Another application demonstrated is separation of the acoustic signature (noise spectra) from several ships. The last application is the estimation of the ship noise modulation spectrum that is related to propeller and shaft rotation by means of short time cross-correlation. Comparison with the conventional detection of envelope modulation on noise method shows a close match. [This work was partially supported by ONR Project N00014-05-1-0632: Navy Force Protection Technology Assessment Project and by the U.S. Department of Homeland Security under Grant No. 2008-ST-061-ML0002.]

10:30

3aUW8. Acoustic communication from a moving source: Data results and model simulations. Aijun Song, Mohsen Badiey (College of Marine and Earth Studies, Univ. of Delaware, Robinson Hall 114, Newark, DE 19716), Julia Hsieh, Daniel Rouseff (Univ. of Washington, Seattle, WA 98105-6698), H.-C. Song, and William Hodgkiss (Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Acoustic communication can provide a flexible way to exchange information among moving platforms in the ocean. However, source-receiver motion poses additional difficulties to high-rate coherent communication. During the KAM08 experiment conducted in the summer of 2008, high-frequency (greater than 10-kHz) communication sequences were transmitted by a towed source and observed by multiple hydrophone arrays in the shallow water region west of Kauai, HI. Using time reversal combining followed by a single channel decision feedback equalizer aided by Doppler and channel tracking, reliable communication was achieved for different source-receiver geometries for the moving source. Further, the ray-based ocean acoustic time series simulation model sonar simulation toolset with a rough sea surface and moving source configuration was used to simulate the acoustic channel and received communication data. Simulated signals and demodulated results were compared with experimental data in order to examine the factors that significantly impede communication effectiveness for moving platforms. [Work supported by ONR 321OA.]

10:45

3aUW9. Passive acoustic classification of vessels in the Hudson River. Michael L. Zucker, Alexander Sedunov, Vladimir Zhdanov, and Alexander Sutin (Davidson Lab., Stevens Inst. of Technol., Castle Point on Hudson, Hoboken, NJ 07030)

Stevens Institute of Technology is conducting research aimed at the development of a low-cost passive acoustic system for detection and classification of underwater and surface threats. The experimental system is comprised of a hydrophone array and a stand alone acoustic buoy. Video and photographs of the passing ships were also captured simultaneously during the acoustic measurements. For classification purposes special attention was paid to extraction of the specific parameters of vessel acoustic signatures in high-frequency band (10–60 kHz) using detection of envelope modulation on noise algorithm. The spectrum of the noise envelope contains shaft and blade frequencies and their harmonics. Parameters extracted include the frequency of the fundamental peak of the ship noise emission, the amplitude modulation of the fundamental tone, the number of peaks found in the spectrum, the slope of the spectral peaks, as well as the calculation of the total

harmonic distortion. Parameters extracted from various records are then plotted together and their statistical analysis was applied for surface ship classification. [This work was supported by the U.S. Department of Homeland Security under Grant No. 2008-ST-061-ML0002].

11:00

3aUW10. Toward acoustic tomography in a shallow estuary. Alexander Sutin, Alan Blumberg, Laurent Fillinger, Nikolay Sedunov, and Alexander Sedunov (Stevens Inst. of Techn., 711 Hudson St., Hoboken, NJ 07030, asutin@stevens.edu)

There is presently much interest in understanding the hydrophysical parameters in a shallow estuary. The importance of predicting the currents in Hudson River was demonstrated during rescue and recovery of the USAirways aircraft crashed in the river on January 15, 2009. Usually hydrophysical parameters are measured by fixed location sensors that cannot provide high spatial resolution. Acoustic tomography, on the other hand, can map the hydrophysical parameters. Recently, Stevens Institute of Technology using their fully equipped underwater facility including two surface vessels has conducted the first feasibility test on acoustic tomography. The test was conducted in the Hudson River, a shallow estuary that is connected to New York Harbor. The test used an estimate of the differential time of flight of an acoustical signal in opposite directions between two nodes. One node was at a fixed location, while the other was on a research vessel drifting on the other side of the river. An application of a frequency sweep signal (30–90 kHz) and a cross-correlation technique allowed measurements of the difference in time of flight of the acoustic signals propagated in opposite directions that is proportional to the current in the river. The average current speed was estimated and correlated well with current measurements available from the Stevens Observing and Prediction System. [This work was supported by ONR Project #N00014-05-1-0632: Na.]

11:15

3aUW11. Experiment results of time-reversal communication in the deep ocean at the range of 300 km. Takuya Shimura, Hiroshi Ochi, Yoshitaka Watanabe (Adv. Marine Tech. RD Program, JAMSTEC, 2-15 Natsushima-cho, Yokosuka 237-0061, Japan, shimurat@jamstec.go.jp), and Takehito Hattori (Nippon Marine Enterprises, Ltd., Yokosuka 238-0004, Japan)

The research is being conducted to realize communication with a long cruising AUV in the deep ocean using time reversal. We have proposed the method of combining time reversal and adaptive equalization. In our previous at-sea experiments, the performance of time-reversal communication was confirmed at the range of 10–100 km with a virtual array. In our latest sea-trial, experiments of passive time-reversal communication were executed at the range of 300 km with a real receiver array. As a probe signal, chirp pulses and *M*-sequence pulses were used. The data transmission speed was 50 and 100 bps. The source was suspended from the research vessel and the 20 channel receiver array was moored. Thus, Doppler effect was induced due to the ship drifting at the speed of approximately 0.25 m/s at maximum. Nevertheless, such effect is compensated and demodulation with no error is achieved with the proposed method. The results of using chirp pulses are inferior to the results of using *M*-sequences because of sidelobes of the chirp correlation. The results of using longer *M*-sequence are degraded due to Doppler effect.

WEDNESDAY AFTERNOON, 28 OCTOBER 2009 REGENCY EAST 3, 1:30 TO 3:00 P.M.

Session 3pAA

Architectural Acoustics, Noise and Musical Acoustics: Acoustics and Theater Consulting: A Special Relationship

Scott D. Pfeiffer, Chair
Threshold Acoustics, LLC, 53 West Jackson Blvd., Chicago, IL 60604

Chair's Introduction—1:30

Panel Discussion

Panelists

.

Robert Campbell Fisher, Dachs Assoc., 22 W. 19th St., 6th Fl., New York, NY 10011

John Coyne, Theater Projects Consultants, Inc., 25 Elizabeth St., S. Norwalk, CT 06854

Todd Hensley, Schuler Shook, Theater Planners, 750 North Orleans, Ste. 400, Chicago, IL 60654

Robert Long, Theater Consultants Collaborative, 6600 Manor Hill Court, Chapel Hill, NC 27516

As a second in a series, this session explores the architectural acoustic consulting process from the point-of-view of the theater consultant. Four respected theater consultants talk about their point-of-view on the interactions within the design team that have led to success, and the obstacles that have reduced the potential outcomes. These brief introductions will be followed by a moderated panel discussion. The session will open with a Chair's introduction, 25 min of introductory comments by the panelists (authors), and a moderated panel discussion.

Session 3pED

Education in Acoustics: Acoustics Education Prize Lecture

Thomas G. Muir, Chair
Univ. of Mississippi, National Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—2:15

Invited Paper

2:20

3pED1. Fundamental acoustics education and applications. James V. Sanders (Dept. of Phys., Naval Postgrad. School, 833 Dyer Rd., Bldg. 232, Monterey, CA 93943)

Teaching acoustics at the graduate level to professional naval officers, who, after graduation, will go back to driving ships, submarines, and airplanes, as well as other professional naval disciplines, offers a unique challenge. The Naval Postgraduate School has been the home for over 50 years of the textbook, *Fundamentals of Acoustics*, originally written by Lawrence Kinsler and Austin Frey and revised in later editions by Alan Coppens and James Sanders. Updating a textbook that is suitable for undergraduate and graduate students in a multitude of disciplines at civilian institutions and also suitable for use by naval officers interested in underwater acoustics continues to be most challenging. Solutions to these and other teaching responsibilities in these environments, including long distance learning, are discussed.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

Micheal L. Dent, Chair
SUNY Buffalo, Dept. of Psychology, Buffalo, NY 14260

Chair's Introduction—1:00

Invited Papers

1:05

3pID1. Biomedical photoacoustics: From sensing to imaging to therapy. Stanislav Emelianov (Dept. of Biomedical Eng., Univ. of Texas at Austin, Austin, TX 78712)

Photoacoustics is a wonder as the lightning and thunder once were. Indeed, the absorbed short laser light pulses can generate sound. Utilizing this photoacoustic effect, researchers are now combining photonics with ultrasound technologies, nanoscience, and molecular biology to develop new approaches in molecular imaging and therapy. In photoacoustic imaging, the tissue is irradiated with non-ionizing laser pulses. The deep-penetrating photons are preferentially absorbed within the tissue and, through thermoelastic expansion, the photoacoustic transients are generated. Using acoustic sensors, these photoacoustic waves are then detected at the tissue surface to assess the strength and distribution of optically induced internal sound sources. Photoacoustic imaging can be seamlessly fused with ultrasound imaging, taking full advantage of the many synergistic features of these systems. Furthermore, using bioconjugated nanoparticles with high optical absorption, events at the molecular and cellular levels can be visualized. In this presentation, the basic physics will be described to introduce biomedical photoacoustics. We will then discuss experimental aspects of photoacoustic sensing and imaging including ultrasound/photoacoustic hardware, signal and image processing algorithms, etc. Techniques to synthesize molecular contrast agents will be overviewed. Finally, we will demonstrate and discuss biomedical and clinical applications of the combined ultrasound and photoacoustic imaging ranging from diagnostics to therapeutics.

1:25

3pID2. Remote sensing with ambient noise. Peter Gerstoft (Univ. of Califonia San Diego, 9500 Gillman Dr., La Jolla, CA 92093-0238) and Martin Siderius (Portland State Univ., Portland, OR)

Ambient noise is ubiquitous in the ocean. For many years, this noise has been considered the unwanted part of the acoustic signal. However, recent studies have shown that the noise itself contains valuable information about properties of the ocean, Earth, and atmosphere. For example, distant storms have been observed using measurements of low-frequency (0.1-Hz) noise that has propagated through the Earth's core. Wind speed over the ocean has also been determined hundreds of kilometers away using noise measurements at coastal observing stations. At higher frequencies, surface wave noise due to breaking waves has been used to image the seabed and reveal details of the sub-bottom structure. In general, these techniques are based on cross-correlations of noise signals measured at different locations. These cross-correlations produce an estimate of the Green's function (impulse response) between the two points which can be used to characterize the medium. With concerns over the impact of anthropogenic sound on the marine environment it is not surprising that remote sensing with naturally occurring noise has become a hot topic in acoustical oceanography. Essential components of noise processing will be described along with examples illustrating applications. [Work supported by ONR.]

1:45

3pID3. Hot topics in noise. Nancy S. Timmerman (Consultant in Acoust. and Noise Control, 25 Upton St., Boston, MA 02118-1609, nancy.timmerman@alum.mit.edu)

Noise as a topic is, perhaps, the most politically charged of all areas of acoustics. This paper will review some of the items of current interest. Recent work in this field includes community noise and vibration from transportation (air and highway) and from sustainable energy projects, in particular, wind turbines; use of an audibility criterion within the National Parks; and use of the "soundscape" to get away from the negative connotations of the word "noise." Codes and standards continuing to be developed include recent criteria for "green building" (LEED) certification for hospitals and schools. Cost credit can be taken for reduction in energy use resulting from noise control. Electronic miniaturization has moved recreational "boom box" noise to in-ear monitors, at the possible detriment of the users. These topics will be addressed as time permits.

WEDNESDAY AFTERNOON, 28 OCTOBER 2009 PECAN, 1:15 TO 3:00 P.M.

Session 3pSA

Structural Acoustics and Vibration: Applications of Structural Acoustics and Vibrations II

Wen L. Li, Chair
Wayne State Univ., Dept. of Mechanical Engineering, 5050 Anthony Wayne Dr., Detroit, MI 39759-8270

Contributed Papers

1:15

3pSA1. Extraction of acoustic signals using blind source separation method. Na Zhu and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

This paper presents the results of acoustic feature extraction using blind source separation. The objective of this study is to separate individual acoustic signals based on measurements of overall signals without knowing the number of sources, their locations, and how signals are mixed. The only requirements are (1) the acoustic sources are statistically independent, namely, knowledge of one source gives no information on that of the other, (2) signals are non-Gaussian, and (3) the number of sources is no more than that of microphones. Numerical simulations are conducted based on input data collected by four microphones in a free field, and results are obtained using fast independent component analysis (ICA) in time domain. Different types of acoustic sources, including human speech, music, impulsive sounds, machine noise, helicopter sounds, etc., are used. The impacts of microphone spacing and locations, source locations, signal to noise ratio, sampling rate, and various methodologies such as maximization of non-Gaussianity, maximal likelihood estimation, minimization of mutual information, etc., on separations of individual sources are examined. Results show that different approaches lead to different levels of success in source separation. In general, satisfactory results can be obtained when sources are spatially isolated and reverberation effects are negligible.

1:30

3pSA2. Spatial correlation in a diffuse field generated by random point masses. Wonjae Choi, Jim Woodhouse, and Robin Langley (Dept. of Eng., Univ. of Cambridge, Trumpington St., Cambridge, CB2 1PZ, United Kingdom, wjc24@cam.ac.uk)

This paper considers spatial correlation in a random subsystem. In statistical energy analysis, a subsystem is modeled as a diffuse field, which can be generated by irregular boundaries or scatterers. This paper considers the case where waves are scattered by irregularities inside the subsystem and the boundaries remain deterministic. The uncertainties are parametrized by number and size of random point masses, and the effect on the spatial correlation function between two points in the random subsystem is investigated in the mid-high-frequency range. The scattering effect of one point mass is introduced with point impedance theory, and then the average effect along a wave path is derived so that the statistical loss factor is obtained. In addition, the correlation function is expressed explicitly in terms of a Green function with the aid of the Lagrange–Rayleigh–Ritz method. In the correlation the first few boundary reflections should be included as a minimum in order to cover the whole area of the system, which leads to local variation within the system. The estimation shows good agreement with Monte Carlo simulation of a random plate.

1:45

3pSA3. The Nyquist spatial sampling requirement in nearfield acoustical holography. Richard Dziklinski, III and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., 5050 Anthony Wayne Dr., Detroit, MI 48202)

Numerical simulations of Fourier acoustics and Helmholtz equation least squares (HELS) based NAH to visualize incoherent point sources separated by 6.25 mm are presented. The acoustic pressure measurements are taken over a 50×50 mm^2 plane at different stand-off distances under various signal to noise ratios (SNRs). The considered source frequencies are well above 9 kHz. For Fourier acoustics based NAH, an 11×11 microphone array with 5 mm microphone spacing is used to collect the input data. Results show that Fourier acoustics based NAH encounters great challenges in locating these sources as the frequency and stand-off distance increase and SNR decreases unless regularization is used. On the other hand, a 5×5 microphone array with 12.5-mm microphone spacing is used for HELS under the same conditions without regularization. Clearly, the Nyquist spatial sampling requirement is violated in this case. Nevertheless, if the goal is to identify the locations of discrete airborne sources rather than reconstructing precise sources amplitudes, HELS can still yield satisfactory results. Results show that HELS is capable of locating the discrete airborne sources with a high spatial resolution ($<\pm$ 2.5 mm) when stand-off distance is 10 mm and SNR is 10 dB.

2:00

3pSA4. Helmholtz equation least squares based panel contribution analysis. Sandeep Mylavarapu and Sean Wu (Dept. of Mech. Eng., Wayne State Univ., Detroit, MI 48202)

This paper presents a methodology for analyzing relative contributions from individual panels toward the sound pressure field inside a large enclosure, for example, an aircraft cabin, automobile passenger compartment, etc., due to sound transmission from external acoustic sources, structural vibrations, or their combinations. In this approach, the field acoustic pressure is correlated directly to the normal component of the time-averaged surface acoustic intensity, which is reconstructed based on acoustic pressures measured by microphones in the near field of a structure using Helmholtz equation least squares method. The normal surface acoustic intensities are calculated on individual panels; therefore their relative contributions toward sound radiation at any field point can be evaluated and ranked. This approach is advantageous over those based on a reciprocally measured transfer function [Hald *et al.*, Internoise (2006); Wolff, Internoise (2007)] in that panel contributions are done by taking a single set of measurements, and results are valid for all interior points. To validate this methodology, numerical simulations are conducted for a spherical enclosure for which the analytic solutions are readily available. Specific examples of numerical simulations include a dilating spherical and a partially vibrating sphere. Finally, experimental results using this methodology for a full-size vehicle are demonstrated.

2:15

3pSA5. Continuous-scan vibrometry technique for broadband vibration measurements with high spatial detail. Matthew S. Allen (Dept. of Eng. Phys., Univ. of Wisconsin-Madison, 1500 Eng. Dr., 535 ERB, Madison, WI 53706)

Recent works by the author and others have shown that one can estimate the mode shapes of a structure with remarkable spatial detail in a short time by sweeping a laser vibrometer continuously over the surface of the structure in a periodic fashion. The highly detailed spatial shapes that this method identifies may aid in locating damage or in understanding mismatch between an analytical model and the experimental hardware. Previous works have developed continuous-scan vibrometry methods for transient response measurements and pure sinusoidal excitation. This work treats measurements from structures excited with a measured, broadband forcing function. The approach amounts to a non-parametric estimator for the state transition matrix (analogous to the transfer function) of the time-periodic system. The method is evaluated using simulated measurements from a time-periodic system and using continuous-scan vibrometer measurements from an air conditioner condenser fan.

2:30

3pSA6. Boundary layer pressure fluctuation measurements in transitional and turbulent flow. Thomas A. Galib (NUWC, Newport, RI 02841)

Pressure fluctuations were measured for external flows over an axisymmetric surface. Measurements were made in laminar, transitional, and turbulent flow using piezoelectric pressure transducers. The entire surface (including the pressure transducers) was coated with a 1/16-in. thick elastomer, in order to ensure a smooth surface. All measurements of boundary layer transition were made in an adverse pressure gradient. Results showed that root-mean-square pressure fluctuation levels during transition were as much as 15 dB higher than for fully developed turbulent flow. The transitional spectra were dominated by Tollmien-Schlichting waves, and associated nonlinearities. The spectra of Tollmien-Schlichting waves during amplification and subsequent transformation to turbulence are shown nondimensionally in comparison with Bakewell's turbulence spectral density.

2:45

3pSA7. On modeling the intra-model, infinite seawater acoustic propagation path of undersea vehicles in statistical energy analysis. Robert M. Koch (Chief Technol. Office, Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841-1708, robert.m.koch@navy.mil)

Statistical energy analysis (SEA) modeling has been utilized for decades toward the prediction of both self- and radiated noise behaviors of undersea vehicles such as submarines, torpedoes, and unmanned underwater vehicles. Since these vehicles operate in a semi-infinite or infinite seawater environment, it is difficult in SEA to account for any "intra-model" acoustic propagation through this infinite seawater path. For example, an important source of undersea vehicle self-noise arises due to acoustic propagation through the seawater environment from the vehicle's aft propulsion system to the vehicle's nose where the array is oftentimes located. To date, there has been no standard method published in the literature for how to accurately handle this important acoustic propagation path in SEA. The present paper investigates and compares several approaches that have been employed for properly accounting for aft-to-forward acoustic propagation through this infinite seawater environment path both for fully-wetted and supercavitating, high-speed undersea vehicle SEA models. A new hybrid method employing both finite acoustic volume and semi-infinite SEA acoustic elements has recently shown promise and is presented for discussion.

Session 3pSC

Speech Communication: Fluid-Structure Interaction in Voice Production: Experiments and Modeling II (Poster Session)

Zhaoyan Zhang, Chair
UCLA School of Medicine, 31-24 Rehabilitation Center, 1000 Veteran Ave., Los Angeles, CA 90095-1794

Contributed Papers

All posters will be on display and all authors will be at their posters from 1:30 p.m. to 2:30 p.m.

1:30

3pSC1. Fully-coupled fluid-structure interaction simulation of vocal folds. Lucy Zhang, Xingshi Wang (MANE, RPI, JEC 2049-110 8th St., Troy, NY 12180, zhanglucy@rpi.edu), and Michael Krane (ARL, Penn State, State College, PA 16804-0030)

In this talk, the human vocal folds are modeled and simulated using a fully-coupled fluid-structure interaction method. This numerical approach is efficient in simulating fluid and deformable structure interactions. The two domains are fully coupled using an interpolation scheme without expensive mesh updating or re-meshing. The method has been validated through rigorous convergence and accuracy tests. The response of the fluid affects the elastic structure deformation and vice versa. The goal of this study is to utilize this numerical tool to examine the entire fluid-structure system and predict the motion and vocal folds by providing constant inlet and outlet pressures. The input parameters and material properties, i.e., elastic and density of the vocal folds used in the model, are physiological. In our numerical results, the glottal jet can be clearly identified; the corresponding pressure field distribution and velocity field are presented.

3pSC2. Viscoelastic properties of phonosurgical biomaterials at phonatory frequencies. Miwako Kimura, Ted Mau (Otolaryngol.-Head and Neck Surgery, Univ. of Texas Southwestern Medical Ctr., Dallas, TX 75390), and Roger W. Chan (Univ. of Texas Southwestern Medical Ctr., Dallas, TX 75390)

The purpose of this study was to compare the functional biomechanical properties of injectable biomaterials commonly used for vocal fold augmentation. Viscoelastic shear properties of cross-linked hyaluronic acid (HA) gel (Juvederm), micronized Alloderm (Cymetra), and atelocollagen (3% bovine non-cross-linked collagen) were determined at phonatory frequencies and compared to those of the human vocal fold cover ($n=13$). The elastic and viscous shear moduli ($G\prime$ and $G\prime\prime$) of all samples were quantified with a linear, controlled-strain simple-shear rheometer as functions of frequency, up to the phonatory range (1–250 Hz). Statistical analysis was performed with one-way analysis of variance and Tukey's tests. Results showed that the injectable materials tested were generally stiffer and more viscous than the vocal fold cover. Significant differences in the viscoelastic functions were observed between micronized Alloderm and the vocal fold cover at some phonatory frequencies. Cross-linked HA and atelocollagen were found to be the closest rheological match to the vocal fold cover. These findings have implications for the functional biomechanical performance of these biomaterials for facilitating vocal fold vibration. [Work supported by NIH.]

3pSC3. High-precision frequency analysis of vocal acoustic spectra. Sam Matteson and Fang-Ling Lu (Dept. of Phys., Univ. of North Texas, 1155 Union Circle #311427, Denton, TX 76203-5017, samuel.matteson@unt.edu)

Earlier work by the authors has demonstrated that the primary overtone series of control subjects is remarkably harmonic; that is, the overtones are very nearly integral ratios of the fundamental. Individuals experiencing dysphonia of various kinds exhibited inharmonicities greater than 10 ¢, however. The current work extends the preliminary investigation and examines the origin of other non-harmonic peaks in high-precision acoustic spectra of both control subjects and those complaining of vocal problems. The authors interpret the non-harmonic peaks in terms of the mechanics of the vocal fold-air stream interaction and comment on the potential of these features for diagnosis of pathology.

3pSC4. Frequency response of synthetic vocal fold models with linear and nonlinear material properties. Stephanie Shaw (Dept. of Commun. Disord., Brigham Young Univ., Provo, UT 84602, stephanie.m.shaw@hotmail.com), Simeon Smith, and Scott L. Thomson (Brigham Young Univ., Provo, UT 84602)

Previous studies have shown the importance of cricothyroid muscle activation in altering fundamental frequency of the human voice. Other studies have explored the nonlinear stress-strain properties of human vocal fold tissue and shown that this also influences vibration frequency. Synthetic models have been and are currently being used to characterize behavior of the human vocal folds. To date, the models have used materials with essentially linear stress-strain properties. The purpose of this research study was to investigate the frequency response of synthetic models fabricated with linear and nonlinear materials as a function of anterior-posterior stretching (as in cricothyroid muscle activation). Methods are described for fabrication of both materially-linear and nonlinear vocal fold models. The mechanism for applying anterior-posterior stretching is described. Onset pressure and fundamental frequency data are given for linear and nonlinear models at 1-mm stretch (about 6% strain) intervals up to 10 mm (about 60% strain). Tensile test results show stress-strain properties for both models, and data are compared with similar tests performed using excised human vocal folds. [Work supported by NIH.]

3pSC5. Characterization of acoustic behavior when prandial material is present in the larynx. Shanmugam Murugappan (Dept. of Otolaryngol., Univ. of Cincinnati, 231 Albert Sabin Way, MSB#6308, Cincinnati, OH 45267, mugam@uc.edu), Suzanne Boyce, Sid Khosla, Lisa Kelchner, and Ephraim Gutmark (Univ. of Cincinnati, Cincinnati, OH 45267-0528)

A perceptible change in phonation characteristics after a swallow has long been considered evidence that food and/or drink material has entered the laryngeal vestibule and is on the surface of the vocal folds as they vibrate. In this paper, we investigate the acoustic characteristics of phonation when liquid material is present on the vocal folds, using *in vitro* excised porcine larynges as a model. Two liquids of different Varibar viscosity (nectar and honey) were studied at constant volume. The presence of material on the folds during phonation was generally found to suppress the higher-frequency harmonics and generate intermittent additional frequencies in the low and high ends of the acoustic spectrum, with differences proportional to viscosity. Perturbation measures showed a higher percentage of jitter and shimmer when liquid material was present on the folds during phonation but they were unable to differentiate statistically between the two fluid conditions. The finite correlation dimension and positive Lyapunov ex-

ponent measures indicated that the presence of material on the vocal folds excited a chaotic system. Further, these measures were able to differentiate the two fluid conditions.

3pSC6. Characterization of the airflow through a scaled physical vocal fold model with a unilateral polyp. Angela C. Seawright (Weldon School of Biomedical Eng., Purdue Univ., 206 S. Martin Jischke Dr., West Lafayette, IN 47907, aseawrig@purdue.edu), Byron D. Erath (Purdue Univ., West Lafayette, IN 47907), and Michael W. Plesniak (The George Washington Univ., Washington, DC 20052)

The formation of vocal fold polyps, caused by trauma, interferes with normal speech, causing hoarseness or breathiness during phonation. The objective of this study is to investigate the disruption of the glottal flow field due to the presence of a unilateral vocal fold polyp. An ellipsoidal polyp was molded using silicone and incorporated onto the right medial surface of a driven, 7.5 times physiological size vocal fold model. Particle image velocimetry was used to observe the air flow through the vocal folds along a coronal section at discrete phases of the phonatory cycle. Air flow was scaled to match physiological values of Reynolds, Strouhal, and Euler numbers. Data were acquired in planes bisecting the polyp and others offset anterior and posterior to the center of the polyp, in order to ascertain any three-dimensionality of the disrupted flow. The resulting velocity measurements and transglottal pressure drops were compared to previous results from normal vocal fold motion. Differences in flow characteristics indicating a change in the phonatory capacity of patients with polyps are reported. [Work funded by NSF Grant No. CBET-0828903.]

Plenary Session and Awards Ceremony

Whitlow W.L. Au, Chair
President, Acoustical Society of America

Presentation of Certificates to New Fellows

David L. Adams
Michael A. Averkiou
Constantin C. Coussios
Diane Dalecki
Mostafa Fatemi
Kullervo H. Hynynen
Michael F. Insana
Boris G. Katsnelson
Robert G. Leisure
Philipos C. Loizou
Patrick J. Loughlin
David A. Mann
David K. Mellinger
Ralph T. Muehleisen
Susan N. Nittrouer
Dennis A. Paoletti
Kevin B. Smith
Andrew J. Szeri
Shin-ichiro Umemura
Oleg A. Sapozhnikov

Announcement of the 2009 Munk Award granted jointly by The Oceanography Society, the Office of Naval Research, and the Office of the Oceanographer of the Navy to James F. Lynch

Presentation of Acoustical Society Awards

Science Writing Award in Acoustics for Journalists to Rachel Ehrenberg
"Stranded: A Whale of a Mystery" in *Science News*, July 2008

Science Writing Award for Professionals in Acoustics to Ingo R. Titze
"The Human Instrument" in *Scientific American*, January 2008

Rossing Prize in Acoustics Education to James V. Sanders

Silver Medal in Acoustical Oceanography to Robert C. Spindel

Silver Medal in Noise to Michael R. Stinson

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Acoustical Oceanography

Robert C. Spindel

2009

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Clarence S. Clay	1993
Herman Medwin	1997
D. Vance Holliday	2004

CITATION FOR ROBERT C. SPINDEL

. . . for implementation of ocean acoustic tomography and basin scale acoustic thermometry

SAN ANTONIO, TEXAS • 28 OCTOBER 2009

The first qualities one encounters in Bob Spindel are his intensity, enthusiasm, and drive. Bob is one of those rare people who can power projects, laboratories, meetings, experiments, and even social gatherings by sheer force of personality. People want to work with Bob because it is immediately evident that he will make good things happen. If I had to pick one word to epitomize Bob Spindel, I would have to go with "leader."

Bob is also one of the savviest, sea-going, ocean acoustic experimentalists this country has produced. Formally trained by Peter Schultheiss at Yale as a signal processor, Bob combined his academic background with his "hobbies" to master the experimental art. Specifically, Bob is an avid car mechanic (rebuilding sports cars), a devoted ham radio operator, a diver, and recently he has taken on watercolors.

Bob's accent, easy-going brashness, curiosity, and humorous cynicism mark him immediately as a New Yorker. Born in New York City, and educated in its public schools, Bob took advantage of the best that New York had to offer educationally, graduating in 1965 from The Cooper Union with a degree in Electrical Engineering. Always having a desire to see far-away places, Bob then trekked all the way to New Haven, Connecticut for graduate work, eventually earning his Ph.D. in 1971 in Electrical Engineering, studying ocean acoustic scattering.

George Veronis, an oceanographer who was on Bob's Ph.D. committee, suggested that a postdoc at the Woods Hole Oceanographic Institution (WHOI) would be an interesting beginning to his professional career, so Bob moved to WHOI, where he would spend the next sixteen years. He initially worked on scattering from nekton, long range sonar and the acoustic detection of microearthquakes on the mid-Atlantic Ridge. The latter involved developing a new mode of operation for the navigation system of deep submersible *Alvin*, and would prove to be germane to Bob's long term concentrations–long range acoustic propagation, ocean acoustic tomography, and ocean thermometry.

Woods Hole quickly recognized the exceptional young talent it had attracted in Bob, and appointed him to the scientific staff in the newly formed Ocean Engineering Department in 1972. By 1982, Bob had climbed through the ranks to Senior Scientist and Department Chairman. As a young scientist at WHOI, Bob worked extensively on long range acoustic propagation and positioning and navigation systems, often in close collaboration with Bob Porter. During that time, Bob invented a Doppler navigation system that he was quite proud of, and used it to make synthetic aperture arrays from drifting receivers. Interestingly, all these ocean acoustics skills were just the ingredients needed for the experimental implementation of a new technology, "ocean acoustic tomography," that was conceived in the mid-1970's by Walter Munk of the Scripps Institution of Oceanography and Carl Wunsch of the Massachusetts Institute of Technology. As fate would have it, Walter Munk's senior engineer at Scripps had just retired, and he was looking for an inventive ocean engineer who could help him implement this scheme. Bob was asked to participate, and a collaboration that has lasted decades, and is still strong today, was begun.

The ocean acoustic "cat-scan" developed by Munk and Wunsch, was a grand challenge to realize, absolutely pushing the state of the art in the late 1970's, when the young Robert Spindel led the experimental efforts in the first tomography demonstration experiments. As we know, the early experiments went well, with a huge part of their success due to Bob being in charge of the instrumentation and mooring efforts. About a decade after the initial demonstrations (late 1980's), and then knowing that ocean tomography basically did work, Bob participated in two other major tomography efforts, the Greenland Sea tomography experiment and the moving ship tomography demonstration. Again, both projects worked well, producing very nice images of deep convection and North Atlantic oceanography.

Bob moved to Seattle in 1987 to direct the Applied Physics Laboratory (APL) of the University of Washington. Shortly thereafter, in the early 1990's, tomography was extended from its original conception as an ocean mesoscale imaging tool to a climate

related usage as a large scale ocean thermometer. In addition to his responsibilities as the Director of APL, Bob was a key player in the ATOC experiment (Acoustic Thermometry of Ocean Climate), and its predecessor, the Heard Island Feasibility Test. ATOC succeeded technically, in spite of the fact that the scope of the experiment was severely reduced by concerns about its possible impact on marine mammals. Though tomography is probably what Bob is best known for, his work over the decades has also included many other topics. Bob has published in array and antenna theory, acoustic propagation through a random medium, the use of the "waveguide invariant," instrument, sensor, and platform development, information technology, shallow water acoustics, and Navy applications of acoustics.

As a last category of "career information," I would mention Bob's previous honors. Bob received the A.B. Wood Medal from the Institute of Acoustics (UK) in 1981, the Gano Dunn Medal from his undergraduate school (The Cooper Union) in 1988, the Institute of Electrical and Electronics Engineers Oceanic Engineering Society Distinguished Technical Achievement Award in 1990, the Walter Munk Award from the Oceanography Society in 2001, and the US Navy Meritorious Civilian Service Award in 2003.

I am enormously pleased to see Bob Spindel be awarded the Silver Medal for Acoustical Oceanography, and be honored by the Acoustical Society as a whole. Bob has given much to all of us, on both a technical, societal, and personal level, and it is very gratifying to see his contribution acknowledged in this manner.

JAMES F. LYNCH

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Noise

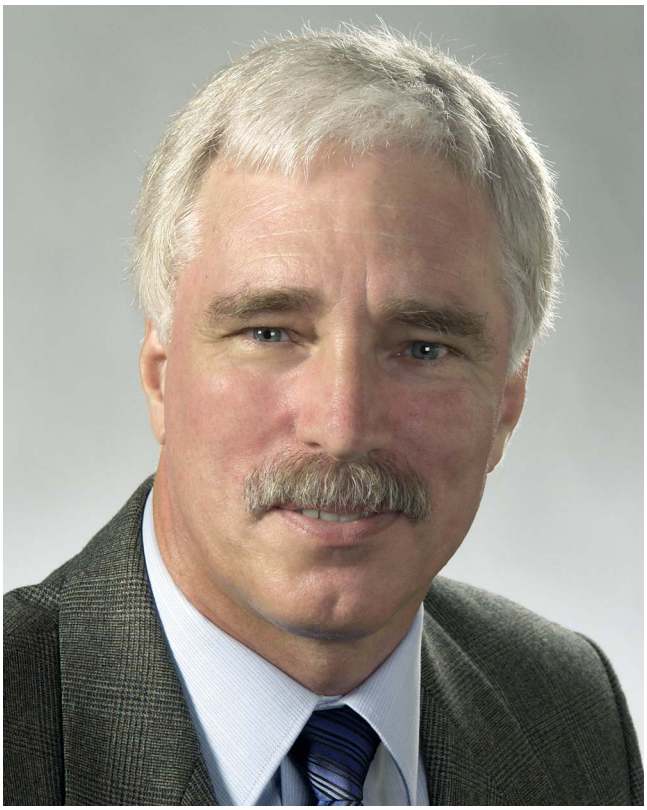
Michael R. Stinson

2009

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

PREVIOUS RECIPIENTS

Harvey H. Hubbard	1978	George C. Mailing, Jr.	1992
Henning E. von Gierke	1981	Kenneth M. Eldred	1994
William W. Lang	1984	Larry H. Royster	1999
Tony F. W. Embleton	1986	Louis C. Sutherland	2002
William J. Galloway	1988	Alan M. Marsh	2006



CITATION FOR MICHAEL R. STINSON

. . . for contributions to outdoor sound propagation, acoustical materials, and ear canal acoustics.

SAN ANTONIO, TEXAS • 28 OCTOBER 2009

Michael R. Stinson is recognized for his contributions to modeling the ear canal, the acoustics of porous materials, outdoor sound propagation, beamforming in reverberant environments and on diffracting surfaces, and, more recently, hearing aid feedback.

Mike received B.Sc. and M.Sc. degrees in physics from Simon Fraser University, British Columbia, in 1971 and 1973, respectively, and a Ph.D. in physics from Queen's University in 1979. He joined the Acoustics Section of the National Research Council (NRC) that same year where he served as Group Leader for the Acoustics & Signal Processing group of the Institute for Microstructural Sciences (IMS) from 2000 to 2009. He currently holds the title of Researcher Emeritus with IMS.

During Mike's initial years at NRC, he was the first to seriously consider the role of the complex vibratory pattern of the tympanic membrane (eardrum), as well as the angle of inclination in the ear canal, on sound power transmission. Important as this is, he is perhaps best known in the field of hearing for his work with Shyam Khanna of Columbia University on ear canal acoustics. Mike's research clearly showed how the sound pressure varies along the length of the ear canal and how this information can be used to estimate the acoustic input to the ear. In this fundamental series of publications, Mike applied relatively simple concepts from linear acoustics to the rather complex geometry of the ear to develop methods for carefully measuring the sound power input to the ear. These contributions are indeed fundamental and have clearly advanced research in hearing. Techniques based on Mike's work, or a closely related Thevenin equivalent approach, are still widely used.

Mike also took an interest in modeling the acoustical behavior of sound tubes, constrictions, resistances, and cavities. His published research results led to the design of some very successful telephony products that are still widely manufactured and still represent the best performance available. Mike's interest then shifted to the study of porous acoustical materials, for which he collaborated with Yvan Champoux of the Université de Sherbrooke. This research involved theoretically modeling and performing experiments with pores of arbitrary shapes. New apparatus was developed to measure the flow resistance, porosity, and tortuosity of porous materials. Prior to this effort, researchers had considered mostly idealized models with slit or circular cylindrical pores. Mike's work helped the community to better understand the role of arbitrary pore shapes and other realistic aspects of porous materials. In particular, his contributions showed the approximations inherent in theories for the acoustical properties of porous media based on pore microstructures and the general relationship between the formulations for viscous and thermal effects.

Beginning in the early 1990's Mike and his colleagues at the National Research Council performed particularly innovative research on impedance of ground surfaces, scattering of sound by atmospheric turbulence, and amplification of sound by surface waves. The research on scattering dealt with coherence of signals in refractive shadow zones across broad frequency ranges. This work went well beyond the classical, line-of-sight scattering theory prevailing at the time, and in the process exhibited remarkable insight into the issues affecting performance of systems for locating outdoor sound sources. The research performed by Mike and his NRC colleagues on surface-wave amplification generated strong interest in the international scientific and defense research communities.

More recently, Mike has returned to hearing research, working with Gilles Daigle to examine performance targets for reducing hearing aid feedback. This research included numerical simulation to determine the effect of handset proximity on the open-loop transfer function and an examination of techniques, such as introducing surface acoustic admittance to the handset, which might reduce the feedback threshold. The research was further expanded to model and measure the sound field in model ear canals occluded by a hearing aid. The measurements led the development of a probe microphone tip having an inner diameter of 0.1 mm and an outer diameter of 0.2 mm!

Mike has been very active in a wide variety of roles for the ASA and other acoustical organizations. He has been chair of the Technical Committee on Noise (2003-2006), Associate Editor of the *Journal of the Acoustical Society of America* for Noise and its Effects and Control (1997-2002), Technical Program Chair of the Ottawa meeting in 1993, member of the Medals and Awards Committee (2000-2003) and of the Publication Policy Committee (1993-2001). Mike continues to serve as Associate Editor for the *Journal of the Acoustical Society of America Express Letters*. He was elected Member of the Executive Council (2007-2010) and is the chair of the joint meeting of the International Congress on Acoustics, the Spring 2013 meeting of the Acoustical Society of America, and the Spring 2013 meeting of the Canadian Acoustical Association to be held in Montreal. He also serves as a member of the Board of the International Commission for Acoustics (2007-2010).

Mike Stinson is a soft-spoken leader in the field of acoustics who has repeatedly demonstrated tremendous technical prowess and a commitment to advancing knowledge and its application. He always exhibits a high level of scientific integrity, and has a selfless record of service, collegiality, and mentoring that is greatly admired by those who have had the privilege to work with him. Mike is an ambassador for the field and an extremely meritorious choice for the Silver Medal in Noise.

GILLES A. DAIGLE
D. KEITH WILSON

Open Meetings of Technical Committees

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held starting immediately after the Social Hours at 7:30 p.m. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings including proposals for special sessions, workshops and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

The Technical Committee on Signal Processing in Acoustics will meet tonight in Rio Grande West.

Session 4aAAa

Architectural Acoustics: The Technical Committee on Architectural Acoustics Vern O. Knudsen Distinguished Lecture

David S. Woolworth, Chair
Oxford Acoustics, Inc., 356 CR 102, Oxford, MS 38655

Chair's Introduction—9:00

Invited Paper

9:05

4aAAa1. Applying Lindsay's acoustical wheel to architectural acoustics. Richard Talaske (Sound Thinking, 1033 South Boulevard, Oak Park, IL 60302, rick@talaske.com)

The acoustic design of performance spaces has advanced over the decades due to the influence of many acoustics disciplines. Psychological acoustics research has helped establish criteria, allowing acousticians to fashion designs which satisfy listeners. Speech perception has advanced within drama spaces, thereby offering more artistic freedom. Structural acoustics and vibration methods have helped facilitate on-stage communication via tactile vibration. The effects of noise on the live event have become better understood, leading to a more satisfying performance experience. And audio systems have been developed, both indoors and outdoors, to embellish architecturally created sound environments. This discussion on architectural acoustics recognizes this confluence of diverse perspectives on acoustics and describes how a broad knowledge base has been channeled into the design of performance facilities for the benefit of patrons and performers alike. The evolution of the design of performance and assembly spaces will be discussed from the perspectives of acoustics, function, and aesthetics. The historical influences of society and technology will be presented as the basis to suggest a glimpse of what performance spaces might be like in the future.

Session 4aAAb

Architectural Acoustics, Noise, and Structural Acoustics and Vibration: Seismic Restraint, Structural Issues, and Building Acoustics

Matthew V. Golden, Chair
Kinetics Noise Control, 6300 Irelan Place, Dublin, OH 43017-0655

Chair's Introduction—10:20

Invited Papers

10:25

4aAAb1. Large building isolation with seismic restraint. James Phillips (Wilson, Ihrig & Assoc., Inc., 5776 Broadway, Oakland, CA 94618)

Isolating a large building such as a concert hall or multi-story condominium necessarily requires structurally separating the isolated building from the foundation. The structural support is then provided by resilient bearings which must properly support the building gravity load and provide seismic restraint and structural stability. The best acoustical results are achieved with natural rubber bearings sized to support the gravity load at columns and loadbearing walls. Lateral support for wind and seismic loads is then achieved, without compromising the isolation, through the installation of lateral support bearings in equal and opposite pairs in both directions. The lateral loads are transmitted to the columns or loadbearing walls via the structural floor above the bearings functioning as a transfer slab. The equal and opposite preloaded pair concept for lateral restraint was developed to achieve an assembly with the bearings always in compression regardless of load or deflection. This avoids any need for embedded plates or bolts which could cause tension in the concrete structural elements. The history of the development of the preloaded lateral restraint rubber bearing concept is presented along with examples of successful application to concert halls and other structures.

10:45

4aAAb2. The impact of structural design decisions on building acoustics. Scott D. Campbell (Structural Anal. Consulting Group, P.O. Box 91364, Louisville, KY 40291, scott@str-analysis.com)

Choices made by the architect and structural engineer at all phases of the design process affect the acoustic and vibration characteristics of a building. The basic structural system locks into place acoustic and vibration properties that are very difficult to change after the fact. Even when a favorable structural system is selected, details such as floor slab design and connection type can adversely affect the acoustic and vibration performance. Current trends toward less expensive structural systems, and hence lighter and more flexible systems, reduce the ability of a building to resist transmission of sound and vibration. In addition, failure to involve an acoustic or vibration consultant early in the process makes it much harder to correct these problems. All of these issues are compounded when the building is located in a high seismic or wind area. The special design requirements are often at odds with proper acoustic and vibration design. This paper will explore the impact of typical structural design options on the acoustic and vibration properties of buildings, and discuss some of the remedies available for mitigating the impact.

11:05

4aAAb3. Acoustical performance of wall assemblies containing wood structural elements. John LoVerde and Wayland Dong (Veneklasen Assoc., 1711 16th St., Santa Monica, CA 90404, jloverde@veneklasen.com)

There has been little historical research within the acoustical community regarding the sound isolation properties of wall assemblies that include structural elements. On most buildings, many of the wall assemblies, including exterior walls, demising walls, and corridor walls, require wood shear panels and/or posts to be incorporated into their design for structural reasons. However, most common acoustical designs have not been acoustically tested with the presence of wood structural elements. Acoustical test results for such wall assemblies are presented. Since a conventional approach is to assume that a plywood shear panel performs similarly to a layer of gypsum board, the walls including a wood shear panel will be compared to equivalent assemblies where that layer is gypsum board. Additionally, field acoustical test results for walls including shear panels shall also be presented and compared with the laboratory tests, and translation of laboratory to field acoustical performance for these constructions will be discussed.

11:25

4aAAb4. Best practice for applying seismic restraint to vibration isolated equipment. James N. Pooler (Kinetics Noise Control, 6300 Irelan Plc, Dublin, OH 43017, jpooler@kineticsnoise.com)

In many areas of the county and throughout the world, seismic restraint is required for mechanical equipment per local building codes and The International Building Code. If seismic restraint is not properly designed or properly installed, vibration isolation efficiency could be affected. Properly designed seismic restraint systems allow for vibration isolation while effectively transmitting the seismic loads into the building structure. This presentation will also use case studies to review the issues and solutions for applying seismic restraint to vibration isolated mechanical equipment.

THURSDAY MORNING, 29 OCTOBER 2009 RIO GRANDE WEST, 8:00 A.M. TO 12:00 NOON

Session 4aAB

Animal Bioacoustics: Natural Soundscapes and Auditory Scene Analysis by Animals I: Natural Soundscapes

Rama Ratnam, Chair
Univ. of Texas at San Antonio, Biology Dept., One UTSA Cir., San Antonio, TX 78249

Chair's Introduction—8:00

Invited Papers

8:05

4aAB1. The soundscape of a nearshore reef near an urban center. Whitlow W. L. Au and Michael Richlen (Hawaii Inst. of Marine Biology, Univ. of Hawaii, 46-007 Lilipuna Rd., Kaneohe, HI 06744)

An acoustic monitoring system in the nearshore waters of the south shore of Oahu is a part of the Kilo Nalu nearshore coral reef observatory (http://www.soest.hawaii.edu/OE/kilonalu/). The underwater housing containing the hydrophone, amplifier-filter, A/D converter, and an ethernet converter is located 690 m from shore at a depth of 22 m. Acoustic signals are digitized at a rate of 24 kHz, and sent to shore on a fiber optic link to a laptop computer and are stored on a disk drive. Snapping shrimp sounds are the most prevalent biological signals and are always present. Other sounds from biological organisms include humpback whales (during their winter season in Hawaii), fish, and dolphins. However, the soundscape during daylight hours is dominated by small boat sounds as a variety of vessels are regularly detected and the signal levels from these boats are considerable higher than the biological sounds. Only humpback whale songs have the chance of being detected in the presence of boat sounds, whereas other biological sounds tend to be overwhelmed. During the day, sounds from other biological organisms can only detected between boat sounds. Boat sounds usually stop after dusk providing the best environment for detecting biological sounds.

8:25

4aAB2. Bird hearing and communication in natural environments. Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

A soundscape, taken from human work, generally refers to the complex of sounds that are integral to and formed by a particular natural environment. Some of this sound is from abiotic sources (e.g., wind, streams, various anthropogenic sources, etc.) while some of this sound may come from conspecifics or other animals. In nature, all sound reaching an animal's ears, whether from biotic or abiotic sources, is shaped by this natural environment. Noise, sound distortion, and frequency-dependent attenuation can increase detection thresholds and interfere with discrimination, recognition, and localization of sound sources. Signals degraded by travel through the environment and/or affected by masking can in turn adversely affect critical behaviors such as mate attraction, territorial defense, parent-offspring communication, predator avoidance, and prey detection. Using laboratory psychophysical data from several species of birds, some interactions among these factors and the implications for communication in a noisy environment are described. Species differences in auditory sensitivity and the characteristics of the environment are major factors in predicting whether a given soundscape limits communication between birds or interferes with detection of other biologically important sounds. [Work supported by NIH grant to R.J.D.]

8:45

4aAB3. Characterizing variation in natural acoustical environments and assessing the consequences of this variation for animal perception and communication. Kurt M. Fristrup (Natl. Park Service, Natural Sounds Program, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

Ambient sound levels present fundamental constraints for auditory perception. An extensive archive of sound level data, 1-s Leq values in 1/3 octave bands, has been collected by the National Park Service to assess conditions across a wide variety of park units and environmental settings. Exceedance values for these data offer one approach to assessing variation, but this ignores the temporal sequence of the measurements. In what sense are sound level patterns "stationary"? This presentation will evaluate several measures of temporal variation in sound levels and evaluate the results in relation to the diel schedule and behavioral tempo of animal activities.

9:05

4aAB4. Non-random patterns of acoustic interactions in chorusing bullfrogs. Andrea M. Simmons, Mary E. Bates, and Jeffrey Knowles (Dept. of Psych., Brown Univ., Providence, RI 02912, andrea_simmons@brown.edu)

Male bullfrogs form choruses to vocally advertise for females and to announce territory occupation to rival males. In dense choruses, calls of individual males may temporally overlap due to the large numbers of vocalizing neighbors. Overlapping calls may be a deliberate communicative strategy, perhaps to form a more salient auditory object in order to effectively guide females to a particular location within the chorus. We used a custom-written MATLAB program to simulate the proportions of overlapping and nonoverlapping calls in five mock choruses, and then compared model output to empirical data from natural choruses of equal size. The simulation assumed that each bullfrog called independently of his neighbors according to a Poisson process so that overlapping calls occur randomly rather than through cooperation. In four of the five empirical recordings, the number of overlapping calls was significantly smaller than the averages produced by corresponding simulations, indicating that more males vocalized together in bouts than predicted by the random simulation. Empirical and simulated data also differed significantly in the remaining chorus, but in the opposite direction. These data suggest that overlapping calls may confer some communicative advantage. [Work supported by NSF Advance program.]

9:25

4aAB5. Dynamical interactions in a green treefrog chorus. Douglas L. Jones (Dept. of Elec. & Comput. Eng., Univ. of Illinois at Urbana-Champaign, 1308 W. Main St., Urbana, IL 61801), Russell L. Jones (Univ. of Illinois at Urbana-Champiagn, Urbana, IL 61801), and Rama Ratnam (Univ. of Texas at San Antonio, San Antonio, TX 78249)

In southeast Texas male green treefrogs (*Hyla cinerea*) assemble in dense choruses and call for several hours every night to attract females. The call is a short "quank" lasting about a 100 ms, with a variable call interval of about 400 ms. Individuals call from dense vegetation surrounding the breeding area, occupying positions that range in elevation from water level to several meters above the ground. The high call-rate and duty cycle combined with their three-dimensional spatial distribution present challenging problems for signalers and receivers and for human observers. It is qualitatively apparent that within a large natural chorus, individuals dynamically adjust their calling to avoid overlap with other callers. However, the dynamical interactions and calling strategies have not been quantified. Here, a 15-microphone array was deployed within the breeding site in the vicinity of one group of callers, and the locations and call patterns of individuals were extracted by acoustic beamforming. The results show that an individual pays attention to only a few of the neighboring males and times his calls to avoid call collision. When a collision occurs one of the individuals will delay the subsequent call by approximately half a call interval to avoid further collision.

9:45

4aAB6. A glimpse of the world through the voice of the echolocating bat. Cynthia Moss (Dept. of Psych., Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742), Chen Chiu (Univ. of Maryland, College Park, MD 20742), and Annemarie Surlykke (Univ. of Southern Denmark, 5230 Odense, Denmark)

Bats echolocating in the natural environment face the formidable task of sorting echoes from closely spaced obstacles and prey, which may be mixed with sonar signals of nearby conspecifics. This talk will review behavioral strategies used by the big brown bat, *Eptesicus fuscus*, to solve such auditory scene analysis tasks. The sonar beam emission pattern of the big brown bat is directional, but broad enough to collect echo information from objects within a 60–90 deg cone, which would enable simultaneous inspection of adjacent objects. Data show, however, that bats shift their sonar gaze to inspect closely spaced objects sequentially and analyze sounds

source direction and distance to track auditory objects as they fly. When bats forage in groups, they face the challenge of sorting echoes from their own calls from echoes and calls of neighboring bats. Research findings demonstrate that bats make adjustments in the spectral characteristics of calls when flying with conspecifics, and the magnitude of these changes depend on the baseline similarity of calls produced by individual bats when flying alone. These recent discoveries from studies of bat echolocation point to the general importance of an animal's action to its analysis of natural scenes.

10:05—10:20 Break

10:20

4aAB7. Bats control the auditory scene by adapting intensity and directionality of echolocation calls. Annemarie Surlykke, Lasse Jakobsen, Signe Brinkloev (Inst. Biology, SDU Univ. Southern Denmark, Odense, DK-5230 Denmark), and Cynthia Moss (Univ. of Maryland, College Park, MD 20742)

Echolocating bats emit short high-frequency calls and use returning echoes to determine direction, distance, and features of objects in the environment. Thus, echolocation is an active orientation system allowing bats to operate with little or no light. In open areas in the wild, bats emit very high intensity search calls of high directionality to increase the range of the sonar. In cluttered habitats (or the laboratory) not only the prey but also vegetation and other obstacles will reflect a fusion of many echoes for each emitted echolocation call. In cluttered habitats, the intensity is lower and the beam is wider to shorten and broaden the angle of view, while the beam axis is still focused very accurately on the target. Thus, by adjusting intensity, directionality, and beam aim of their calls, bats can actively filter and control the "sound-picture" of their surroundings that they perceive through echolocation.

10:40

4aAB8. Defocusing of echo-delay images for clutter rejection in echolocating bats. Mary E. Bates (Dept. of Psych., Brown Univ., Box 1853, 89 Waterman St., Providence, RI 02912) and James A. Simmons (Brown Univ., Providence, RI 02912)

Echolocating big brown bats emit ultrasonic FM sounds containing two prominent harmonics (FM1 and FM2) and perceive target distance from echo delay. Each broadcast and its echoes are processed by time-domain mechanisms incorporating neuronal delay-lines for spectrogram correlation. Pulse-to-pulse variability requires the bat to store a template of the FM sweeps in each broadcast to recognize its exact echoes. Ordinarily, gradations of mismatch, or decorrelation, between echoes and broadcasts would lead to a correspondingly graded decline in delay accuracy. However, when the relation between FM1 and FM2 is gradually disrupted, bats abruptly lose delay acuity through a process that amounts to deliberate defocusing of the image. Bats actually have two computational pathways for perception of delay—one involving overall delay to determine target range, and the other involving delay differences between glint reflections to determine shape, extracted from the interference spectrum between the overlapping reflections. Defocusing occurs in the shape, or fine delay part of the image, possibly due to failure of the spectrally based transformation process to converge upon a simple solution that would depict a manageably small number of glints from a regular pattern of interference notches. [Work supported by ONR and NIMH.]

Contributed Papers

11:00

4aAB9. The effect of different schedules of reinforcement on the structure of the black-capped chickadees' chick-a-dee call. Darren S. Proppe (Dept. of Biological Sci., Univ. of AB, Edmonton, AB T6G 2E9, Canada proppe@ualberta.ca) and Christopher B. Sturdy (Univ. of Alberta, Edmonton, AB T6G 2E9, Canada)

It is known that adult songbirds can modify the output and structure of their vocalizations in different behavioral contexts. However, whether changes in the abiotic environment (i.e., food availability) also affect vocal structure is less understood. Here we test whether the chick-a-dee calls of the black-capped chickadee differ between birds on two different schedules of reinforcement (continuous and partial) and in two different behavioral contexts (singing alone and mildly alarmed by human intrusion). Our results reveal differences in the composition and pitch of chick-a-dee calls from birds in different behavioral contexts and on different schedules of reinforcement. However, calls from birds on the different schedules of reinforcement become more similar when evoked by human intrusion. We suggest that these results may represent differences in the energy dedicated to call production. If vocal structure varies consistently with changes to the abiotic environment, we may be able to detect environmental changes (i.e., warming and reduction in available forage) through analysis of vocal traits. While we demonstrate differences in call structure based on schedules of reinforcement, results from field studies in multiple regions will divulge whether these vocal changes are consistent and useful for analysis of abiotic conditions.

11:15

4aAB10. Understanding the relationship between marine mammals and their environment in the Bering Sea. Jennifer L. Miksis-Olds, Susan E. Parks (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, jlm91@psu.edu), and Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105)

Passive aquatic listening (PAL) instruments were deployed on subsurface moorings along the 70-m isobath of the Bering Sea shelf. PALs operate according to an adaptive sampling protocol as opposed to recording continuously, thus enabling data collection for a full year at a sampling rate of 100 kHz. Year long time series indicate that marine mammal vocalizations dominate the regional soundscape in the late winter/early spring time period. Species acoustically identified include North Pacific right, killer, beluga, humpback, fin, and bowhead whales, plus bearded seal and walrus. Synoptic measurements of prey abundance with three-frequency active acoustic sensors, ice cover, and physical oceanographic processes provide vital information for understanding the relationship between marine mammals and their continually changing environment. These data represent important baseline measurements of environmental variability, marine mammal species richness, and temporal distribution of animals in an area of the world that is rapidly transforming due to global climate change. [Research supported by ONR.]

11:30

4aAB11. Killer whale discrete pulsed call variation. Dawn M. Grebner, David L. Bradley, Dean E. Capone, Susan E. Parks, Jennifer L. Miksis-Olds (Grad. Prog. in Acs., Penn State Univ., State College, PA 16804, dmg302@psu.edu), and John K. B. Ford (Pac. Biol. Station, Nanaimo, BC V9T 6N7, Canada)

Animals can increase and diversify vocal complexity and call content by altering internal features within a call. The Northern Resident orca pods, residing off British Columbia, each have their own dialect of structurally discrete and highly stable pulsed calls. The objective of this study is to determine if there are distinctive internal acoustic features within the defined envelope of a single discrete pulsed call (N04) which could potentially relay the signaler's behavioral circumstance. Orca discrete pulsed calls are highly complex with varying time-frequency slopes and multiple sidebands. This analysis began with the parsing of the N04 call into different subtypes based on distinctive changes in time-frequency slopes found in the call spectrograms. Call subtypes were verified using discriminant analysis, and changes in slope trends (ascending, descending, or constant frequency) at designated locations along the calls were compared. Variations in slopes were found between subtypes predominantly in the calls' front and terminal regions. Clear and reliable acoustic cues within a discrete pulsed call could not only provide receivers with the physical location and group affiliation of the signaler but also would alert receivers to the signaler's behavioral state or prey catch which would be vital information for a prey-sharing species.

11:45

4aAB12. Environmental assessment of offshore wind power generation near Rhode Island: Acoustic and electromagnetic effects on marine animals. James H. Miller, Gopu R. Potty, Kathleen Vigness Raposa, David Casagrande, Lisa Miller, Steven E. Crocker, Robert Tyce, Jonathan Preston, Brian Roderick (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett, RI 02882, miller@uri.edu), Jeffrey A. Nystuen (Univ. of Washington, Seattle, WA 98105-6698), and Peter M. Scheifele (Univ. of Cincinnati, Cincinnati, OH 45267)

An offshore wind farm is planned for Rhode Island coastal waters. The developer has proposed to deploy wind turbines in two stages: 5 turbines in shallow waters 5 km south of Block Island and 100 turbines in deeper waters 30 km to the east. As part of the planning of the proposed offshore wind power generation project under the Rhode Island Special Area Management Plan, ambient acoustic and electromagnetic measurements were made in the area. Two passive acoustic listener (PAL) systems were deployed within 4 km of Block Island from October 6 to November 11, 2008. Data from the PALs were used to compute the ocean acoustic noise budget and other statistics by source. Transmission loss measurements were also made to support the noise budget calculation. Measurements of airborne noise from a 1.5-MW land-based wind turbine already in operation in Rhode Island were made. To support the electromagnetic effect study, an underwater magnetometer was towed at the two proposed sites and over an operational underwater 23-kV power cable. A preliminary assessment of the effects of the offshore wind farm on marine animals at these sites will be presented. [Funding provided by the RI Office of Energy Resources.]

THURSDAY MORNING, 29 OCTOBER 2009 REGENCY EAST 2, 8:25 A.M. TO 12:00 NOON

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration, Physical Acoustics and Underwater Acoustics: Multiple Scattering of Waves: From Theory to Application I

Guillaume Haiat, Cochair
CNRS, LRO 10, Ave de verdun, 75010 Paris, Cedex, France

Purnima Ratilal, Cochair
Northeastern Univ., Electrical and Computer Engineering, 360 Huntington Ave., Boston, MA 02115-5000

Chair's Introduction—8:25

Invited Papers

8:30

4aBB1. Multiple scattering effects on communication and imaging through random media. Akira Ishimaru, Sermsak Jaruwatanadilok, and Yasuo Kuga (Dept. of Elec. Eng., Univ. of Washington, Box 352500, Seattle, WA 98195, ishimaru@u.washington.edu)

Communication through turbulence and particulate matters such as rain are seriously affected by multiple scattering. We present a study on multiple-input-multiple-output system in random medium and its channel capacity making use of the stochastic Green's functions and the mutual coherence function (MCF). The channel matrix and the eigenvalues are given explicitly in terms of the medium characteristics, and the channel capacity is given in terms of signal-to-noise ratios, eigenvalues, and the antenna gain characteristics. MCF is given in terms of the power spectrum for turbulence and the optical depth, phase function, and albedo for particulate matters. Next, we consider imaging through random medium such as biological media. We present several space-time array signal processing imaging techniques which include time-reverse imaging, time-reverse multiple signal classification, Capon minimum variance, modified beam former, and SAR. We discuss the advantages and the disadvantages of these techniques with numerical examples. The effects of array sizes, Fresnel size, bandwidth, the distance, and the medium characteristics on the transverse and longitudinal resolutions are clarified. This study is aimed at combining the propagation and scattering and the array signal processing for communication and imaging in a random complex environment such as turbulence and biological media.

9:00

4aBB2. Anderson localization of ultrasonic waves in three dimensions. John Page, Anatoliy Strybulevych, Hefei Hu (Dept. of Phys. and Astronomy, Univ. of Manitoba, Winnipeg, MB R3T 2N2, Canada), Sergey Skipetrov, and Bart van Tiggelen (Univ. Joseph Fourier and Laboratoire de Physique et Modélisation des Milieux Condensés, CNRS, 38042 Grenoble, France)

There is currently a resurgence of interest in Anderson localization, where multiply scattered waves become "trapped" through interference in a very strongly scattering medium. This interest is fueled by theoretical and experimental advances, especially for classical waves, where unambiguous experimental evidence for localization in three dimensions has remained elusive until recently. In this talk, progress in demonstrating the localization of ultrasound in a "mesoglass," made by assembling aluminum beads into a three-dimensional (3-D) elastic network, is summarized. Measurements of the time-dependent transmission of the ultrasonic intensity, as well as the dynamic transverse confinement of the waves due to localization, are discussed. The data are well described by a new self-consistent theory of the dynamics of localization, providing an important validation of this theoretical approach and enabling the localization length to be measured. Finally, evidence for non-Gaussian statistics of the transmitted intensity is discussed, consistent with values of the Thouless conductance g less than 1. This is the first time that these three different fundamental aspects of Anderson localization (time-dependent transmission, transverse confinement of the waves, and statistics) have been studied simultaneously, providing very convincing evidence for 3-D localization of ultrasound in these materials.

9:25

4aBB3. Numerical study of a spectral approach to solve integral equation formulation of the high-frequency acoustic multiple scattering problem by disks. L. Xavier Antoine, Karim Ramdani, and Bertrand Thierry (Institut Elie Cartan Nancy (IECN), Universit Henri Poincaré, Nancy 1, Equipe INRIA Corida, Bureau 307, B.P. 239, F-54506 Vandoeuvre-lès-Nancy Cedex, France, xavier.antoine @iecn.u-nancy.fr)

This aim of this talk is to propose a new numerical method for solving the multiple scattering problem of an acoustic wave by circular cylinders. The situation where the wavelength is small compared to the characteristic size of the cylinders is analyzed in detail here. It is known that this framework implies specific numerical difficulties since it leads to the numerical solution of a large size linear system. The proposed approach is based on computing the scattered field as the solution of an integral equation. Writing this integral equation in a Fourier basis and using a suitable truncation of these Fourier series and a preconditioned Krylov iterative solver (GMRES), an efficient and robust numerical procedure is built. Computational simulations will be provided for solving high frequency multiple scattering problems by random configurations of circular obstacles.

9:50

4aBB4. The effective wavenumber equation for an elastic medium containing cylindrical scatterers. Andrew Norris (Dept. of Mech. and Aerosp. Eng., Rutgers Univ., Piscataway, NJ 08854) and Jean-Marc Conoir (UPMC Univ. Paris 06, F-75005 Paris, France)

We will discuss effective wavenumber equations for waves in an elastic solid containing randomly distributed inclusions in a half-space. The multiple scattering approach used is similar to the one proposed by Fikioris and Waterman for acoustic waves, and will be reviewed in that context. The characteristic equation, the solution of which yields the effective wavenumbers of coherent elastic waves, is obtained in an explicit form without the use of any renormalization methods. Two approximations are considered. First, formulas are derived for the effective wavenumbers in a dilute random distribution of identical scatterers. These equations generalize the formula obtained by Linton and Martin for acoustic waves. Second, the high-frequency approximation is compared with the Waterman and Truell approach for elastic waves. The Fikioris and Waterman approach, in contrast with Waterman and Truell's method, shows that *P* and *SV* waves are coupled even at relatively low concentration of scatterers. Reflection coefficients of *P* and *SV* waves incident on the interface of the half-space containing randomly distributed inclusions are also discussed. These expressions depend on frequency, concentration of scatterers, and the two effective wave numbers of the coherent waves propagating in the elastic multiple scattering medium.

10:15—10:30 Break

Contributed Papers

10:30

4aBB5. Experimental study of the resonant motion of biological cells. Youngsoo Choi, Won-Suk Ohm (Dept. of Mech. Eng., Yonsei Univ., 262 Seongsanno, Seodaemun-gu, Seoul 120-749, Korea), and Yong Tae Kim (Korea Res. Inst. of Standards and Sci., Daejeon 305-340, Korea)

The two lowest natural modes of vibration of a spherical cell suspended in a host fluid have been known to be quadrupole and octupole oscillations [Zinin *et al.*, Phys. Rev. E **72**, 061907 (2005)], although no experimental confirmations have been reported to date. In this paper, an attempt to ultrasonically excite and measure the quadrupole oscillation of a spherical cell is described. Cells suspended in an ultrasonic standing wave field are put into resonance by a series of shock pulses, and the resulting acoustic emission is measured by a needle hydrophone. Both the characteristic frequency and the overall radiation pattern of the observed acoustic emission serve as evidence of the quadrupole cell motion.

10:45

4aBB6. Analysis of human fibroadenoma using three-dimensional impedance maps. Alexander Dapore (Bioacoustics Res. Lab., Univ. of Illinois at Urbana-Champaign, 405 North Mathews Ave., Urbana, IL 61801, adapore2@illinois.edu), Lauren A. Wirtzfeld, Michael R. King, Sandhya Sarwate (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801), Josephine Harter (Univ. of Wisconsin-Madison, Madison, WI 53706), Michael L. Oelze (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801), Timothy J. Hall (Univ. of Wisconsin-Madison, Madison, WI 53706), Minh N. Do, and William D. O'Brien, Jr. (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Three-dimensional impedance maps (3DZMs) are virtual volumes of acoustic impedance values constructed from histology to represent tissue microstructure acoustically. From the 3DZM, estimations can be made for ultrasonic backscatter and scatterer properties, such as effective scatterer diameter (ESD). Additionally, the 3DZM can be exploited to visualize and identify possible scattering sites, which may aid in the development of more

effective scattering models to better represent the ultrasonic interaction with underlying tissue microstructure. In this study, 3DZMs were created from several human fibroadenoma samples. ESD estimates were obtained using the fluid-filled sphere form factor model. These estimates were made using two regions of interest (ROIs) sizes: cubes of side lengths of 300 and 150 μm. This estimation technique allowed a better understanding of the spatial distribution and variability of the estimates throughout the volume. For a collection of ten 3DZMs, the ESD was estimated to be 94 ± 47 μm with the large ROI and 52 ± 19 μm when using the small ROI. The 3DZMs were then investigated visually to identify possible scattering sources. This visualization and comparison resulted in the identification of possible ultrasonic scattering sources within human fibroadenomas. [Work supported by NIH Grant CA111289.]

11:00

4aBB7. Simulation of ultrasonic scattering in breast tissue based on cell and tissue morphology. Timothy E. Doyle, Vern Hart, and Brady Ambrose (Dept. of Phys., Utah State Univ., 4415 Old Main Hill, Logan, UT 84322, tim.doyle@usu.edu)

High-frequency ultrasonic spectroscopy is a promising approach for the *in vivo* detection of cancer in applications such as margin assessments during breast cancer surgery. To advance this approach, a multiple scattering model was developed to more fully understand and analyze ultrasonic backscatter data in the 10–100-MHz region from breast tissues in various stages of tumor progression. The model uses an iterative multipole approach to simulate the multiple scattering of shear and longitudinal waves between nucleated cells in a histologically accurate tissue representation. Previously, concentric spheres were used to model the nucleus and cytoplasm of cells organized in complex, hierarchical tissue structures consisting of alveolar cavities and up to 2075 cells [Doyle *et al.*, J. Acoust. Soc. Am. **125**, 1751–1767 (2009)]. The simulations used addition theorems to model multiple scattering and showed significant spectral changes that correlated with tumorigenesis. To increase the fidelity of the simulations, the scattering model has been expanded to incorporate spheroidal and cuboidal cells, noncentered nuclei, mixed tissue types, and duct-like channels representing the three-dimensional luminal structure in breast tissue. Theory, simulation results, and experimental data will be presented. [Work supported by NIH 1R21CA131798-01A1.]

11:15

4aBB8. Comparison between independent and multiple scattering models to predict velocity dispersion in trabecular bone. Guillaume Haiat (CNRS, B2OA UMR 7052, 10, avenue de verdun, 75010 Paris, France) and Salah Naili (Universit Paris 12—Val de Marne, 94010 Creteil Cedex, France)

Transverse transmission devices are used clinically to assess fracture risk. However, the physical interaction between trabecular bone and ultrasound remains unclear. Unlike most soft tissues, negative values of velocity dispersion have been measured in trabecular bone. The origin of negative dispersion is still a matter of debate. The aim of this paper is to propose a model predicting the frequency dependence of phase velocity as well as its physical determinants. A two-dimensional homogenization model accounting for the coupling of independent scattering effects with viscoelastic absorption is employed to calculate the dependence of phase velocity on frequency, bone volume fraction (BV/TV), and trabecular thickness (TbTh). A simple biphasic description of TB in which the trabeculae are assumed to be identical and parallel cylinders are considered. The first step of the method is to compute the attenuation coefficient in the entire frequency range using a suitable decomposition via Bessel functions. Then, the Kramers–Kronig relationships are used to derive the frequency dependence of phase velocity. The results obtained assuming, independent scattering are compared with a model accounting for multiple scattering [Haiat *et al.*, J. Acoust. Soc. Am. (2008)], which is shown to be in better agreement with experimental results obtained in the literature.

11:30

4aBB9. Experimental and theoretical investigations of ultrasound contrast agent dynamics between parallel plates and within rectangular channels. Todd A. Hay (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029), Valeria Garbin, Michel Versluis (Univ. of Twente, 7500 AE Enschede, The Netherlands), Yurii A. Ilinskii, and Mark F. Hamilton (The Univ. of Texas at Austin, Austin, TX 78713-8029)

The behavior of ultrasound contrast agent (UCA) microbubbles in confining environments is relevant for applications including ultrasound imaging and targeted microbubbles. In this presentation, measurements from experiments investigating the radial dynamics of UCA microbubbles positioned between parallel plates and within rectangular microchannels will be compared with model simulations. The radial motion of several UCA bubbles was recorded with high-speed imaging at 15 million frames per second over a range of insonation frequencies at various offset distances from the plates or channel walls. The behavior of each individual bubble was observed in both parallel plate and microchannel environments, thereby isolating effects due purely to the confinement. Several plastics and polymers with well-defined material properties were used to construct the plates and channels, making the dataset suitable for model validation. Measurements show that the oscillation amplitude and natural frequency of the bubbles depend on their position with respect to the walls as well as the material properties and geometry of the confining surfaces. Comparisons with various models for parallel plate and microchannel geometries will be presented. [Work supported by NIH DK070618.]

11:45

4aBB10. High-frequency wave propagation in suspensions of elastic particles: Experimental validation of a time domain numerical simulation. Guillaume Haiat (Laboratoire de Recherches Orthopediques, CNRS, Univ. Paris 7, UMR CNRS 7052 B2OA, 10 avenue de Verdun, 75010 Paris, France), Belfor Galaz (Univ. Paris 6, 75006 Paris, France), Romain Berti, Nicolas Taulier (Univ. Paris 6, 75006 Paris, France), Jean-Jacques Amman (Universidad de Santiago de Chile (USACH), Santiago, Chile), and Wladimir Urbach (Univ. Paris 6, 75006 Paris, France)

Sound propagation in suspensions of elastic particles immersed in a fluid is a difficult problem due to (i) the random distribution of the particles and (ii) possible multiple scattering and mode conversion effects. The ultrasonic propagation in suspensions has mostly been studied in the frequency domain and time domain approaches may also be of interest. Here, two-dimensional finite difference time domain numerical simulations of ultrasonic propagation in suspensions of polystyrene 6-μm-diameter microspheres immersed in water were performed around 50 MHz. The numerical results were compared with the Faran analytical model, leading to a maximum difference of 15%, which constitutes a validation of the numerical approach. Experiments were carried out in transmission and backscattering modes. The attenuation coefficient 50 MHz [$\alpha(50)$], the ultrasonic velocity (V) and the relative backscattered intensity (I_B) were measured for different concentrations (from 2 to 25 mg/ml). Each experimental ultrasonic parameter was successfully compared to the numerical results obtained by averaging the ultrasonic parameters derived from 15 different spatial distributions of the micro-disks. The good agreement between the experimental and numerical results constitutes a validation of the approach, which may be used in the future in different applications such as ultrasound contrast agent.

Session 4aEA

Engineering Acoustics: Magnetostriction in Galfenol and its Applications

Stephen C. Thompson, Chair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Chair's Introduction—8:25

Invited Papers

8:30

4aEA1. Iron-gallium (Galfenol) transduction alloys: Magnetic and mechanical properties. James B. Restorff, Marilyn Wun-Fogle (Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700), Arthur E. Clark (Clark Assoc., Adelphi, MD 20783), Thomas A. Lograsso (U.S. Dept. of Energy, Ames, IA 50011-3020), and Gabriela Petculescu (Univ. of Louisiana, Lafayette, LA 70504)

Galfenol, $Fe_{100-x}Ga_x$ ($\sim 5 < x < \sim 20$), is a unique magnetostrictive alloy that combines high saturation magnetostrictions with robust mechanical properties. [Clark *et al.*, "Magnetostrictive properties of body-centered cubic Fe-Ga and Fe-Ga-Al alloys," IEEE Trans. Magn. **36**, 3238 (2000)]. These materials can be welded and machined. Additionally, alloys treated with a stress annealing process retain the maximum value of magnetostriction up to tensile loads of approximately 38 MPa. [Wun-Fogle *et al.*, "Stress annealing of Fe-Ga transduction alloys for operation under tension and compression," J. Appl. Phys. **97**, 10M301 (2005)]. Room temperature $Fe_{81}Ga_{19}$ single crystal properties are $M_s \cong 1.7$ T, $(3/2)\lambda_{100} \cong 400 \times 10^{-6}$, $c' \cong 20$ GPa, and tensile strength >500 MPa. In this presentation, temperature, field and stress dependent magnetic and elastic properties such as permeability, piezomagnetic d constants, Young's modulus and coupling factors will be presented for several Fe–Ga alloys.

8:55

4aEA2. Large-magnetostriction ternary alloys obtained from Galfenol through partial substitution of Ga. A. O. Mandru, G. Petculescu (Univ. of Louisiana, Lafayette, LA 70504), K. B. Hathaway (G/J Assoc., Annapolis, MD 21401), T. A. Lograsso (Ames Lab., Ames, IA 50011), A. E. Clark (Clark Assoc., Adelphi, MD 20783), M. Wun-Fogle, and J. B. Restorff (Naval Surface Warfare Ctr., W. Bethesda, MD 20817-5700)

In this paper the magnetoelastic characteristics of ternary Fe–Ga–X (X=Ge, Sn) alloys are compared with those of Galfenol (Fe–Ga) within the same range of solute concentration. New magnetostriction and elastic modulus data for Fe–Ga–Ge at six different compositions are presented. The most notable case is that of $Fe_{86.9}Ga_{9.9}Ge_{8.7}$. Although the tetragonal magnetostriction, λ_{100}, of Fe–Ge is lower overall than that of Fe–Ga, the magnetostriction of the aforementioned ternary alloy equals the value of Galfenol at the same solute concentration, $Fe_{86.9}Ga_{13.1}$. The theory proposed for explaining the measured λ_{100} values in the ternary alloys takes into account the total number of electrons each solute contributes to the alloy. The valence of Fe is taken to be "1" from first-principles calculations. Previously published results on Fe–Ga–Sn [Summers *et al.*, "Magnetostriction of binary and ternary Fe-Ga alloys," J. Mater. Sci **42**, 9582–9594 (2007)] concur with this theory (Sn and Ge belong to the same periodic table group). The Sn ternary alloys are of economic importance for large-scale applications given that significant magnetostriction can be achieved while using less expensive solutes. It is also notable that, when compared with Galfenol at a Ga concentration of x=6.2 at. %, the ternary alloy $Fe_{88.1}Ga_{6.2}Ge_{5.7}$ has a magnetostriction that is larger by a factor of 2.

9:20

4aEA3. Galfenol technology, state-of-the-art. Eric Summers and Julie Slaughter (ETREMA Products, Inc., 2500 N. Loop Dr., Ames, IA 50010, eric.summers@etrema.com)

Galfenol is a recently discovered magnetostrictive material containing iron (Fe) and gallium (Ga) in various ratios based on desired properties. This alloy system exhibits several unique advantages over legacy smart materials such as Terfenol-D and PZT. Galfenol can be readily machined using conventional machining techniques, it can be formed into various geometries using standard metal working practices such as forging and rolling, alloys can be produced with large internal pre-stresses negating the necessary presence of external pre-load mechanisms, and Galfenol can be welded to other ferrous materials without issue. All of these attributes make Galfenol an attractive new material for the applications engineer. In this presentation, the current state-of-the-art of Galfenol technology will be discussed along with future technology development. Static and dynamic magnetic properties will be presented for various Galfenol alloys; in addition, mechanical properties and available processing routes will be discussed. The work presented includes progress in transitioning this technology from university and laboratory-scale to small-scale production and the impact of material processing and availability on transducer applications.

9:45

4aEA4. Galfenol low frequency slotted cylinder transducer. Richard Meyer, Jr. (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804, rmeyer@psu.edu), Eric Summers, Julie Slaughter (ETREMA Products, Ames IA 50010), Charles Allen, and Stephen C. Thompson (The Penn State Univ., State College, PA 16804)

The ability of Galfenol to be readily formed into various geometries can be exploited by device designers in their efforts to create innovative sources and sensors. In this research forged Galfenol alloy is utilized as the active source actuating a slotted cylinder transducer (SCT). Analytical modeling was utilized to determine the geometry necessary to achieve the notional performance specifications: 210-dB source level, 750-Hz resonance frequency, and 200-Hz bandwidth. The final form factor was a Galfenol SCT with overall dimensions of 7.75-in. diameter $\times$ 12-in. length with a graphite composite shell material and Galfenol drivers comprising a rib-like structure within the shell. The advantages this device has over existing SCTs are reduced cost, ease in manufacture and assembly, and improved reliability. Magnetic and mechanical models were combined in order to conduct three dimensional FEA analysis. The magnetic circuit was modeled in COMSOL MULTIPHYSICS and the acoustic model was built in ATILA. This presentation will include the predicted SCT performance based on the modeling effort as well as details on material manufacturing. Future work will include design optimization to further improve output and bandwidth.

10:10—10:25 Break

10:25

4aEA5. Acoustic cilia arrays using electrochemically-synthesized Galfenol nanowires and heterostructures. Madhukar Reddy, Patrick D. McGary, Mazin Maqableh (Elec. and Comput. Eng., Univ. of Minnesota, 200 Union St. SE, Minneapolis, MN 55455, stadler@umn.edu), Patrick Downey, Alison Flatau (Univ. of Maryland, College Park, MD 20742), and Bethanie J. H. Stadler (Univ. of Minnesota, Minneapolis, MN 55455)

A novel cilia design is introduced that incorporates a newly discovered magnetostrictive family of Fe–Ga alloys, also called Galfenol, into self-assembled, vertical nanopores to fabricate nanocantilevers. The mechanical response of the cilia to stimuli can be detected without direct contact by detecting their magnetic output via giant magnetoresistive sensors, similar to hard drive heads which can also be grown in nanoarrays. The electrochemistry of alloys involving non-ideal metals is very complex, so finite element modeling was used to simulate growth parameters and successful growth was achieved. A variety of nanowire geometries were produced, including segmented Galfenol with alternating nonmagnetic material to overcome the shape anisotropy of the nanowires. In addition, the mechanical properties of this new material were measured to determine the effects of nanostructuring. Finally, the nanowires were actuated using an external field to show their potential actuators as well as sensors. [Work supported by ONR.]

10:50

4aEA6. A galfenol force-based energy harvester. Julie Slaughter, Jim Kosterman, Barry Arbetter, and Eric Summers (ETREMA Products, Inc., 2500 N. Loop Dr., Ames, IA 50010, julie.slaughter@etrema.com)

A force-based energy harvester utilizing the magnetostrictive material, Galfenol, has been developed. Energy harvesters currently available primarily utilize a mass-spring concept where optimal energy harvesting efficiency occurs around a distinct resonance frequency. Deviation from the tuned frequency significantly degrades energy harvesting performance. Galfenol was engineered into a bolt-like configuration with a Galfenol rod welded to threaded stainless steel end pieces and pickup coils wound around the Galfenol. Dynamic forces are applied axially to the Galfenol resulting in a change in flux inducing voltage in the pick-up coils. Because dynamic force is the mechanism generating power, the energy harvester works across a broad range of frequencies. Two different electronic packages were developed; a voltage doubler and quadrupler, which provided efficient energy conversion based on the frequency range of interest. The configuration of the Galfenol bolt resulted in an energy harvester capable of harvesting energy over a broad frequency range. The development and the testing of this unique energy harvester will be discussed. Specific points of discussion will be the optimum device configuration, both hardware and associated electronics that provided the largest output power. In addition, the voltage and output power results will be presented as a function of frequency and force.

11:15

4aEA7. Variable compliance split-cylinder transducer using Galfenol for frequency control. Stephen C. Thompson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804), Julie Slaughter, Eric Summers (ETREMA Products, Inc., Ames, IA 50010), Charles Allen, and Richard J. Meyer (The Penn State Univ., State College, PA 16804)

The size of conventional split ring transducers is much smaller than a wavelength at operating frequency leading to a high Q resonance and limited operating bandwidth. This work investigated modifying the split ring structure to allow its resonant frequency to be adjusted under active control so that it is always operated at or near its resonance. Because the transducer is operated at resonance, the phase of its input impedance is approximately constant, simplifying the design of the transmit power amplifier and reducing the size and weight of the system. A high Q system is desirable for an actively tuned system resulting in very high efficiency and allowing the use of less expensive shell materials. The two variable compliance designs studied involve the use of Galfenol stiffener bars located at the nodal point/hinge point of the shell. By activating all or a portion of these bars, the compliance (stiffness) of the shell can be varied with a corresponding shift in resonance frequency. Two methods of compliance control are being investigated, one that switches the stiffness "on" and "off," and one that allows continuous control using the "delta-E" effect in Galfenol.

Contributed Paper

11:40

4aEA8. Multiphysics modeling of laminated Galfenol transduction devices. Scott P. Porter, Stephen C. Thompson, and Richard J. Meyer, Jr. (Appl. Res. Lab., The Penn State Univ., State College, Pennsylvania, 16804)

Galfenol is an attractive giant magnetostrictive material for transducer design because of its competitive strain capabilities, mechanical robustness, and high magnetic permeability. The latter two properties are particularly interesting as they allow exploration of a design space inaccessible to other active materials, both giant magnetostrictive and piezoelectric. Especially noteworthy is the ability to roll Galfenol into thin sheets from which laminated assemblies may be economically produced. Lamination is an effective mitigation scheme against eddy current losses and in this paper the authors explore transducers based on this construction. Much of this work builds on the designs of World War II era magnetostrictive transducers developed at the Harvard Underwater Sound Laboratory. Modeling these transducers as electro-magneto-mechano-acoustical devices can be first achieved in a limited sense with a one-dimensional model; this model is valuable for rapid simulation during the early design process. Many of the physical phenomena present in the transducer, however, are inherently beyond the limitations of the one-dimensional model. For this reason, the authors have turned to a multiphysics approach for a comprehensive transducer simulation. In this modeling environment, the coupled differential equations describing the transduction are implemented in a three-dimensional finite element analysis.

THURSDAY MORNING, 29 OCTOBER 2009 LIVE OAK, 8:30 TO 11:50 A.M.

Session 4aMU

Musical Acoustics: Musical Ability and Other Abilities: Innate and Environmental Influences

Diana Deutsch, Chair
Univ. of California, San Diego, Dept. of Psychology, 9500 Gilman Dr., La Jolla, CA 92093

Chair's Introduction—8:30

Invited Papers

8:35

4aMU1. Singing in the brain: Professional singers, occasional singers, and out-of-tune singers. Gottfried Schlaug (Music and Neuroimaging Lab., Dept. of Neurology, Beth Israel Deaconess Medical Ctr. and Harvard Med. School, Boston, MA 02215, gschlaug@bidmc.harvard.edu)

Speaking and singing are the two most complex communication systems of the human being, and perhaps in all of nature, both of which call on extensive cortical and subcortical resources in the brain. Singing is a combination of language and music that taps brain centers involved in both forms of communication and may rely more on right than left fronto-temporal brain structures. I will present results of research studies that examined the neural correlates of singing in professional singers, in occasional singers, and in individuals that sing out of tune. These "tone-deaf" singers lack some of the connections between auditory and motor brain regions that play a critical role in the feedback and feedforward control of vocal communications. These same brain structures are enhanced in professional singers and the duration of training as a singer correlates with gray and white matter enhancements seen in their brains. Ongoing studies are examining the development of these brain regions in children with and without musical training.

9:05

4aMU2. Effects of musical training on brain development. Laurel Trainor (Dept. of Psych., McMaster Univ., 1280 Main St. West, Hamilton, ON L8S 4K1, Canada, ljt@mcmaster.ca)

The auditory cortex builds representations of the auditory environment based on the characteristics of its input. I will present EEG and MEG data showing that, compared to non-musicians, adult musicians have enhanced responses to isolated musical tones, melodic sequences, and simultaneous melodies. Whether these differences are the result of musical experience is explored through training studies in adults and longitudinal studies comparing children engaged in musical lessons with children engaged in other activities. It will be shown that auditory sensory representations undergo a protracted developmental trajectory, not reaching adult states until well into the teenage years. Furthermore, the effects of musical experience can be seen in children as young as 4–5 years of age for both sensory and attentional components of the event-related potential. High-frequency oscillatory EEG activity (induced gamma band), which is associated with executive control, attention, and memory, emerges after 1 year of music lessons in children of this age, but is not seen in children who are not taking music lessons. Thus any benefits of musical training for other cognitive domains is likely mediated by the development of superior executive functioning.

9:35

4aMU3. Musical ability and cognitive abilities. Glenn Schellenberg (Dept. of Psych., Univ. of Toronto at Mississauga, Mississauga, ON L5L 1C6, Canada, g.schellenberg@utoronto.ca)

Does music make you smarter? Associations between music and cognitive functioning are notable only if the benefits apply reliably to nonmusical abilities and if music is unique in producing the effects. Such associations could arise either from music listening or music lessons, and there is no reason to believe that observed associations between cognitive abilities and music listening should parallel those involving music lessons. The available evidence indicates that music listening leads to enhanced performance on a variety of cognitive tests, but that such effects are short-term and stem from the impact of music on arousal level and mood, which, in turn, affect cognitive performance; experiences other than music listening have similar effects. Music lessons in childhood tell a different story. They are associated with small but general and long-lasting cognitive benefits that cannot be attributed to obvious confounding variables such as family income and parents education. Nonetheless, such associations may not be evident in samples of adults with many years of music lessons or among professional musicians. Moreover, the mechanisms underlying observed associations between music lessons and cognitive abilities have yet to be determined.

10:05—10:20 Break

10:20

4aMU4. Evidence for a genetic contribution, and a chromosome 4 locus, to musical aptitude. Harald Goring (Dept. of Genetics, Southwest Foundation for Biomedical Res., 7620 NW Loop 410, San Antonio, TX 78227-5301, hgoring@sfbrgenetics.org)

The genetic influences on mental diseases, skills, and behaviors are essentially entirely unknown. We have performed a family study to investigate the genetic etiology of musical aptitude. We focused on three measures of musical aptitude that in our opinion likely are more reflective of innate ability than of musical training: a test for auditory structuring ability [Karma music test (KMT)] commonly used in Finland, and the Seashore pitch and time discrimination subtests (SP and ST, respectively) used internationally. These phenotypes were assessed in 15 Finnish, multigenerational families comprising 234 individuals. Using variance component-based methodology, we obtained significant evidence of a genetic contribution to all phenotypes, with heritability estimates of 42% for KMT, 57% for SP, 21% for ST, and 48% for a combined music test score. To map any underlying major loci, we performed multipoint genome-wide linkage analysis, using 1113 microsatellite markers. We obtained nominally significant linkage evidence on chromosome 4q22 (lod=3 .33) for the combined music test score. This locus seems to largely influence KMT, which yielded a lod score of 2.91 in the region. Our results suggest that genetic factors in aggregate influence musical aptitude, with chromosome 4 appearing to harbor a major locus.

10:50

4aMU5. The development of domain-specific musical knowledge. Erin Hannon (Dept. of Psych., Univ. of Nevada, Las Vegas, 4505 Maryland Pkwy., Box 455030, Las Vegas, NV 89154-5030, erin.hannon@unlv.edu)

Music and speech are both complex sound structures that unfold over time, yet they undoubtedly rely on fundamentally dissimilar perceptual processes and knowledge. Given that naive infant listeners must discover and learn meaningful structures in music and speech in parallel, a natural question to ask is whether early auditory learning is influenced by domain-specific or domain-general constraints. I will discuss two current studies exploring and comparing how infants and adults respond to speech-music similarities and dissimilarities. Because music and speech share certain rhythmic properties, one set of experiments explores the extent to which adults and infants can classify speech and music stimuli based on the rhythm. After being familiarized to low-pass filtered speech or instrumental songs from French-speaking or English-speaking cultures, adults fail to classify novel stimuli across domains (e.g., French and English speech after familiarization to songs), whereas 10-month-old infants succeed. A second set of experiments examines whether infants and adults exhibit domain-specific biases when required to infer rule-like patterns from sound sequences in a linguistic versus musical context (e.g., rules based on syllables versus pitch). Results across both studies indicate that domain-specific knowledge and learning biases may increase from infancy to adulthood.

11:20

4aMU6. The genesis of absolute pitch: Theories and experimental findings. Diana Deutsch (Dept. of Psych., Univ. of California, San Diego, 9500 Gilman Dr. 0109, La Jolla, CA 92093-0109, ddeutsch@ucsd.edu))

This paper reviews theories and findings concerning the genesis of absolute pitch, defined as the ability to name a musical note without the aid of a reference note. Given the rarity of absolute pitch, there has been much speculation concerning its genesis. Four theories are evaluated: First, absolute pitch can be acquired at any time through practice; second, it is a rare genetic trait which rapidly becomes apparent as circumstances allow; third, absolute pitch is subject to a critical period that is specific to music; fourth, the faculty as generally documented is essentially language-based, and so is subject to the same critical period as are other aspects of language. In this context, the evidence considered that absolute pitch is present implicitly in most individuals, even those who are unable to label musical notes when these are presented in isolation. Findings concerning the neurological underpinnings of absolute pitch are also reviewed.

Session 4aNS

Noise: Advancements in Noise Control in Latin America

Sergio Beristain, Chair
Mexican Institute of Acoustics, P.O. Box 12-1022, Narvarte, Mexico City, 03001 DF, Mexico

Chair's Introduction—8:20

Invited Papers

8:25

4aNS1. Noise control evolution in Mexico. Sergio Beristain (Lab. Acoust., E.S.I.M.E. ipn, P.O. Box 12-1022, Narvarte 03001, DF Mexico, Mexico, sberista@hotmail.com)

Noise legislation in Mexico started in the mid-1970s, when most of the standards and recommendations available nowadays were issued, ranging from noise classification, sound level meter characteristics, to noise pollution and labor noise limits. Several of those standards were modified, with very little updating in the mid-1990s, but most of them were just transposed to the national standards format, without any major technical change, because the original issuing office was a federal depending ministry, but not the official standards office. More recently one of those standards was improved in only 3 dB, and a noise monitoring system for Mexico City is under preparation. From the non legislation point of view, some posters have been published in newspapers and billboards, and radio and TV noise control programs have been broadcasted.

8:45

4aNS2. Mexico City urban noise control network. Felipe Rolando Menchaca Garcia, Luis Pastor Sanchez Fernandez, and Sergio Suarez Guerra (Comput. Res. Ctr., Natl. Polytechnical Institue, Mexico City, Mexico 03001)

This paper presents an IP network model for monitoring urban noise emitted from mobile and fixed sources. The model is applying to build a monitoring network in the historic center of Mexico City. Design of network system includes measuring station design, network design, central of monitoring design, measuring politics and functionality, event detection modules, central database of historic information, report module design, and administration system. Mexico City Urban Noise Control Network is going to be an important tool to enforce application of local noise regulations, detection of infractions for excessive noise, detection of security events like gun shots, and monitoring of noise pollution levels. Measuring stations are developed to detect average standard levels and individual events, to generate warnings and alarms, and to detect some specific events. Network is developed over a wireless Network.

9:05

4aNS3. Factors that have positive or negative influence in the noise levels in the great cities of Mexico. Fernando J. Elizondo-Garza (Acoust. Lab., FIME-UANL, P.O. Box 28 F, Cd. Universitaria, San Nicolas, NL, 66450, Mexico, fjelizon@hotmail.com)

In this paper the noise condition in the largest cities of Mexico is described by means of an "analysis of opposed." The historical, sociocultural, political, economic, and technological factors that have positive or negative influence in the noise generation are considered, which imply a fight of opposites, in such a way that the current situation will be the resultant of all of their influences. By visualizing the noise pollution problem in this way, although not a quantitative approach, represents a very useful qualitative tool, especially to settle down strategies to reduce the noise levels in the cities. Unfortunately, in some cases, the inertia of the negative factors—which are the producers of noise—may cause that, in the short term, some positive proposals to reduce the noise levels may not result feasible.

9:25

4aNS4. Noise source identification in transportation systems in Brazil using beamforming technique. Samir Gerges, William Fonseca (Mech. Eng., UFSC, Trindade, Florianopolis, SC, Brazil), Jose Bento Coelho, and Miguel M. Neves (IST, 1049-001 Lisbon, Portugal)

Beamforming with acoustic phased arrays has become a popular technique for locating and quantifying noise sources in complex system such as environmental noise, aircraft noise, internal and external automobile noise, beside other applications. The signal processing performs traditional delay-and-sum beamforming and accounting for the different propagation paths from the source and to the different microphones. The sources may pass the acoustic camera with high speeds, allowing the beam to focus and follow the bypassing source past the microphone array. The signal processing can zoom into any part of the studied objector of any frequency band. Ultimately, image quality depends on the number of microphones and their configuration. When the microphones cover a big surface, the resolution gets higher; however, the image may contain noise at high frequencies. Reducing microphone distances from the source removes the specks but at the same time worsens the resolution. So the best way to guarantee high-quality images is to use as many microphones as possible. This paper presents the results of tests with 32-channel from National Instrument phased array system developed by the UFSC-Brazil. Measurement results are shown for motorcycle, car motor, and external automobile noise during pass-by test.

9:45

4aNS5. Regulation of noise in Chile: Progress and challenges. Jaime Delannoy (Instituto de Acustica, Brown Norte 290 Nunoa, Santiago 779-0569 Chile jdelannoy@vtr.net), Igor Valdebenito, and Roberto Quezada (CONAMA)

In the past 20 years Chile has shown a growing concern for noise pollution. There are regulations for stationary noise sources and also for the surface public transportation. There was a change in the exchange rate to calculate the dose of noise in occupational environments (from $q=5$ to $q=3$). The System of Environmental Impact Assessment sets out to mitigate the potential noise impact of each new project. However, the authorities have not been possible to reach an agreement to regulate noise pollution caused by the construction of buildings. In addition, future noise regulations for airports and for noise of private transport are major challenges for the country.

10:05—10:25 Break

10:25

4aNS6. Progress of noise control in Chile–South America. Christopher H. Rooke (Silentium, Alcalde Pedro Alarcón 877, San Miguel, Santiago de Chile, crooke@silentium.cl)

The Chilean experience can be interesting to other countries in Latin America. A short review from an educational point of view will include: Universities offering undergraduate acoustic programs and post graduate programs in Chile; non state institutions providing acoustic quality standards; short review of national standards and laws related to noise control in Chile: environmental noise law, hearing protection law, new laws, and standards at project stage at state institutions. Difficulties found on law application and challenges of state inspectors and institutions. How law implementation, undergraduate programs, community and business community awareness, and several factors had allowed a development of acoustic consultants, solution manufacturers, and installers and several other progresses. What is still missing and the long-term way to achieve higher levels of education—related to noise pollution and control—more effective laws and standards, state inspection, political will, consultant development, tested products and solutions, and general awareness of noise as a health problem where all of us had something to say and do to achieve better quality of life.

Contributed Paper

10:45

4aNS7. Shortcomings of the Brazilian standards on community noise. Jose Augusto Nepomuceno and Daniel Mandu (Acustica & Sonica, Rua Fradique Coutinho 955 s 01, 05416-011 Sao Paulo, Brazil, janepomuceno@yahoo.com)

The Brazilian standard NBR 10151 was revised on 2000 to establish guidance for evaluation of noise in inhabited areas aiming the comfort of community. While this is a technical standard, it is often used as a "federal legislation." States or cities have their own community noise regulation. The NBR 10151 has shortcomings that should be resolved. For example, there is no mention about parameters as Ldn or percentile levels or about evaluation of tone content or sampling methods. The only evaluation parameter is the LAeq. The urban zones are characterized as "rural," "residential (restricted) hospital and schools," "mixed area, mostly residential," "mixed area, mostly commerce and administration," "mixed area, mostly recreation and leisure," and "mostly industrial." There is no guidance how to apply the classification . There are several communities around factories that would be better classified in other ways. The approach in cities and state legislation differs. Again, there are shortcomings as lack of definitions or vague descriptions in these legislations. The authors discuss the Brazilian standards applied to factories and freight harbor. In these situations the shortcomings made them not helpful at all. The Brazilian standards are compared to international regulations used in these projects.

THURSDAY MORNING, 29 OCTOBER 2009 BLANCO/LLANO, 9:00 TO 10:00 A.M.

Session 4aSAa

Structural Acoustics and Vibration: Structural Acoustics for Damping Materials

J. Stuart Bolton, Chair
Purdue Univ., School of Mechanical Engineering, West Lafayette, IN 47907-2031

Contributed Papers

9:00

4aSAa1. Analysis of structural damping performance in passenger vehicles chasis. Marcio Furukava, Samir Gerges, Miguel M. Neves, and Bento J. L. Coelho (CAPS, IST, TULisbon, Lisbon, Portugal)

Viscoelastic material is commonly used in the passive treatment of structural vibration of the vehicle structure, to introduce damping and reduce vibration level at resonance frequencies. The effectiveness of such damping treatment depends on design parameters such as location, area, thickness, and choice of the type of the damping material. This paper describes the experimental proceeds to evaluate two different layer damping material performances, when they are applied to a passenger vehicle chassis.

9:15

4aSAa2. Negative stiffness metamaterial elements for enhanced material damping capacity. Lia B. Kashdan, Michael R. Haberman, Preston S. Wilson, and Carolyn C. Seepersad (Appl. Res. Labs. and Dept. of Mech. Eng., The Univ. of Texas at Austin, 1 Univ. Station C2200, Austin, TX 78712-0292, haberman@arlut.utexas.edu)

Recent research has shown that the introduction of small quantities (1%–2% by volume) of constrained negative stiffness inclusions into a viscoelastic host material leads to an effective energy absorbing behavior greatly exceeding that of conventional materials [Lakes *et al.*, Nature (London) **410**, 565–567 (2001)]. The root cause of this increased absorptive

capacity is the ability of the negative stiffness materials to do work on the viscoelastic host when perturbed by an incident disturbance such as a sound wave or vibrational motion [Koutsawa *et al.*, Int. J. Mech. Mater. Des. **5**, 29–44 (2007)]. A buckled beam constrained to an S-shaped configuration has a strain energy density at a local maximum and will accordingly display negative stiffness. This negative stiffness behavior is non-transient and broadband in nature. The buckled beam structure is therefore a compelling candidate for the construction of new metamaterials for noise reduction, anechoic coatings, and transducer backing materials. This work presents ongoing research that exploits the geometric freedom provided by selective laser sintering to design and construct bistable beam elements. Design configurations are presented together with dynamic test results quantifying the damping capacity via system transmissibility and the resulting decrease in mechanical quality factor at resonance.

9:30

4aSAa3. Noise reduction of honeycomb sandwich panels with acoustic mesh caps. Christina Naify and Steve Nutt (Dept. of Mater. Sci., Univ. of Southern California, 3651 Watt Way VHE 416, Los Angeles, CA 90089)

The noise reduction effectiveness of acoustic mesh septa caps in composite honeycomb sandwich panels was evaluated. The acoustic mesh in the cells acted as a lightweight resonator in weight-critical aerospace structures. Nylon mesh was inserted into the honeycomb cells of the sandwich panels, and the acoustic absorption (AA) and transmission loss (TL) of the panels were measured using a normal incidence impedance tube. Single septa inserted into each cell yielded consistent improvements of 2 dB in TL over low- to mid-frequencies (50–1600 Hz). A double layer of septa further increased these acoustic properties in the mid-frequency range by an additional 2 dB. Varying the size of the mesh-weave had a significant effect. AA of the panels improved by a factor of 2 for the single septa and was further increased by perforating the panel facesheet. Results of the facesheet perforation were consistent with predicted values. Over low- to mid-frequencies, septa yielded measurable sound-mitigating results for a small weight addition with no change to the mechanical performance of the panels.

9:45

4aSAa4. Vibration damping of metal foam-polymer interpenetrating phase composite. Nassif E. Rayess (Univ. of Detroit Mercy, 4001 W. McNichols Rd., Detroit, MI 48221, rayesna@udmercy.edu)

The damping and basic dynamic properties for a novel type of multifunctional hybrid material known as metal foam-polymer composite are investigated. This material is obtained by injection molding a thermoplastic polymer through an open cell aluminum foam, in essence creating two contiguous morphologies, an aluminum foam interconnected "skeleton" with the open pores filled with a similarly interconnected polymer substructure. This coexistence of both materials allows each to contribute its salient properties (e.g., the plastics contributing surface toughness and the metal foams contributing thermal stability). Basic damping testing results are presented for various aluminum foam porosities and pore sizes as well as for three types of polymers. A basic mathematical model of the damping is also presented. The integrity of the interface between the aluminum foam and the polymer is discussed in terms of its effect on the overall material damping.

THURSDAY MORNING, 29 OCTOBER 2009 BLANCO/LLANO, 10:30 TO 11:45 A.M.

Session 4aSAb

Structural Acoustics and Vibration: Structural Acoustics and Vibrations for Micro/Nano Devices

Henry A. Scarton, Chair
Rensselaer Polytechnic Inst., Dept. of Mechanical, Aerospace, and Nuclear Engineering, Troy, NY 12180

Contributed Papers

10:30

4aSAb1. Characterizing energy dissipation in singlewalled carbon nanotube polycarbonate composites. Nikhil A. Koratkar, Jonghwan Suhr (Dept. of Mech., Aerosp. and Nuclear Eng., RPI, Troy, NY 12180), Amit Joshi, Ravi S. Kane, Linda S. Schadler, Pulickel M. Ajayan, and Steve Bartolucci (RPI, Troy, NY 12180)

In this study, singlewalled carbon nanotube and bisphenol-A-polycarbonate composite beams were fabricated by a solution mixing process and dynamic (cyclic) load tests were performed to characterize energy dissipation. We report up to an order of magnitude (>1000%) increase in loss modulus of the polycarbonate system with the addition of 2% weight fraction of oxidized singlewalled nanotube fillers. We show that the increase in damping is derived from frictional sliding at the nanotube-polymer interfaces. The nanoscale dimensions of the tubes not only results in large interfacial contact area, thereby generating high damping efficiency, but also enables seamless integration of the filler materials into the composite structure.

10:45

4aSAb2. Temperature activated interfacial friction damping in carbon nanotube polymer composites. Jonghwan Suhr, Wei Zhang (Dept. of Mech., Aerosp. and Nuclear Eng., RPI, Troy, NY 12180), Pulickel M. Ajayan, and Nikhil A. Koratkar (RPI, Troy, NY 12180)

Effect of temperature on interfacial sliding in single-walled carbon nanotube polycarbonate composites is investigated experimentally. We show that interfacial slip at the tube-polymer interfaces can be activated at relatively low dynamic strain levels (~0.35%) by raising temperature to ~90°C. We attribute this to increased mobility of the polymer chain backbones at elevated temperatures and thermal relaxation of the radial compressive stresses at the tube-polymer interfaces. These results show the potential of polymer nano-composites as high temperature damping materials for vibration and acoustic suppression in a variety of dynamic systems.

11:00

4aSAb3. Nonlinear sensitivity of capacitive micromachined ultrasonic Lamb wave transducers. Li-Feng Ge (School of Electron. Sci. and Technol., Anhui Univ., 3 Feixi Rd., Hefei, Anhui 230039, China, lfge@ahu.edu.cn)

A capacitive micromachined Lamb wave transducer using the second-order bending mode of high-aspect rectangular diaphragms to excite and receive Lamb waves in a silicon plate was presented [L. Ge, Proc. SPIE 6358, 1 (2006)]. Thus, the width of the diaphragm is made equal to the wavelength of the *A*0 wave, and all the transducer parameters can be determined by the TDK model for CMUTs with high-aspect rectangular diaphragms. The sensitivity, i.e., a ratio of the relative frequency shift to the change in mass loading of the plate, is derived and is a nonlinear function, which is the same as that for capacitive ultrasonic mass resonators [L. Ge, IEEE IUS, 1034 (2008)]. Also, it is a general nonlinear sensitivity expression available for both capacitive and piezoelectric gravimetric sensors. The area density increment, caused by the adsorbed or bound molecules or the fluid loading, is

usually comparable with the area density of the silicon plate. Therefore, it is significant to use the nonlinear expression for determining the sensitivity accurately. The capacitive Lamb wave transducers are very attractive as gravimetric sensors for chemical and biological sensing since they have much simple structure and high sensitivity. [Work supported by NSFC (60774053).]

11:15

4aSAb4. A piezoelectric microelectromechanical systems ultrasonic radiator. Benjamin A. Griffin, Matthew D. Williams, and Mark Sheplak (Dept. of Mech. and Aero. Eng., Univ. of Florida, 231 MAE-A, Gainesville, FL 32611, griffo@ufl.edu)

The development of a piezoelectric micromachined ultrasonic radiator is presented. The transmitter is formed by a radially non-uniform circular composite diaphragm. The diaphragm consists of a layer of piezoelectric material with annular electrodes. Transduction of the diaphragm occurs when an electric field is induced between annular electrodes across the piezoelectric material creating mechanical strain in the annular region. The strain couples into force and moment resultants that generate diaphragm deflection. A time varying electric field induces vibrations of the diaphragm which in turn generates acoustic waves. Included in the overall systems design of the transmitter is a unique package with back cavity depth control. A systems model incorporates electrical, mechanical, and acoustic components using composite plate mechanics, lumped element modeling, and acoustic theory. Comparison to experimental results is conducted through mechanical and acoustic characterizations. A laser vibrometer system measures diaphragm velocity from which the fundamental mode shape and resonant frequency are extracted. Farfield acoustic measurements are compared to the extrapolation of vibrometer measurements via Rayleigh's integral. The characterization results show mismatch between devices as well as the equivalent circuit model. Experimental results of the variable back cavity depth show a qualitative match to the systems model.

11:30

4aSAb5. Diffraction based optical microelectromechanical systems microphone and its application to the measurement of sound intensity. Baris Bicen (G. W. Woodruff School of Mech. Eng., Georgia Inst. of Tech., 801 Ferst Dr. NW, Atlanta, GA 30332-0405, baris@gatech.edu), Weili Cui, Quang Su, Ronald Miles (SUNY at Binghamton, Binghamton, NY 13902-6000), and Levent Degertekin (Georgia Inst. of Tech., Atlanta, GA 30332-0405)

Diffraction based optical MEMS microphones and force feedback capabilities of these microphones are presented earlier [Bicen *et al.*, J. Acoust. Soc. Am. **123**, 3230 (2008)]. Further investigation is done in order to see the capabilities of these optical microphones in measuring the sound intensity. The biomimetic diaphragms of these microphones are designed for responding to pressure gradient which can be used for extracting the particle velocity required to measure the sound intensity. In order to show the velocity measuring capability of these microphones, an impedance tube method is used because the velocity and pressure at each point in the impedance tube can be found accurately. The comparison of the measured velocities and sound intensities with the optical microphone and commercially available sound intensity probes is presented. Model results are discussed in order to show the effect of the different mode shapes on the accuracy of the pressure gradient measurements. Optimization of these biomimetic diaphragms will result in the measurement of the sound intensity in a larger bandwidth without using any spacers and additional microphone adapters like in commercial sound intensity probes. [Work supported by National Institutes of Health (NIH) and the Catalyst Foundation.]

THURSDAY MORNING, 29 OCTOBER 2009 RIO GRANDE EAST, 8:00 TO 11:25 A.M.

Session 4aSC

Speech Communication: Perceptual Learning and Adaptation in Speech

Tessa C. Bent, Chair
Indiana Univ., Speech and Hearing Science, 200 S. Jordan Ave., Bloomington, IN 47405

Chair's Introduction—8:00

Invited Papers

8:05

4aSC1. How listeners represent dialectal variation. Tanya Kraljic (Dept. of Psych., Univ. of Pennsylvania, 3401 Walnut St., Ste. 400A, Philadelphia, PA 19104, kraljict@sas.upenn.edu)

Successful speech recognition requires that listeners map acoustically variant pronunciations onto the same meaningful interpretation. As listeners are exposed to a talker's speech, they learn the relevant mapping, and their perception shifts to accommodate the talker's pronunciations. Recent work on the perceptual learning of idiolectal variation has made it possible to investigate precisely what it is that changes in perceptual processing as listeners adapt to speakers. These studies suggest that phonemic representations themselves are rapidly modified to reflect the idiolectal variants produced by a talker. However, dialectal variation appears to be handled differently. Several studies will be presented that address the representation of dialectal forms, including a recent experiment that directly compared learning of a dialectal variant to learning of an idiolectal variant with the same acoustic properties. Listeners adjust to talkers' dialects, but unlike with idiolectal variation, such adjustments do not reflect an underlying modification of the phonemic representations of the canonical counterparts. The implication of these results for dialect representation, as well as for the perceptual learning process, will be discussed. [This work was supported by NIH NRSA Award No. 7 F32 HD052342-03.]

8:35

4aSC2. Perceptual adaptation to foreign-accented English. Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, abradlow@northwestern.edu)

Foreign-accented speech presents an interesting challenge for native listeners due to its deviations from native accented speech along multiple acoustic-phonetic dimensions. Nevertheless, since these deviations arise primarily from interactions of the native (L1) and target (L2) language sound systems, they are highly systematic both within individuals and across talkers from the same L1. Moreover, since numerous features of English are typologically rare (e.g., coda clusters and large vowel inventory), various foreign-accents of English frequently converge on similar specific deviations (e.g., cluster reduction and vowel category conflation). Thus, despite a high degree of deviation from native talker norms, the systematicity of foreign-accented speech should facilitate listener adaptation. We examined adaptation at the level of the individual (talker-dependent), the L1 (talker-independent), and the population (accent-independent). First, we showed variation in the speed and extent of talker-dependent adaptation depending on the talker's baseline English speech intelligibility. Second, we demonstrated that talker-independent adaptation required exposure to multiple talkers of the accent. Finally, we demonstrated that exposure to multiple accents (e.g., Korean-, Hindi-, Mandarin-, and Romanian-accented English) can facilitate accent-independent adaptation, i.e., adaptation to both a trained accented (Mandarin-accented English) and an untrained accent(Slovakian-accented English). This work builds a strong case for highly flexible speech perception mechanisms.

9:05

4aSC3. Perceptual learning of non-native speech sounds: The importance of features from an attention to dimensions perspective. Alexander, L. Francis (Dept. of Speech, Lang. and Hearing Sci., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, francisa@purdue.edu)

Research on the acquisition of non-native speech sounds is frequently conducted within the framework of theories that were originally developed to account for aspects of cross-language speech perception (e.g., PAM, NLM, and the SLM). Such theories typically focus on the influence of native phonetic categories on the perception of non-native speech, and second language learning researchers may tend to adopt this category-focused view although these theories make few strong predictions about learning. Interestingly, despite their focus on the role of categories, all of these theories incorporate the same assumption: That learning (of both first- and second-language contrasts) progresses by reweighting of phonetic features. This talk presents an overview of results from a series of studies that support this core assumption, suggesting that features, not categories, should play the central role in structuring our understanding of phonetic learning, especially with respect to such short-term adult learning as may be obtained in the laboratory or classroom. Taking such a perspective facilitates connections between theories of linguistic and non-linguistic perceptual learning, reduces dependence on category-based models of speech perception and production, and may provide a more principled basis for making predictions about what non-native contrasts should be easy or difficult to learn.

9:35

4aSC4. Perceptual learning of cochlear implant simulated speech: Cognitive findings and clinical implications. Jeremy L. Loebach (Dept. of Psych., St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057, jeremyloebach@gmail.com)

Adaptation to the acoustic world after cochlear implantation requires significant adjustment to degraded auditory signals. The cognitive mechanisms underlying the mapping of such signals onto preexisting internal representations (postlingually deafened individuals), or the formation of novel internal representations (prelingually deafened individuals), are not well understood. Therefore, understanding the mechanisms of perceptual learning is critical to providing efficient training and (re)habilitation for new cochlear implant users. The advent of noise and sinewave vocoders to model the output of a cochlear implant speech processor has increased the tools available to investigate perceptual learning of speech in normal-hearing listeners. A fundamental question is whether training should focus exclusively on speech perception (synthetic approach), or whether training on extralinguistic or nonspeech auditory information (analytic approach) promotes more robust perceptual learning and generalization to novel materials and signals. In this talk, I will present data from two recent studies on the perceptual learning of sinewave-vocoded speech by normal-hearing participants that focused on different levels of linguistic/extralinguistic and nonspeech acoustic information. I will discuss the implications these data have for understanding perceptual learning and the cognitive mechanisms that mediate speech perception.

10:05—10:25 Break

10:25

4aSC5. Disordered speech as a testing ground for listener learning and adaptation. Julie M. Liss (Dept. Speech and Hearing Sci., Arizona State Univ., Tempe, AZ 85287, julie.liss@asu.edu), Andrew J. Lotto (Univ. of Arizona, Tucson, AZ 85721), and Kaitlin Lansford (Arizona State Univ., Tempe, AZ 85287)

Speech production disorders arise from a wide variety of anatomical defects, neurological injury, and disease. Depending on the etiology, the speech manifestations may be quite regularized—in the same way that a foreign accent presents consistently—or highly variable with unpredictable speech interruption and degradation. Indeed, even within an individual speaker, one will find both relatively consistent and variable aspects of speech abnormality. A growing body of literature shows that listeners enjoy intelligibility gains with just brief auditory exposure to disordered speech that is accompanied by lexical feedback, and with repeated unsupervised exposures. These benefits have been shown to be both talker-specific and applicable to groups of people who presumably share relevant acoustic-perceptual speech features. While many of these studies have been designed with an eye toward clinical and communicative implications, there is much to be learned about the mechanisms underlying these perceptual improvements with naturally impoverished speech. The purpose of this presentation is to describe the use of clinical populations as a testing ground for hypotheses of listener learning and adaptation. Speech production disorders particularly lend themselves to the study of episodic memory, rhythmic expectancy, and talker-specific versus talker-general effects. [Work supported by NIDCD, NIH.]

10:55

4aSC6. Recent findings on early word recognition and learning: Representational change in adapting to variability in voice, foreign accent, and dialect. Rachel Schmale (Dept. of Psych., North Park Univ., 3225 W. Foster Ave., Box 16, Chicago, IL 60625, rschmale@northpark.edu)

A core aspect of recognizing and learning words is accommodating variable acoustic forms. Thus, to become proficient learners, infants and toddlers must adapt to the inherent variability of speech, where words frequently vary in form depending on a host of talker characteristics (e.g., voice, emotional state, foreign accent, dialect, speaking rate). Four sets of experiments examined the impact of variability in voice, foreign accent, and dialect on the abilities of English-learning infants and toddlers to recognize and learn words. The findings reveal evidence about the flexibility of early representations across development. In the face of subphonemic and suprasegmental variability, young infants and toddlers seem to encode and store fine acoustic-phonetic details about the words they hear, which impedes their ability to recognize and learn dissimilar instances of the same word. Alternatively, older infants and toddlers more readily accommodate extraneous acoustic variability by abstracting across dissimilar word forms, leading to more successful recognition and learning. Although younger children display more difficulty adapting to irrelevant acoustic variation, preliminary findings reveal benefits to exposure to variability in inducing more abstract, flexible representations. [Alejandrina Cristià, Elizabeth Johnson, and Amanda Seidl.]

THURSDAY MORNING, 29 OCTOBER 2009 RIO GRANDE CENTER, 8:10 TO 11:55 A.M.

Session 4aUW

Underwater Acoustics and Acoustical Oceanography: Progress in Modeling Complex and Dynamic Acoustic and Oceanographic Characteristics of Continental Shelves and Slopes I

Preston S. Wilson, Cochair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin, TX 78712

David P. Knobles, Cochair
Univ. of Texas at Austin, Applied Research Labs., 10000 Burnet Rd., Austin, TX 78758

Chair's Introduction—8:10

Invited Papers

8:15

4aUW1. Surface and deep ducting of sound parallel to strong nonlinear internal wave fronts. Ying-Tsong Lin, Timothy F. Duda (Appl. Ocean Phys. & Eng. Dept., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, ytlin@whoi.edu), and D. Benjamin Reeder (Naval Postgrad. School, Monterey, CA 93943-5738)

Acoustic ducting effects observed in a nonlinear internal wave field on the shelf of the South China Sea northeast of Tung-Sha Island (water depth of ~120 m) are investigated. Environmental measurements from stationary instruments in the area show that the nonlinear internal waves have large amplitudes (over 40 m) and strong curvature. A principal component interpolation method is used to reconstruct the time-dependent three-dimensional internal wave field. A three-dimensional parabolic equation model is then employed to simulate monochromatic 400-Hz acoustic propagation parallel to the internal wave fronts for comparison to field data. The passing internal waves alternately elevate and depress the thermocline to form strong three-dimensional near-bottom acoustic ducts *between* the wave troughs and strong surface ducts of mixed layer water in the wave troughs. These alternating, depth-dependent acoustic ducts create asynchronous, vertically-bifurcated sweeping beams of ducted acoustic energy. The three-dimensional acoustic propagation modeling results are shown to mimic the field observations to a strong degree and elucidate the propagation physics in this highly complex environment.

8:35

4aUW2. Optimal adaptive sampling for continental shelf acoustic forecasting. Kevin D. Heaney (OASIS Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039)

Shallow water acoustic propagation variability is driven by bathymetry and geo-acoustic and oceanographic variabilities. At the shelf-break, the rapid change in bathymetry often leads to a separation of shelf-water with open ocean sea-water, leading to a dynamic shelf-break front. In order to accurately predict ocean acoustic propagation (forecasting), an improved understanding of the dynamic oceanographic features is required. To this end, a system of adaptive sampling, utilizing oceanographic and acoustic measurement systems, dynamic oceanographic models, and data assimilation, is being developed. In this paper we will present the approach to optimal oceanographic glider placement and optimal acoustic measurements for reducing the oceanographic and acoustic forecast uncertainty. The results from three experiments will be presented where oceanographic sampling was combined with acoustic measurements in a region modeled with a high-fidelity dynamical ocean model. [Work supported by ONR Code 32OA and SPAWAR.]

8:55

4aUW3. Empirical dependence of acoustic transmission scintillation statistics on bandwidth, frequency, and range in the Gulf of Maine. Mark Andrews and Purnima Ratilal (Dept. of Elec. and Comput. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115)

The scintillation statistics of broadband acoustic transmissions are determined as a function of signal bandwidth B, center frequency f_c, and range with experimental data in Georges basin in the Gulf of Maine. The received signal intensity is shown to follow the gamma distribution implying that the central limit theorem has led to a fully saturated field from independent multi-modal propagation contributions. The gamma distribution depends on the mean intensity and the number of independent statistical fluctuations or coherent cells in the received signal. The latter is calculated for the matched filter, the Parseval sum, and the band-passed center frequency, all of which are standard ocean acoustic receivers. The number of fluctuations in the received signal is found to be an order of magnitude smaller than the time-bandwidth product TB of the transmitted signal, and to increase monotonically with relative bandwidth B/f_c. A computationally efficient numerical approach is developed to predict the mean intensity and the corresponding broadband transmission loss of a fluctuating, range-dependent ocean waveguide by range and depth averaging the output of a time-harmonic stochastic propagation model. This model enables efficient and accurate estimation of transmission loss over wide areas, which has become essential in wide-area sonar imaging applications.

9:15

4aUW4. Impact on acoustic propagation by internal waves and tides in the region of shelf and slope. Chi-Fang Chen, Yung-Sheng Linus Chiu, and Yuan-Ying Chang (Dept. of Eng. Sci. and Ocean Eng., Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei 106, Taiwan, chifang@ntu.edu.tw)

Internal waves (IWS) and tidal activities were investigated in the South China Sea (SCS) and the northeastern seas of Taiwan. These oceanic processes cause large fluctuations and impact on underwater acoustic propagation. These effects include two-dimensional (2-D) and three-dimensional (3-D) effects. The 2-D and 3-D effects are related to the angle between the directions of sound propagation and IW front. When the IW front are from 20 to 90 deg with respect to the acoustic propagation direction, acoustic mode coupling is the dominant factor which could be sufficiently predicted by the $N\times$2D simulations. Acoustic energy is exchanged between modes and is re-distributed among the water columns. However, when the angles between the wave front and the acoustic wave propagation are 0–20 deg, the horizontal refraction effect dominates over mode coupling and the fully 3-D calculation is needed. The acoustic energy would be refracted, as a consequence, resulting in energy focusing and defocusing. These effects are clearly seen by the series of data collected in the SCS and the region of Northern East of Taiwan. Computer modeling results are used to manifest experiment data results in this research. [This work is supported by National Science Council of Taiwan.]

9:35

4aUW5. Accounting for spatiotemporal variability of shallow water waveguides in the geoacoustic inverse problem. Kyle M. Becker (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804-0030, kmbecker@psu.edu)

Reliable acoustic field predictions are essential for sonar performance modeling and matched field processing, particularly related to the geoacoustic inverse problem. On continental shelves and slopes, properties of the water column can vary widely in both time and space, while the seabed can be heterogeneous over scales of only a few kilometers and sometimes less. This combined spatiotemporal variability of the acoustic waveguide was a hallmark of the Shallow Water Experiment 2006 (SW06). This is illustrated using data collected during SW06 for an acoustic field measured on a fixed receiver array for a source towed along a radial track and repeated several times. In these experiments, the large degree of variability in the observed acoustic fields was attributed to large-scale changes in the water column that occurred over the course of a day. Properly accounting for this variability proved to be essential for obtaining robust geoacoustic inversion results for the seabed. Specifically, the spatiotemporal properties of the water column had to be accounted for in the forward models from which objective functions were derived. Methods for capturing the observed water-column variability are discussed as used in the inverse problem as well as for the more general problem of model-data comparison.

9:55—10:10 Break

Contributed Papers

10:10

4aUW6. Simultaneous inversion of seabed and water column sound speed profiles in range-dependent shallow water environments. Megan S. Ballard (Graduate Program in Acoust., Penn State, P.O. Box 30, State College, PA 16804-0030, msd200@psu.edu) and Kyle M. Becker (Penn State, State College, PA 16804-0030)

It is well known that the properties of the water column and seabed affect acoustic propagation in shallow water. As a result, numerous inversion schemes have been developed to estimate environmental properties. Many of these algorithms address the problem of estimating properties of the seabed alone, while assuming properties of the water column are known. An example of such an inversion scheme is perturbative inversion which uses modal wavenumbers to obtain sound speed in the seabed as a function of depth [Rajan *et al.*, (1987)]. The inversion algorithm involves solving an ill-posed problem, with regularization used to stabilize the solution, resulting in a smoothed version of the true sound speed profile. To simultaneously estimate water column and seabed properties, qualitative regularization is used to resolve the discontinuity in the sound speed profile at the seafloor and approximate equality constraints are used to constrain the solution in portions of the water column for which the data alone are insufficient. The primary advantage of simultaneously inverting for water column and sediment sound speed profiles is that poor knowledge of *a priori* information about water column is not aliased into the solution for the seabed. [Work supported by NDSEG and ONR.]

10:25

4aUW7. Spatial and temporal properties of water column dynamics in the East China Sea. Chad Smith, Michelle Kingsland, and David Bradley (Graduate Program in Acoust., The Penn State Univ., Appl. Res. Lab., P.O. Box 30, State College, PA 16804-0030, cms561@psu.edu)

In late August, early September of 2008, 36 continuous hours of data from a conductivity, temperature, and depth chain tow were collected. The

sensor chain spanned the lower portion of the upper mixed layer, a sharp thermocline, and the upper part of the isothermal lower water depth. The sound velocity field, calculated using the Chen–Millero formulation, is the result of a combination of events, including internal waves and density neutral intrusions. The features of the sound field are examined in terms of time variability and (oceanic) wave number behavior. Together with an accompanying paper that discusses sound propagation for these conditions, a detailed description of the impact of water column dynamics in a littoral coastal shelf is presented. [Acknowlegment of work supported by the Office of Naval Research, Contract No. N00014-08-1-0455.]

10:40

4aUW8. Sound propagation calculations for dynamic conditions in the East China Sea. Kingsland Michelle, Smith Chad, and Bradley David (Graduate Program in Acoust., Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804-0030, mbm224@psu.edu)

Based on the sound field discussed in detail in a companion paper, sound propagation behavior is calculated using the range-dependent acoustic model parabolic equation method. A number of source/receiver depth combinations are examined, for frequencies in the range of 200–1000 Hz, and for ranges to 30 km. Impact of the temporal and spatial dynamics of this coastal shelf region will be discussed. Of particular interest is the variability produced by both internal waves and intrusive water masses. [Acknowledgment of work supported by the Office of Naval Research, Contract No. N00014-08-1-0455.]

10:55

4aUW9. Temporal coherence of normal modes in an ocean waveguide. T. C. Yang (Naval Res. Lab., Code 7120, Washington, DC 20375)

Temporal coherence is a measure of sound correlation in time, and the temporal coherence time refers to the time it takes for the signal to become uncorrelated. Over 20 experiments have been conducted over the past 40 years to measure the temporal coherence of sound propagation to a long distance. The experimental results in shallow and deep water were recently analyzed by Yang [J. Acoust. Soc. Am. **120**, 2595–2614 (2006); J. Acoust. Soc. Am. **124**, 113–127 (2008); J. Acoust. Soc. Am. **125**, 1247 (2009)]. A theoretical treatment is given in this paper for the temporal coherence of normal modes using the path integral approach and also the perturbation expansion of the coupled normal-mode equation to the first order of sound-speed-perturbation squared. For a typical shallow water and deep water environment with internal waves present, the coherence time of the acoustic field (averaged over the source and receiver depths) decreases as $-3/2$ power of frequency due to mode coupling, as is observed in data. Not including mode-coupling, the coherence time of the acoustic field decreases as -1 power of frequency, same as that predicted for acoustic rays. The coherence time decreases as $-1/2$ power of range in all cases. (Work supported by the Office of Naval Research.)

11:10

4aUW10. Statistics of acoustic intensity fluctuations caused by non-linear internal waves of elevation in shallow water in the South China Sea. D. Benjamin Reeder (Dept. of Oceanogr., Naval Postgrad. School, 833 Dyer Rd., Monterey, CA 93943, dbreeder@nps.edu)

In April 2007, a 6-day long field experiment was conducted to measure internal waves and their acoustic effects at the edge of the continental shelf in the Northern South China Sea. Very large internal solitary waves of depression originating in the Luzon Strait generate trains of internal elevation waves as they encounter a subaqueous bluff at the shelf edge in 120-m water depth. The acoustic transect, from the moored acoustic source transmitting 400-Hz *m*-sequences to two vertical line arrays moored at 3- and 6-km ranges, was oriented roughly parallel to the crests of the passing internal elevation waves, so as to observe the strong horizontal refraction effects which occur under such conditions. Statistics of and correlations between the internal wave characteristics and their associated effects on acoustic propagation are presented. [Work supported by ONR.]

11:25

4aUW11. The effects of surface variability on received signal statistics. Jeremy M. Joseph, R. Lee. Culver, and Alex W. Sell (Grad. Prog. Acoust. & Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA 16804)

Acoustic propagation through the ocean can result in significant variability in the received signal. The degree of variability depends on the signal frequency and source-receiver range as well as medium variability along the propagation path. Our overall research goal is to develop the capability to classify passive sonar signals using the statistics of the received signal, and to support this goal, we require the capability to simulate statistically-valid signals. When the propagation path interacts with a rough ocean surface, the received signal will vary with time. We have employed the rough surface RAM parabolic wave equation code (Rosenberg, J. Acoust. Soc. Am. **105**, 1999) to predict the statistics of low-frequency received signals that interact with the surface (such as in a shallow ocean channel) in order to calculate the effects of the time-varying rough surface scattering on frequency and amplitude statistics. Using a temporally-evolving rough surface, frequency and amplitude statistics are calculated for several ranges and depths. Received signal parameter statistics are compared with published analytical and empirical models. [Work sponsored by ONR Undersea Signal Processing.]

11:40

4aUW12. Effects of non-adiabaticity in modal propagation through random sound speed perturbations. Kevin D. LePage (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

Modal propagation through sound speed fluctuations in shelf environments can be modeled adiabatically or fully coupled through polynomial chaos (PC) expansions of the modal amplitudes or complex phase. The importance of modal scattering versus adiabatic effects is explored for random sound speed fields in the mid-frequency regime. It is shown that the importance of inter-modal scattering is related to the interference wavelength between the modes in relation to the horizontal wavelengths of the sound speed fluctuations as well as the vertical structure of the sound speed fluctuations with respect to the shape of the interacting modes. Modal scattering is also highly affected by the relative amplitudes between the modes with strong scattering into weaker modes. While these effects are well known, the coupled mode PC equations give unique insight into the detailed physics of these controlling factors.

Session 4pAA

Architectural Acoustics: Room Acoustics Potpourri—Absorption, Diffusion, Testing, and Modeling

Pamela J. Harght, Cochair
BAi LLC, 4006 Speedway, Austin, TX 78751

David S. Woolworth, Cochair
Oxford Acoustics, Inc., 356 CR 102, Oxford, MS 38655

Contributed Papers

1:30

4pAA1. Effect of macroscopic parameters on sound absorption and sound transmission loss of porous materials. Paresh Shravage (Electro-acoustics Res. Lab., Dept. of Phys., N. Wadia College, Mangaldas Rd., Pune, India 411001) and Keith de'Sa (N. Wadia College, Pune, India 411001)

Sound absorbing materials are indispensable nowadays for noise control applications. The acoustical behavior of these acoustic materials is governed by five physical (e.g., porosity, flow resistivity, tortuosity, viscous characteristic length, and thermal characteristic length) as well as three mechanical parameters (e.g., Young's modulus, Poisson ratio, and loss factor). The characterization of these porous materials is very crucial as it plays a crucial role in design and development stage itself for predicting acoustic behavior of multilayer porous materials for higher sound absorption and transmission loss. This prediction depends on measurement accuracy of macroscopic physical parameters which are very difficult to measure except porosity and flow resistivity, which is the only standardized test until today; also availability of such rigs is also a problem for manufactures as they are available only at specialized test laboratories. This paper presents effect all of these parameters on sound absorption coefficient and transmission loss of the porous materials using Johnson–Champoux–Allard and Biot model for poroelastic materials. It also discusses results of simulation with effect of each parameter on acoustic behavior of the sound absorbing materials.

1:45

4pAA2. Analysis of a Schroeder diffusor using the Biot–Tolstoy theory. Nicolás Ramírez, Maria Paz Raveau, Osvaldo Ramírez, and Christopher Feuillade (Laboratorio de Acústica, Universidad Tecnológica de Chile-INACAP, Sede Pérez Rosales, Brown Norte 290, Ñuñoa 779-0569, Santiago, Chile, ramamaiden@hotmail.com)

The Biot–Tolstoy (BT) theory for the impulse response of a point source to a rigid wedge [J. Acoust. Soc. Am. **29**, 381–391 (1957)] has been used to investigate the scattering properties of a QRD Schroeder acoustic diffusor. The BT is a fully three dimensional time-domain method, which facilitates individual identification and evaluation of the three components which give rise to the diffused field: reflections from the faces of the diffusor, diffractions from the edges of the wells, and "reflection-diffractions," which combine the previous two phenomena. The computed impulse response readily enables frequency analysis over a broad frequency band with an upper limit determined by the signal sampling rate. Comparisons with published theoretical results for a diffusor at long range, obtained using a boundary element technique, show very good agreement, both for the diffusion coefficient and the polar scattering patterns as a function of frequency. Comparisons of the results of experimental measurements, with BT predictions, for short range diffusor configurations also show promising agreement. The results of both of these tests of the BT method will be presented, and the overall performance assessed.

2:00

4pAA3. Optimizing scattering from stage canopy arrays: determining density, size, shape and position. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com) and Trevor J. Cox (Univ. of Salford, Greater Manchester M5 4WT, United Kingdom)

The reflecting surfaces around a stage area play an important part in enabling ensemble among musicians. Energy must be reflected back to the stage to enable musicians to hear themselves and others and so achieve the correct rhythm, intonation, balance, and timbre. Gade summarized the current understanding of stage acoustic requirements; in particular, when stage reflections should arrive and how loud they should be. The study presented in this paper concentrates on the design of overhead canopy arrays, in particular, investigating effects of density, size, shape, and position. This aspect has surprisingly been little researched, although work by Rindel investigated similar questions to those being posed here. The advent of modern computer processing power allows stage canopy arrays and towers to be optimized using iterative algorithms, based on wave-based acoustic prediction models. A shape optimization program has been developed to optimize both the shape and tilt of overhead canopy arrays. Previously, the height and density have been specified based on experimental evidence and lighting/scenery considerations. The program has recently been updated to address the question of optimal canopy density and size, and some results from this work are presented below.

2:15

4pAA4. Determining the directional scattering coefficient from polar responses. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774, pdantonio@rpginc.com)

A method to measure the random incidence scattering coefficient has been described in ISO 17497-1. The method involves averaging the impulse responses for the sample in different orientations to isolate the specular component of the scattering. As such, the method does not distinguish between samples in which there is depth variation in only one direction, as in a cylinder, and textured surfaces with depth variation in many directions. Since computer model algorithms currently do not distinguish between these two types of surfaces, this may be an acceptable approximation. However, as computer models evolve to distinguish between these types of surfaces, directional scattering coefficients will be required. It is possible to calculate or measure the directional scattering coefficient from scattered polar responses, using either the correlation scattering coefficient described by Mommertz, in which the polar responses of a scattering sample and a flat surface are correlated, or the ratio of the energy in the specular zone and the total energy. Examples of calculated and measured directional scattering coefficients will be presented and discussed.

2:30

4pAA5. The effects of nonideal microphone directivity patterns on directional impulse response measurements. J. James Esplin, Brian T. Thornock, and Timothy W. Leishman (Acoust. Res. Group, Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602)

An acoustician can identify and treat problematic surfaces to reduce or eliminate unwanted reflections only if he knows the origins of those reflections. Several measurement techniques exist for the purpose of identifying these origins, including the Polar ETC method, which requires six cardioid impulse response measurements along Cartesian axes. This presentation will explore two implementations of the method using either a microphone positioner and six sequential cardioid measurements (as originally intended) or four simultaneous measurements from a tetrahedral subcardioid microphone array (originally intended for Ambisonic recordings, but also usable to synthesize the six cardioid measurements). It will compare the two approaches and investigate typical errors introduced by nonideal cardioid directivity patterns. The presentation will also discuss the capabilities of a new method, based on a Cartesian array of seven omnidirectional microphones, and explore the effects of nonideal omnidirectional patterns.

2:45

4pAA6. An inexpensive approach to measure one-third octave reverberation times. Mark H. Holdhusen (Univ. of Wisconsin-Marathon County, 518 S. 7th Ave., Wausau, WI 54401, mark.holdhusen@uwc.edu)

Reverberation time is a common measurement taken in room acoustics. However, the software used in portable sound-level meters can be rather expensive, making it challenging for those on a smaller budget to take these necessary measurements on site. The work here outlines an inexpensive method to instantaneously measure one-third octave reverberation times using a portable experimental setup. The setup consists of a PCB random-incidence microphone along with a National Instruments USB data acquisition system and a notebook PC. Using LABVIEW along with MATLAB software, code was written to process acoustic measurement data to achieve one-third octave reverberation times with almost no additional processing time. The entire setup costs approximately 2000 dollars with educational discounts. This setup and procedure give similar results to the aforementioned sound-level meters, but at a fraction of the cost.

3:00

4pAA7. Deconstructing the interaural cross-correlation function. William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., East Lansing, MI 48824) and Brad Rakerd (Michigan State Univ., East Lansing, MI 48824)

Because it is a measure of the similarity of signals in a listener's left and right ears, the long-term interaural cross-correlation (IACC) function within one-third octave bands provides an estimate of auditory envelopment as well as a guide to the ability to localize a sound. The peak height of the IACC, known as the coherence, is particularly relevant. The IACC can be written mathematically in terms of the amplitudes of spectral components in the left and right ears and in terms of the interaural differences in the component phases. The ensemble-average IACC, as measured with an artificial head in real rooms, can be further analyzed in terms of the amplitudes—their statistical distributions and their interaural correlations—and in terms of the interaural phases—their distributions and correlations with amplitudes. The nature of these statistical relationships depends on the frequency range through the ratio of a characteristic wavelength to the head radius. Analysis of this kind helps solve the long-wavelength problem. In the long-wavelength limit, the IACC must be large, but that does not mean that sound localization is good. [Work supported by the NIDCD Grant DC00181.]

3:15

4pAA8. Time-frequency test signal synthesis for acoustic measurements during music concerts. Gang Ren (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627, garen@ece.rochester.edu), Mark F. Bocko, and Dave Headlam (Univ. of Rochester, Rochester, NY 14627)

When occupied by an audience, a musical performance space exhibits different acoustic characteristics compared to when it is empty. To obtain acoustical measurements with an audience present would require them to sit through an entire session of acoustic test signals, which is not practical. The dynamics of room acoustics parameters during an on-going concert may also be important to the mixing and recording engineers. These acoustic measurement scenarios demand an analysis tool for performing the designated tests during music concert sessions while not being noticeable by the audience. An acoustic measurement technique employing test signals generated during a musical performance is proposed. This method employs an adaptive time-frequency synthesis algorithm that determines the energy vacancies in the spectrogram of the on-going performance and automatically generates low-energy test signals that fill in the vacancies. The energy level of the test signal can be low enough to be masked by the music while high enough to be measurable above the noise floor. Several classes of test signals are proposed and their implementations in live musical events are demonstrated. Different system identification methods for estimating architectural acoustics parameters from the recorded test signals are also discussed and compared.

3:30

4pAA9. Statistical spectrogram modeling and analysis for blind estimation of room acoustics from musical recordings. Gang Ren (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627, garen@ece.rochester.edu), Mark F. Bocko, and Dave Headlam (Univ. of Rochester, Rochester, NY 14627)

An algorithm employing Markov random field modeling has been applied to spectrographic representations of musical recordings to uncover acoustical features of the recording environment. For recorded music, the reverberation pattern is most visible at the onset/offset edges of harmonic components in the spectrogram. Edge features extracted through image analysis algorithms can be mapped to the acoustical features of the recording space, but the correspondence is complicated by the inherent randomness of the musical "test signal." This is exacerbated in the media production process because a final mix can be generated from separate recordings from multiple acoustic spaces. The Markov random field modeling algorithm utilizes data-modeling techniques to estimate the probabilistic links between edge analysis results and room acoustical features and further to identify the probabilistic nature of any time-variation in latent room acoustics. Our algorithm obtained enhanced room acoustics feature extraction performance by allowing a gradual refinement of the room acoustics feature vectors through dynamic fusion of prior estimates and knowledge from current musical segments. This on-line approach is also more computationally efficient compared to batch processing. Multiple variants of the proposed algorithm are demonstrated and compared to existing blind room acoustics parameter estimation methods.

3:45—3:55 Break

3:55

4pAA10. A comparison study between the transport equation model and diffusion equation model for room-acoustic predictions in long spaces. Yun Jing, Ning Xiang (Graduate Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., Troy, NY 12180), and Edward Larsen (Univ. of Michigan, Ann Arbor, MI 48109)

This paper compares two numerical models: diffusion equation model and transport equation model, for room-acoustic predications in long spaces. Mathematically, the diffusion equation is the asymptotic approximation to the transport equation, therefore is expected to be less accurate. In addition, the transport equation intrinsically includes the direct sound field while the diffusion equation does not. However, due to its simple form of the diffusion equation, it has certain computational advantages over the transport equation model, e.g., it is less computationally expensive. When comparing the two models with the ray-tracing based method, the transport equation is shown to be more precise, especially when the walls are highly absorptive. Nevertheless, the diffusion equation model performs fairly well provided that the absorption on the walls is low. In this case, the diffusion equation model might be preferred over the transport equation model due to a low computational load.

4:10

4pAA11. Investigation of the just noticeable difference of the clarity index for music, C80. Meghan J. Ahearn, Matthew J. Schaeffler, Robert D. Celmer, and Michelle C. Vigeant (Acoust. Prog. and Lab., Dept. of Mech. Engr., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117, vigeant@hartford.edu)

The just noticeable difference (JND) of the clarity index for music, C80, has been reported to be approximately 1 dB, but there is limited research to support this value. A subjective study was conducted to verify this JND using a total of 51 musically trained subjects. Test signals were created using digital delays, equalizers, and reverberation-units, and sent out to eight loudspeakers distributed throughout an anechoic chamber. Three motifs and two C80 base-cases were tested: (1) had a C80 of −1 dB (1 kHz) with a 2.1-s reverberation time (RT), while (2) had a C80 of +3 dB (1 kHz) with a 1.6-s RT. Signals were presented in pairs with the first signal being the base-case and the second having a positive difference ranging between 0.5 and 3.0 dB. Control cases with no C80 differences were also presented. Results from all 51 subjects did not reveal a clear relationship between the percentages who heard a difference versus the difference in decibels. However, when the data were filtered to include 17 of the subjects, clearer trends resulted that suggest a higher JND of approximately 1.6 dB. Implications for future research will be discussed. [Work supported by the Paul S. Veneklasen Research Foundation.]

4:25

4pAA12. Evaluating the practical performance and benefits of various loudspeaker styles for party wall and floor/ceiling assembly field sound transmission testing. Marlund E. Hale (Adv. Eng. Acoust., 663 Bristol Av., Simi Valley, CA 93065, mehale@aol.com)

The instrumentation and equipment used in field sound transmission testing of multi-family residence party walls and floor ceiling assemblies have steadily improved. ASTM E336 suggests the use of omnidirectional speakers for such performance tests. When directional box speakers are used, they need to be placed in and face toward a corner of the sound source room. Since many multi-family residential complexes are two stories and higher, the need to move heavy equipment (including massive speakers) from room to room and floor to floor can be a challenge. It is therefore desirable to optimize weight and performance. This paper compares the functionality and performance of the standard directional box-type loudspeaker (90 dB at 1 m, 350-W maximum input, 90 lbs), an omnidirectional dodecahedron spherical loudspeaker (110 dB at 1 m, 480-W maximum input, 30 lbs) and a prototype omnidirectional dodecahedron spherical loudspeaker (120 dB at 1 m, 1000-W maximum input, 35 lbs). In addition to comparing the acoustical output of these different loudspeaker systems, their tendency to introduce vibrations into the source room floor is also discussed.

4:40

4pAA13. Subjective evaluation of low-frequency impact sounds on floor-ceiling assemblies. Bradford Gover, John Bradley, Trevor Nightingale, and Berndt Zeitler (Natl. Res. Council Inst. for Res. in Construction, 1200 Montreal Rd., Ottawa, ON K1A 0R6, Canada brad.gover@nrc-cnrc.gc.ca)

To investigate meaningful ratings and criteria for the acoustical performance of building partitions, it is necessary to understand the subjective judgments of listeners when hearing transmitted or radiated sounds. Impact sounds on floors caused by footstep noise (walking, running, etc.) are particularly bothersome and are not well characterized. A series of recordings of the sound radiated into the room below a floor-ceiling test assembly when an impact occurs on the topside of the assembly (floor surface) is ongoing. The primary impact sources considered are real adult walkers, and a heavy rubber ball, standardized in JIS A 1418-2. The binaural recordings are presented to listening test participants, who rate subjective attributes of the sounds such as annoyance. In this way the floor-ceiling test assembly is rated (and can be ranked) subjectively, and metrics derived from purely physical measurements can be assessed and interpreted. Preliminary results from measurements on a range of lightweight wood framed assemblies are discussed.

4:55

4pAA14. Insertion loss: Train and light-vehicle horns and railroad-crossing sound levels. Raymond M. Brach (Aerosp. & Mech. Eng. Dept., Univ. of Notre Dame, Notre Dame, IN 46556, rbrach@nd.edu) and Matthew Brach (Brach Eng., Granger, IN 46530)

Studies and analyses of sound levels inside road vehicles frequently require estimation of the sound levels of moving or changing positions of the source (such as a train horn) and receiver (road vehicle). The study presented here focuses on two of the components of the sound transmission from the source to the interior of an automobile. One part of this paper is the development of a method to predict the attenuation of the sound level over the path from the source to the vehicle. The method is based on the classical sound decay equation (including variable directivity and theoretical 6-dB drop-off per doubling of distance) but is modified to accommodate different drop-off rates including experimentally measured values. An example using a typical train horn sound power level is provided. The second portion of the sound path covered in this paper is the attenuation of the sound level due to transmission through vehicle bodies and relates to annoyance of vehicle exterior noise. Insertion loss values and insertion loss spectra are measured for seven different light vehicles. An unusual property of insertion loss spectra is observed and studied. It is shown that direct subtraction of measured band-filtered levels can provide misleading overall levels. A method of correction of the spectrum is presented.

Session 4pAB

Animal Bioacoustics: Natural Soundscapes and Auditory Scene Analysis by Animals II: Auditory Scene Analysis by Animals

Richard R. Fay, Cochair
Loyola Univ., Parmly Hearing Inst., 6430 N. Kenmore, Chicago, IL 60626

Micheal L. Dent, Cochair
SUNY Buffalo, Dept. of Psychology, Buffalo, NY 14260

Chair's Introduction—1:30

Invited Papers

1:35

4pAB1. Auditory scene analysis in fish. Richard Fay (Parmly Hearing Inst., Loyola Univ. Chicago, Chicago, IL 60626)

Auditory scene analysis (ASA) is a capacity of auditory perception that results in the hearing out or segregating of sound sources as individual objects or events when heard in the presence of other, simultaneous sources. Fish, humans, starlings, and nonhuman primates have been shown to have this capacity. I argue that any animal that hears probably has this capacity, and that it is not restricted to large-brained animals, nor to the perception of species-specific communication sounds. The demonstrations of ASA in goldfish include the capacity to hear out the frequency of pure tones when heard as a mixture of tones, and the hearing out of two, simultaneous pulse trains based on temporal and spectral differences, analagous to hearing out different musical instruments playing the same note. So far, harmonicity has been ruled out as a factor in ASA in goldfish, and spectral resolution has been shown to be necessary. ASA is a powerful phenomenon of human and animal hearing, and probably determines what we mean by "hearing." ASA probably evolved very early as a fundamental feature of the sense of hearing.

1:55

4pAB2. The complex auditory scene at leks: Female perception of and response to male aggregations and predators in the acoustic Lepidoptera. Michael D. Greenfield and Sylvain Alem (IRBI CNRS UMR 6035, Parc de Grandmont, Univ. François Rabelais de Tours, 37200 Tours, France)

The fundamental mechanisms by which females respond to and choose males based on the parameters of their calling and courtship songs are now rather well understood in many acoustic insects and anurans. However, they remain relatively unexplored in more complex, natural settings, where displaying males may gather in leks of varying size. We addressed several aspects of this problem in an acoustic moth, *Achroia grisella* (lesser waxmoth; Lepidoptera: Pyralidae), where females show response thresholds and preference functions based on the pulse rate and pulse amplitude of male song. We report that females adjust to increasing lek size by lengthening their exploration trajectories while maintaining the same level of evaluation of each male. We then discuss the potential indirect (genetic) benefits that may accrue to females from their behavioral adjustments. The "auditory scene" at leks may include the sounds of natural enemies (echolocation calls of various species of gleaning bats) as well as displaying males. We note how female *A. grisella* discriminates between conspecific males and potential predators, both of which broadcast in the same (ultrasonic) frequency band. We then show how females perceive and respond to the full auditory scene and adjust their evaluation of males when visiting leks at which (synthetic) bat echolocations are heard.

2:15

4pAB3. Birds and models: Not as different as you might think. Christopher B. Sturdy, Michael R. W. Dawson, Lauren M. Guillette, Carly M. Nickerson, Tara M. Farrell, Marisa Hoeschele, Laurie L. Bloomfield, and Isabelle Charrier (Dept. of Psych., Univ. of AB, P-217 Biological Sci. Bldg., Edmonton, AB T6G 2E9, Canada csturdy@ualberta.ca)

For some time, the Sturdy laboratory group has been studying chickadee vocal production and perception using a variety of approaches. These include, among others, bioacoustic analyses of vocalizations, operant conditioning studies, and, more recently, artificial neural networks. This multidisciplinary approach has been very fruitful. The addition of artificial neural networks to the standard empirical approaches has significantly enhanced the understanding of songbird behavior and has provided models of bird operant conditioning behavior, perception, and cognition, allowed the investigation of questions that would be difficult to carry out with animal studies, honed research questions and foci, and has inspired further empirical studies. This talk will provide a longitudinal review of these and related research findings capitalizing on this data-model/model-data interplay. Topics discussed will include models of bird note type perception, models that have directed the formation of hypotheses about important perceptual features in note types, and models that have inspired further empirical studies of note type perception and have been used to explore a classic cognitive phenomenon, peak-shift, in a multimodal, note-type continuum.

2:35

4pAB4. A test of Gestalt grouping principles in frogs: Auditory illusion or sensory bias? Mark A. Bee (Dept. of Ecology, Evolution, and Behavior, Univ. of Minnesota, 1987 Upper Buford Circle, St. Paul, MN 55108, mbee@umn.edu) and Folkert Seeba (Carl von Ossietzky Universitaet Oldenburg, 26111 Oldenburg, Germany)

The study of auditory illusions has been important in elucidating the rules of auditory grouping in humans, which adheres to principles first described by Gestalt psychologists to explain the formation of visual objects. For example, the human auditory system perceptually restores short, deleted segments of speech and other sounds (e.g., tones) when the resulting gaps are filled by a potential masking noise. Results from studies of such illusions suggest that the Gestalt principle of "continuity" allows humans to perceive complete auditory objects in the face of incomplete or degraded acoustic information in the presence of noise. The present study tested the hypothesis that female treefrogs experience the perceptual restoration of discrete pulses in the male sexual advertisement signal when they are deleted and replaced by a potential masking noise. While added noise restored some attractiveness to degraded signals, there was no evidence that the frogs experienced the illusion of perceiving actual pulses that were missing. Instead, the added noise functioned as an acoustic ornament that made some signals more attractive than others as a result of an inherent sensory bias for greater sensory stimulation. Whether such sensory biases themselves adhere to Gestalt principles of auditory grouping remains to be demonstrated.

2:55

4pAB5. Processing of clustered sensory signals in gray treefrogs (*Hyla versicolor*): Behavioral responses in the laboratory and field. Kevin W. Christie (Neurosci. Program and Beckman Inst., Univ. of Illinois at Urbana-Champaign, 405 N. Matthews Ave., Urbana, IL 61801, kwchrist@illinois.edu), Johannes Schul (Univ. of Missouri-Columbia, Columbia, MO 65211), and Albert S. Feng (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Anuran mating choruses may consist of hundreds of sympatric and conspecific males calling in addition to other biotic and abiotic noise, representing a complex auditory scene whose characteristics change with distance. Detecting and localizing individual males in the chorus poses a challenge to the female, especially at long ranges. We carried out a series of recordings and behavioral experiments that are aimed at determining the distance at which female treefrogs (*Hyla versicolor*) use the chorus as an acoustic cue for their orienting responses. We analyzed natural choruses recorded at distances of 1–100 m and used these to create stimuli for phonotaxis experiments evaluating how the chorus attracts and provides cues for localization by females. In addition, we performed phonotaxis tests in the field near natural choruses to compare the frog's behavior under laboratory and field conditions. We found that females utilize acoustic cues for orientation toward choruses at a distance up to 50 m, but additional cues may be used to ensure orientation in the field at distances $>$50 m. Additional experiments revealed that temporal coherence of the male call, as opposed to a general reduction in stimulus intensity, was the most important factor in female phonotactic behavior.

3:15

4pAB6. Behavioral measures of auditory streaming in humans and ferrets. Christophe Micheyl (Dept. of Psych., Univ. of Minnesota, N640 Elliott Hall, Minneapolis, MN 55455, cmicheyl@umn.edu), Ling Ma, Pingbo Yin, Shihab Shamma (Univ. of Maryland, College Park, MD 20742), Andrew J. Oxenham, and (Univ. of Minnesota, Minneapolis, MN 55455)

The perceptual organization of sound sequences into auditory streams, or "auditory streaming," is an essential component of the analysis of acoustic scenes by humans and other animals. The development of behavioral paradigms for measuring auditory streaming in human and non-human listeners is an important step toward a better understanding of the similarities and differences in sound perception between humans and other species. In this talk, we will present the results of recent studies performed in collaboration between our groups at the universities of Minnesota and Maryland concerning performance-based measures of auditory streaming in humans and ferrets. In particular, we will describe two complementary temporal discrimination tasks, one of which can be used to induce and measure stream segregation, while the other can be used to promote and measure stream integration in a listener. Because these tasks use similar stimuli, they are especially suitable for use in studies of neural correlates of streaming percepts without stimulus confounds. In addition, we will describe the results of behavioral measures of auditory streaming and informational masking in ferrets, using stimuli and tasks that were directly inspired by psychoacoustical experiments in humans. [Work supported by NIDCD R01 DC 07657.]

3:35—3:50 Break

3:50

4pAB7. Network interactions and the emergence of representations for natural communication signals. Timothy Gentner (Dept. of Psych., UC San Diego, 9500 Gilman Dr., MC 0109, La Jolla, CA 92093, tgentner@ucsd.edu)

Current understanding of experience dependent plasticity in sensory cortices turns on the notion that as signals acquire behavioral relevance, they elicit increasingly robust neural responses in single cells and/or drive increasingly large neural populations. Similar facilitative effects are seen in the songbird auditory forebrain region CMM when birds learn to recognize conspecific songs. Here, I describe a discrete population of neurons in the adjacent auditory region NCM where the opposite occurs. Following song recognition training, learned songs elicit significantly lower extracellular responses from single NCM neurons than do unfamiliar songs. Moreover, this effect is directly tied to associative learning, as control songs presented in the absence of behavioral contingencies also elicit robust responses. These findings are consistent with an active role for (1) broad scale response suppression and (2) contributions from irrelevant signals in shaping the learned sensory representation for natural, spectro-temporally complex, acoustic communication signals.

4:10

4pAB8. Dendritic mechanisms of auditory stream segregation in an insect. Johannes Schul (Biological Sci., Univ. of Missouri, 207 Tucker Hall, Columbia, MO 65211, schulj@missouri.edu)

Auditory stream segregation (AudSS) is the process of grouping the auditory scene into coherent perceptual objects. We recently described AudSS in an insect: In Neoconocephalus katydids, a first-order auditory interneuron (TN-1) segregates bat cries with slow pulse rates from conspecific communication signals with fast pulse rates. This insect system provides a powerful opportunity to study neuronal mechanisms of AudSS in an identified neuron with testable behavioral function. TN-1 adapts rapidly to fast pulse rates and responses are completely suppressed. However, if a slow pulse rate of sufficiently different carrier frequency is presented simultaneously, TN-1 responds reliably to the slow rate pulses. The suppression of the fast pulse rate on TN-1, i.e., the critical component of AudSS, is caused by two dendritic processes intrinsic to TN-1. The first mechanism is transient and likely calcium mediated, while the second mechanism is sustained throughout the stimulation and likely sodium mediated. TN-1 responses to slow pulse rates, occurred when the slow and fast pulse rate were transmitted by different groups of receptor cells, which project onto different regions of the TN-1 dendrite. Based on these results, we propose a mechanism of dynamic dendritic compartmentalization to underlie AudSS in the katydid TN-1.

Contributed Papers

4:30

4pAB9. Range is pitch, shape is pitch: How much of biosonar is adapted from passive hearing? James A. Simmons (Dept. of Neurosci., Brown Univ., Box G-LN, Providence, RI 02912, james_simmons@brown.edu)

Echolocating big brown bats perceive target range from overall echo delay. Neuronal networks acting as delay-lines retain the volley of spikes evoked by each broadcast for coincidence-comparisons with volleys of spikes evoked by echoes. These circuits are similar to delay-line /coincidence-detectors proposed for determining periodicity pitch of unresolved frequency components. Bats use resolved frequency components to detect harmonic coherence as the criterion for deciding whether multiple-harmonic sounds are true echoes. These circuits involve FM1-FM2 facilitatory neurons in circuits similar to what has been proposed for spectral pattern matching based on detection of coincidences of spikes at harmonic frequencies, another route for extracting periodicity pitch. Additionally, bats perceive small range separations between target glints from the frequencies of spectral notches produced by interference between overlapping reflections at small time offsets. This psychophysical effect is similar to determining the pitch of comb-filtered noise using neuronal circuits for identifying modulation-rates of rippled spectra. Neuronal circuits for detecting nulls at specific frequencies in echoes yield responses tuned to specific glint separations in time. Bats may use the different mechanisms proposed for determining the pitch of pulse trains, complex sounds, and rippled noise to perceive integrated images of targets. [Work supported by ONR and NIMH.]

4:45

4pAB10. Auditory streaming of birdsong in budgerigars and zebra finches. Kristen A. Garcia (Dept. of Psych., Univ. at Buffalo - SUNY, 206 Park Hall, Buffalo, NY 14260, kgarcia@buffalo.edu), Siddarth Rajaram (Dept. of Cognit. and Neural Systems, Boston Univ., Boston, MA 02215), Kamal Sen (Dept. of Biomedical Eng., Boston Univ., Boston, MA 02215), Barbara G. Shinn-Cunningham (Dept. of Cognit. and Neural Systems, Boston Univ., Boston, MA 02215), and Micheal L. Dent (Dept. of Psych., Univ. at Buffalo-SUNY, Buffalo, NY 14260)

The ability to segregate or group together sounds accurately in the environment is vital to the survival of all animals and should be highly conserved across many different species. To date, evidence of auditory streaming has been shown in a wide range of organisms, from insects to birds to primates, but these studies have usually used simple stimuli emitting from only one location. The current study utilizes categorization techniques to determine the role of intensive, spatial, temporal, and spectral cues on streaming in budgerigars (*Melopsittacus undulatus*) and zebra finches (*Taeniopygia guttata*). Both species were trained to differentially peck keys in response to either a synthetic zebra finch song consisting of five syllables (whole song) or to the same song with the fourth syllable omitted (broken song). Probe trials were then inserted, where possible cues for streaming were tested by altering characteristics of the fourth syllable. Preliminary results showed that the birds were most likely to report hearing a whole song when the fourth syllable in the probe trials was played closer in space, at the same intensity, and with the same spectral characteristics as the rest of the song, suggesting that these cues are important for auditory streaming in birds.

5:00

4pAB11. Acoustic and perceptual categories for vocal signals in budgerigars (*Melopsittacus undulatus*). Hsiao-Wei Tu (Dept. of Psych., Univ. of Maryland, College Park, MD 20742, htu@psyc.umd.edu), Michael S. Osmanski (Dept. of Biomedical Eng., Johns Hopkins Univ., 720 Rutland Ave., Baltimore, MD 21205), and Robert J. Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742)

The budgerigar has an extraordinarily complex, learned, vocal repertoire consisting of both long rambling warble and a number of shorter calls produced as single utterances. The contact call is the most frequently produced vocalization by budgerigars, and it bears a strong resemblance to short (about 150-ms), frequency-modulated elements that make up 25% of the sounds of warble. Using signal processing and psychophysical techniques we examined whether these two vocalizations represent different phonological systems or whether contact calls are used in two different situations. Warble call elements and contact calls were recorded from four male budgerigars. For each bird, contact calls produced as single utterances were different from contact call-like warble elements on a number of acoustic dimensions. Psychophysical tests showed that contact calls and call-like warble elements represent distinct perceptual categories. Finally, comparison of percent correct and response latencies across tests indicates that birds perceive the similarity between contact calls and call-like warble elements produced by the same bird. In sum, these birds show perceptual categories that correspond to acoustic categories in these complex vocalizations and they perceive the similarity in vocalizations from different acoustic categories produced by the same bird. [Work supported by NIH/NIDCD R01DC000198 to R.J.D.]

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration, Physical Acoustics, and Underwater Acoustics: Multiple Scattering of Waves: From Theory to Application II

Guillaime Haiat, Cochair
CNRS, LRO 10, Ave de verdun, 75010 Paris, Cedex, France

Purnima Ratilal, Cochair
Northeastern Univ., Electrical and Computer Engineering, 360 Huntington Ave., Boston, MA 02115-5000

Invited Papers

1:30

4pBB1. Effect of multiple scattering and modal dispersion on population density imagery of Atlantic herring shoals with ocean acoustic waveguide remote sensing. Mark Andrews (Notheastern Univ., Dept. of Elec. and Comput. Eng., 360 Huntington Ave., Boston, MA 02115)

Wide area sonar such as ocean acoustic waveguide remote sensing (OAWRS) has been shown to be a useful tool for instantaneously imaging large shoals of fish distributed over continental shelf regions. Here, we show that the population density of fish groups can be accurately estimated by using incoherent averaging of the matched filtered returns. Numerical Monte-Carlo models are applied to simulate the active imaging system and determine the statistics of the received matched filtered intensity scattered off Atlantic herring shoals in the Gulf of Maine. The model includes multiple scattering of the dense fish groups and uses a range-dependent acoustic model to simulate modal dispersion in the fluctuating ocean environment. We illustrate the effects of modal dispersion on the sonar imagery for various seafloor types. We determine the conditions for when multiple scattering is significant for Atlantic herring shoals, which depends on the fish density, imaging frequency, and mean target strength. Results of the model are compared to OAWRS sonar imagery collected during the 2006 Gulf of Maine Experiment.

1:55

4pBB2. Multiple scattering through clouds of gas bubbles in liquids: Accounting for clustering in the higher-order moments. Thomas C. Weber (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824), Anthony P. Lyons, and David L. Bradley (The Penn State Univ., State College, PA 16804)

The presence of bubble clouds can change the sound speed and attenuation in liquids, both of which are often treated as first moment quantities that can be derived using a multiple scattering approach for an effective medium. The same multiple scattering framework can be used to examine higher-order statistical moments, accounting for the large variability that is often observed in acoustic signals in the presence of bubble clouds. When clustering of bubbles (preferred concentrations) is present, multiple scattering theory shows that the statistical moments should change. These cluster-induced, frequency dependent changes in sound speed, attenuation, and higher-order moments can be examined both from a theoretical standpoint and with the aid of numerical simulations. Simulations have shown that while point measurements of the acoustic field for non-clustered bubble clouds appear normally distributed, clustering introduces both a non-zero skewness and a higher than expected kurtosis, depending on the number density of the bubble clouds and the nature of the clustering. This type of behavior is expected to be present both for forward scattering and backscattering scenarios.

Contributed Papers

2:20

4pBB3. Hamiltonian formalism and cellular algorithm for modeling large numbers of interacting bubbles. Derek C. Thomas, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., The Univ. of Texas at Austin, Austin, TX 78713-8029)

Large, dense clusters of bubbles play an important role in the efficacy of shock-wave lithotripsy. Although methods exist for simulating bubble interaction dynamics within a cluster, for moderately sized clusters ($\sim$1000 bubbles), they are computationally expensive and scale poorly to larger clusters. Currently under development is a method to reduce the computational complexity of the problem and permit direct simulation of large bubble clusters. The algorithm functions by partitioning the cluster into small cells. Bubble interaction within each cell is assumed to be instantaneous and is modeled using a Hamiltonian formulation [Ilinskii *et al.* J. Acoust. Soc. Am. **121**, 786 (2007)] that has been extended to include a correction for radiation damping. The cellular sub-clusters are coupled via the pressure field they generate, and the time delays associated with finite acoustic propagation speed are taken into account. Use of the Hamiltonian formulation reduces computational complexity and numerical error for clusters containing large numbers of bubbles, and it is thus well suited for application to lithotripsy. The accuracy of the method will be validated by comparison with results obtained from an existing model obtained using Lagrangian formalism. [Work supported by the ARL:UT McKinney Fellowship in Acoustics and NIH DK070618.]

2:35

4pBB4. Scattering and radiation force of helicoidal Bessel beams incident on spheres and the movable rigid sphere case. Philip L. Marston (Phys. and Astronomy Dept., Washington State Univ., Pullman, WA 99164-2814, marston@wsu.edu)

Understanding the scattering properties and coupling to modes of spheres in beams gives insight into several issues. Beam parameters may be selected to suppress or highlight the coupling to specific modes. When considering the radiation force on spheres by acoustic beams, it is also helpful

to remember that while the axial wave-vector component has similarities to propagating waves, the transverse components may have similarities to standing waves. The present work examines the scattering and radiation force for the simplest Bessel beam having an axial null [P. L. Marston, J. Acoust. Soc. Am. **125**, 3539–3547 (2009); F. G. Mitri, J. Phys. A: Math. Theor. **42**, 245202 (2009)]. This is an acoustic vortex beam having a unit magnitude topological charge m. For some elastic and fluid spheres such beams allow conditions giving negative as well as positive axial forces. Negative forces are found to be correlated with reduced backward hemisphere scattering (as previously reported in the $m=0$ case). A noteworthy example is the case of a movable rigid sphere since for low ka the scattering properties are similar to metallic objects in water and the elastic modulus only weakly affects the scattering and radiation force. [Work partially supported by ONR.]

THURSDAY AFTERNOON, 29 OCTOBER 2009 LIVE OAK, 1:40 TO 3:05 P.M.

Session 4pED

Education in Acoustics: Historical Teaching and Research Acoustic Apparatus

Andrew A. Piacsek, Cochair
Central Washington Univ., Dept. of Physics, 400 E. University Way, Ellensburg, WA 98926-7422

James M. Sabatier, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, 1 Coliseum Dr., University, MS 38677

Chair's Introduction—1:40

Invited Papers

1:45

4pED1. The acoustical apparatus of Rudolph Koenig. Thomas B. Greenslade, Jr. (Dept. of Phys., Kenyon College, Gambier, OH 43022, greenslade@kenyon.edu)

Rudolph Koenig was the pre-eminent maker of acoustic apparatus for research and instruction in the last third of the 19th century. His apparatus is still to be found in lecture demonstration collections in many physics departments in North America. Much of the apparatus that he brought to Philadelphia to display at the Centennial Exhibit 1876 remained on these shores, with part going to the United States Military Academy at West Point, and the remainder to the Physics Department at the University of Toronto. I have visited these collections and many others to photograph many examples of the apparatus that he designed and sold to generate and analyze acoustic signals.

2:10

4pED2. Chladni patterns to holograms: Apparatus for modal analysis of vibrating plates. Andrew Morrison (Phys. Dept., Northwestern Univ., 2145 Sheridan Rd., Evanston, IL 60208) and Thomas Rossing (Stanford Univ., Stanford, CA 94305)

Since Chladni popularized the use of powder patterns in the 18th century, they have been widely used to display and analyze the modes of membranes, plates, and other vibrating structures. Although Chladni patterns have been largely replaced by optical and computer methods for modal analysis, most physics apparatus shelves still house one or more sets of "Chladni plates." We present a brief history of modal analysis of vibrating structures and describe apparatus used by various teachers and researchers, past and present.

Contributed Papers

2:35

4pED3. The Rubens tube. Kent L. Gee (Dept. of Phys. and Astronomy, Brigham Young Univ., N243 Eyring Sci. Ctr., Provo, UT 84602, kentgee@byu.edu)

In 1905, Heinrich Rubens and Otto Krigar-Menzel published a paper describing a unique acoustics teaching apparatus. They developed a flammable gas-filled tube with holes in the top that revealed the acoustic standing wave behavior inside the pipe via the height of flames above the pipe. Interestingly, their article holds the distinction of being printed immediately following Einstein's Nobel-prize winning paper on the photoelectric effect. From that auspicious beginning, the "Rubens tube" has been used for over a century in the teaching of acoustical resonance behavior. This paper will describe the tube's operation, examples of its use, and commentary on and investigations involving the flame tube found in the literature.

2:50

4pED4. The Harvey Fletcher years at Brigham Young University. Kent L. Gee, William J. Strong, and Scott D. Sommerfeldt (Dept. of Phys. and Astronomy, Brigham Young Univ., Provo, UT 84602, kentgee@byu.edu)

Following his long and illustrious career at Bell Laboratories, Harvey Fletcher retired to Provo, UT and continued his career at Brigham Young University. Building on his research at Bell Labs, Fletcher designed and constructed a large anechoic chamber at BYU and also pursued numerous research projects with students at the university. These projects often required the development of research equipment that was very innovative in its time, much of which is still in existence today. This paper will summarize some of the research carried out during those years, and the measurement equipment and apparatus developed by Fletcher and his students to obtain the desired measurements.

Session 4pPA

Physical Acoustics: Materials Characterization and Related Topics

Joseph R. Gladden, Chair
Univ. of Mississippi, Dept. of Physics, 1 Coliseum Dr., University, MS 38677

Contributed Papers

2:15

4pPA1. Ultrasonic diffraction grating spectroscopy: Variables affecting the penetration depth. Margaret S. Greenwood (Pacific Northwest Natl. Lab., P.O. Box 999, MS K5-26, Richland, WA 99352, margaret.greenwooe @pnl.gov)

Initial measurements of the penetration depth were presented at the ASA meeting in Portland. The grating unit is in contact with a slurry and send and receive transducer fastened to the back surfaces, as described previously [M. S. Greenwood, POMA **4**, 045017 (2008)]. At the critical frequency, a peak appears in the signal to the receive transducer. Attenuation is observed by comparing the peak height for the slurry with that for water. The penetration depth is defined as the perpendicular distance from the diffraction grating where the attenuation occurs. The measurement uses a vessel in which the grating surface is in a horizontal plane and in contact with a slurry of polystyrene spheres. When the mixer is turned off, the particles start to fall and the peak in the receive transducer increases until it has the same size as for water. By recording the time for the change in signal and determining the terminal speed, the penetration depth is determined. Data will be presented to show how the frequency and the number of grooves insonified affect the penetration depth. Results will be compared with the theory for the interference of N wavelets from N grooves.

2:30

4pPA2. Monitoring stress in materials using diffuse ultrasonic backscatter. Chris Kube (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, W317.4 Nebraska Hall, Lincoln, NE 68588-0526, ckube@huskers.unl.edu), Goutam Ghoshal (Univ. of Illinois, Urbana, IL 61801), and Joseph A. Turner (Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0526)

Monitoring stress in structures is very important for various safety reasons. One of the challenging problems is to estimate stress nondestructively in various structural components due to changes in loading conditions so that accurate maintenance can be taken before failure. In this presentation, a non-destructive method of monitoring and measuring uniaxial tensile stress in metals using diffuse ultrasonic backscatter is discussed with a focus on the experimental implementation. Ultrasonic backscatter occurs when incident ultrasonic energy is scattered from heterogeneities (in this case grain boundaries) in the material. When an applied load is introduced, third-order elastic behavior of the grains results in changes to the received backscattered energy. Uniaxial loading experiments on samples of steel, aluminum, and a magnesium alloy highlight the utility of this approach for monitoring structural components under stress. Different scattering modes are shown to have different sensitivities to the applied load. Specific application of this technology for quantification of rail stress is also discussed. These results are anticipated to impact ultrasonic nondestructive evaluation of heterogeneous media under stress. [Work supported by US FRA.]

2:45

4pPA3. Two-dimensional display of nonlinearity parameter *B/A* automatically measured with focused Gaussian beam. Shigemi Saito (School of Marine Sci. and Tech., Tokai Univ., Orido, Shimizu-ku, Shizuoka 424-8610 Japan, ssaito@scc.u-tokai.ac.jp) and Jung-Ho Kim (GW Corp., Shinjuku-ku, Tokyo 169-0072 Japan)

A liquid-like sample to fill the 1-mm gap between an acoustic window of polystyrene plate and a tungsten reflector is set in water within the focal region of an 18.6-MHz focused Gaussian beam, whose width is 0.2 mm in the sample. The sound speed is measured with the time of flight. The density is obtained from the reflection coefficient at the interface between the polystyrene and sample. Emanating a dual frequency wave of fundamental and its double frequencies, two attenuation coefficients are estimated from the amplitudes at those frequencies in the sound transmitted through the sample. The magnitude of velocity dispersion is also derived from the relative phase difference of two frequency components. Based on the finite amplitude method, the nonlinearity parameter B/A is finally determined with the second harmonic amplitude. All the process including the setting of the reflector surface on the focal plane is automatically conducted in 5 s with LABVIEW programs. Scanning the beam over a 3×3-mm^2 area, the measured B/A is displayed in a two-dimensional image of 256 segments.

3:00

4pPA4. Structural transitions in self-healing materials utilizing resonant ultrasound spectroscopy. Austin Ricci, Kenneth A. Pestka, II (Dept. of Phys., Rollins College, Winter Park, FL 32789), and Stephen J. Kalista, Jr. (Washington and Lee Univ., Lexington, VA 24450)

Poly(ethylene-co-methacrylic acid) copolymers are among a unique class of materials that can self-heal after exposure to extreme environmental stimulus, such as impact from a bullet. Recent experiments on these copolymers indicate that two distinct time scales govern the evolution of the healing process: a melt state, immediately after the damage, and a slowly evolving polymer interdiffusion. A unique approach will be presented for characterizing the time-scale of the interdiffusion state via observation of the system's resonant spectrum.

3:15

4pPA5. Elastic moduli of single crystal palladium hydride near the tri-critical point. Joseph R. Gladden, III (Dept. of Physics, Univ. of Mississippi, University, Mississippi 38677)

Palladium has long been studied as a model metal hydride system because of the high absorption rates and levels of hydrogen. Even moderately high pressures and temperatures can result in almost one to one hydrogen to palladium atom ratios. Depending on the ambient temperature and hydrogen pressure, the structural phase of hydrogen atoms in the palladium lattice can have three different forms: alpha, beta, and a mixed phase (alpha + beta). There exists a triple point in the pressure versus temperature phase diagram near 290° C and 400 psi. A recent model for absorption/desorption hysteresis in metal hydride systems predicts a sharp drop in the shear modulus near this triple point. Results from recent elastic constants and acoustic attenuation measurements as a function of both temperature and pressure near the triple point will be shown [R. B. Schwarz and A. G. Khachaturyan, Acta Mater. **54**, 313 (2006)]. The experimental procedures for resonant ultrasound spectroscopy measurements at high temperatures and pressures will be also discussed.

3:30—3:45 Break

3:45

4pPA6. Stress dependent diffuse ultrasonic backscatter coefficient for polycrystalline media. Joseph A. Turner (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, W317.4 Nebraska Hall, Lincoln, NE 68588-0526, jaturner@unl.edu) and Goutam Ghoshal (Univ. of Illinois, Urbana, IL 61801)

The diffuse ultrasonic backscatter coefficient (BSC) is a fundamental quantity associated with scattering in heterogeneous media. During a pulse-echo scattering experiment, the energy that scatters back to the source transducer is proportional to the BSC. Theoretical models of the BSC can often be derived from first principles if the heterogeneities can be appropriately modeled. In this case, the statistical properties of the microstructure can be determined in closed form. Polycrystalline materials, such as steel or aluminum, are one class of materials for which the BSC can be determined, typically by assuming that the material is in a stress free state. In this presentation, this fundamental approach is extended to the case in which an applied stress is added to the material. The nonlinear behavior of the grains is included in the analysis such that dependence of the BSC on stress is derived. The stress dependent BSC for different scattering modes relevant to specific experimental configurations is also discussed. Numerical results are presented for materials of common interest and for which single crystal third-order elastic data are available. [Work supported by US FRA.]

4:00

4pPA7. Sensitivity analysis of measurements associated with nonlinear material constants for piezoelectric crystals. Haifeng Zhang (Dept. of Eng. Tech., Univ. of North Texas, 3940 N. Elm St., Denton, TX 76207, haifeng.zhang@unt.edu) and Joseph A. Turner (Univ. of Nebraska, Lincoln, NE 68588)

A robust understanding of nonlinear effects associated with langasite single crystals (electroelastic effect, force-frequency effect, acoustoelastic effect, etc.) is critical for the design of numerous acoustic devices. These nonlinear effects are closely related to the nonlinear material constants. It is common for these constants to be determined through measurements of force-frequency and electroelastic effects or through measurements of ultrasound wave speed changes with respect to external loads/electrical fields. However, due to the complexity of the anisotropy of langasite single crystals, a careful selection of samples with an optimal design is critical for the accurate measurements of these material constants. In this presentation, the first quantitative analysis for the sensitivity of these measurements with respect to extraction of the nonlinear material constants is discussed as applied to langasite single crystals. The sensitivity of each of the third-order elastic and piezoelectric constants to the force-frequency, electroelastic, and acoustoelastic effects of langasite resonators is analyzed. The analysis provides information with respect to the measurement uncertainty for each constant given a set of resonator cuts such that an optimal set can be chosen. These results provide an important reference for the determination of nonlinear material constants of langasite and other piezoelectric materials.

4:15

4pPA8. Granite rock acoustic measurements of nonlinearity and slow dynamics recovery. Claes Hedberg and Kristian Haller (Blekinge Inst. Tech., 371 79 Karlskrona, Sweden, claes.hedberg@bth.se)

Through monitoring the resonance frequency of a granite bar, the nonlinearity and the slow recovery dynamics have been investigated. Nonlinearity was measured through a constant strain protocol, and slow dynamics was measured through a conditioning and relaxation cycle protocol. It is noteworthy that both tests show the same behavior—an increase in resonance frequency with strain for low amplitudes, which at higher strains turns into the normally expected decrease with strain.

4:30

4pPA9. Acoustic parameter estimates and confidence intervals for gravel at low frequencies. Michael J. White (US Army Engineer Res. Development Ctr., P.O. Box 9005, Champaign, IL 61826, michael.j.white@usace.army.mil), George W. Swenson, Jr., Todd A. Borrowman, and George Z. Gertner (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Rigid grain materials such as gravel offer the possibility of achieving low impedances and moderate attenuation for low-frequency acoustic waves. Using a standing wave impedance tube, we have estimated acoustic properties (wavelength, absorption coefficient, acoustic resistance, and reactance) for bulk compressional waves within and above gravel layers of several depths. In this paper we develop confidence intervals for the parameter estimates obtained by fitting measured data to a plane wave model using complex Fourier pressures. Comparisons to the three microphone method and microstructural impedance models are also made.

4:45

4pPA10. Experimental results of a two-dimensional sonic device. Daniel M. Bock, Keehoon Kim, Martin Lopez, and Bryan Molina (Physical Optics Corp., 20600 Gramercy Pl., Bldg. 100, Torrance, CA 90501)

The authors present the design and experimental results of a meta-material multilayered structure in a two-dimensional (2-D) matrix described by Torrent and Sánchez-Dehesa (2007) which can be applied to the fabrication of acoustic lenses and cloaks by tailoring the material anisotropy. The structure consists of a 2-D matrix made of two isotropic materials (low-density foam and aluminum) that are evenly spaced in air with varying diameters in order to control the acoustic impedance. We simulated a 2-D matrix in order to design an acoustic lens with a focal length of 18 cm and a gain of 13 dB in a finite element analysis program. Based on these simulation results, we designed and fabricated a 2-D acoustic lens. We measured the properties of the acoustic lens, having a focal length of 17 cm and a focal point gain of 14 dB (*re* 20 μPa). The excellent agreement between the simulation and experimental results supports the use of a 2-D meta-material structure for the control of acoustic signals in any arbitrary application and can be used for the design of more complex structures in the future.

Session 4pSA

Structural Acoustics and Vibration and Physical Acoustics: Structural Acoustics for Poroelastic Materials

Noureddine Atalla, Chair
Univ. of Sherbrooke, GAUS, Mechanical Engineering, Sherbrooke, QC, J1K 2R1, Canada

Invited Papers

1:30

4pSA1. Inverse characterization of poro-elastic materials based on acoustical input data. J. Stuart Bolton and Kwanwoo Hong (Ray W. Herrick Labs., Purdue Univ., West Lafayette, IN 47907-2031, bolton@purdue.edu)

It has become popular to estimate the Biot parameters of poro-elastic media by using inverse methods. That is, measurements of absorption coefficient and/or transmission loss are used as inputs to an optimization procedure that adjusts the Biot parameters to match the measurements. Software packages are available to perform these operations. To be successful, the models used to match the data must accurately represent the physics of the measurement. However, existing software are based on transversely infinite, layered systems. Such models cannot represent data measured in standing wave tubes when the elasticity of the edge-constrained sample is significant. Here, an approach based on a finite element model of the sample in the tube will be described. It will be demonstrated that this approach makes it easier to estimate the stiffness parameters of a porous sample. Further, a technique involving singular value decomposition of a sensitivity matrix derived from a linearized representation of the absorption coefficient and transmission coefficient near the solution point will be introduced. This procedure gives a clear indication of how many parameters can be estimated from a given data set and gives guidance as to the type of data and the frequency ranges that give the best results.

1:55

4pSA2. Indirect acoustical characterization of sound absorbing materials. Raymond Panneton and Yacoubou Salissou (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, Sherbrooke, PQ J1K 2R1, Canada)

This paper discusses methods to evaluate the dynamic properties of sound absorbing materials (fibers, foams, etc.) based on impedance tube measurements, and the use of indirect methods to retrieve the main macroscopic material parameters from the dynamic properties. For the methods to be successful, the measured dynamic properties need to follow accurately an equivalent fluid behavior and show little noise, particularly at low frequencies. To improve the quality of these measurements, a modification is proposed to an existing three-microphone transfer matrix approach. The modified approach can also be applied to deduce with precision the transmission loss of homogeneous and non-homogeneous symmetrical or non-symmetrical samples. A comparison of the methods and the obtained macroscopic properties (porosity, tortuosity, static viscous and thermal permeabilities, and characteristic lengths) is discussed.

2:20

4pSA3. Acoustic radiation of a vibrating wall covered by a porous layer: Transfer impedance concept and effect of compression. Nicolas Dauchez, Olivier Doutres, and Jean-Michel Genevaux (Laboratoire Acoustique, UMR CNRS 6613, Univ. du Maine, 72095 Le Mans Cedex 9, France, nicolas.dauchez@univ-lemans.fr)

The acoustic radiation of a vibrating wall covered by a porous layer is investigated. The porous layer is described by the one-dimensional Biot model: it accounts for the propagation of the two longitudinal waves in the porous layer within its thickness. The transfer impedance concept [Doutres *et al.*, J. Acoust. Soc. Am. **121**, 206–213 (2007)] allows to determine the effect of the porous layer on the acoustic radiation of the vibrating wall. This impedance differs from the surface impedance that can be measured in a plane wave impedance tube. It is shown that the radiation efficiency of the structure increases in the vicinity of the first resonance of the skeleton in its thickness and decreases for higher frequencies. Experimental validation is performed with a baffled piston covered by a foam or a fibrous layer. Only the radiation model using the transfer impedance shows close agreement with experimental data. Finally, effect of compression of the porous layer is investigated by both analytical and experimental approaches, in the case of fibrous materials used in aeronautical applications.

2:45

4pSA4. Modeling noise control treatments attached to sandwich-composite panels. Haisam Osman (United Launch Alliance, 9100 E. Mineral Circle, Centennial, CO 80112) and Noureddinne Atalla (Univ. de Sherbrooke, Sherbrooke, PQ J1K 2R1, Canada)

This paper discusses the air-borne and structure-borne insertion losses of acoustic materials attached to sandwich-composite panels. The main structure is modeled using an analytical model using discrete layer theory. In this model, all layers of the panel are assumed orthotropic, and both symmetric and anti symmetric modes of propagation are accounted for. For both air-borne and structure-borne excitations, the transfer matrix method with finite size effect correction is used to account for the effect of the noise control treatment. Moreover, power balance is used to highlight the effect of the acoustic treatments in terms of added damping. A three dimensional finite element modeling based on the mixed (u,p) formulations for poroelastic materials and a sandwich element for the structure is presented and used to validate the model. Several examples including sandwich composite panels and various noise control treatments (foams versus fibers, effect of a resistive screens, etc.) are presented and compared.

3:10—3:25 Break

3:25

4pSA5. Investigation of airborne versus structure-borne insertion loss of acoustic materials attached to a flat panel. Dilal Rhazi and Noureddine Atalla (GAUS, Dept. of Mech. Eng., Univ. de Sherbrooke, Sherbrooke, Qubec J1K 2R1, Canada)

This paper discusses methods to predict airborne and structure-borne insertion loss of acoustic materials (fiber, foams, etc.) in single wall and double wall configurations attached to various structures. Specifically, a transfer matrix method based approach is proposed and compared to finite element and statistical energy analysis based predictions. Based on several numerical and experimental results, it is shown first that this method presents a simple and general model that can handle easily and automatically practical configurations. Next, a systematic comparison of structure-borne versus airborne excitation is presented for various structures and sound packages. In particular, the effect of damping and nature of coupling between the sound package and the main structure is discussed.

3:50

4pSA6. Modeling porous-elastic materials in periodically mount connected double panel partitions. Julien Legault and Noureddine Atalla (Dept. of Mech. Eng., Universit de Sherbrooke, 2500 boul. de l'Universit, Sherbrooke, PQ J1K 2R1, Canada, julien.legault @usherbrooke.ca)

In aeronautic double panel partitions, the trim panel (interior panel) is usually attached to the ring frames that are ribbed to the skin panel (exterior panel) with periodically spaced resilient mounts. Moreover, porous-elastic materials are inserted in the cavity between these panels to provide absorption. Due to their periodic nature, these partitions allow the use of space-harmonic expansion formulations for their vibration and sound transmission analysis. This paper shows how equivalent fluid and porous-elastic models can be incorporated in such formulations to assess the influence of sound absorbing materials. The equations of the model are derived and simulation results are presented for materials providing different absorptions. Application is focused on the relative effects of absorption versus the mounts' resilience on the transmission loss of the system.

4:15

4pSA7. Experimental and numerical comparison of acoustic performance of sound packages with viscoelastic damping or equivalent mass as treatments to flat aluminum panel. Esen Cintosun, Tatjana Stecenko (MTI Polyfab Inc., 7381 Pacific Circle, Mississauga, ON L5T 2A4, Canada), and Noureddine Atalla (Univ. de Sherbrooke, Sherbrooke, PQ, J1K2R1, Canada)

Acoustic performance parameters of airborne and structure-borne insertion loss were compared experimentally and numerically for sound packages with viscoelastic damping or equivalent mass. Transfer matrix method and finite element method were used to perform the numerical analysis. Viscoelastic material damping and equivalent mass (as part of sound packages) were compared as treatments to an aluminum flat panel. The sound packages were made up of either fiberglass or foam in addition to viscoelastic material damping or equivalent mass. The viscoelastic material damping used in this study is constraining layer damping (CLD). The equivalent mass was a solid material with the same surface weight of CLD. As part of the analysis, aluminum panel with and without sound package was subjected to diffuse acoustic field and point force mechanical excitations. The experimental and numerical results both show the same trends. [Work supported by MTI Polyfab Inc.]

Contributed Paper

4:40

4pSA8. Acoustic wave transmission loss mechanisms in poroelastic media, modeling, and experiment. Hasson M. Tavossi (Dept. of Phys., Astronomy & Geosciences, Valdosta State Univ., 1500 N. Patterson St., Valdosta, GA 31698)

Among factors that determine acoustic wave transmission loss in poroelastic media are incident wave energy loss due to acoustic impedance mismatch, between pore fluid and solid frame, and their density ratio, in addition to frequency dependence of dissipation in viscous boundary layers of fluid-solid interface for compression and surface waves. Model for prediction of transmission loss mechanisms includes the use of dimensionless Prandtl and Strouhal numbers for momentum and heat transfers, respectively. Other mechanisms considered are transmission losses by resonant and wave localization in the pore spaces. Wave filtering by poroelastic media as well as transmission loss by mode conversion, scattering, diffraction, interference, absorption, and thermal conduction are considered. Mechanical and electrical models of low-pass, band-pass, and high-pass filters applied to poroelastic media lead to the type of parameters to be included in this model as a function of frequency and pore size. Experimental measurements to test the prediction on acoustic wave transmission loss in similar poroelastic media of constant uniform porosity, but of different thicknesses and pore sizes, are presented to test transmission loss as a function of pore size, media thickness, and acoustic wave frequency.

Session 4pSC

Speech Communication: Perception and Spoken Language Processing (Poster Session)

Kristin J. Van Engen, Chair
Northwestern Univ., Linguistics Dept., 2016 Sheridan Rd., Evanston, IL 60208

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:15 p.m. and contributors of event-numbered papers will be at their posters from 3:15 p.m. to 5:00 p.m.

4pSC1. Perceptual constancy, reverberation, and grouping: Within-band effects. Anthony Watkins, Simon Makin, and Andrew Raimond (School of Psych. and Clinical Lang. Sci., Univ. of Reading, Reading RG6 6AL, United Kingdom)

Perceptual constancy effects are observed when differing amounts of reverberation are applied to a context sentence and a test-word embedded in it. Adding reverberation to members of a "sir"-"stir" test-word continuum causes temporal-envelope distortion, which has the effect of eliciting more sir responses from listeners. If the same amount of reverberation is also applied to the context sentence, the number of sir responses decreases again, indicating an "extrinsic" compensation for the effects of reverberation. Such a mechanism would effect perceptual constancy of phonetic perception when temporal envelopes vary in reverberation. This experiment asks whether such effects precede or follow grouping. Eight auditory-filter shaped noise-bands were modulated with the temporal envelopes that arise when speech is played through these filters. The resulting "gestalt" percept is the appropriate speech rather than the sound of noise-bands, presumably due to across-channel "grouping." These sounds were played to listeners in "matched" conditions, where reverberation was present in the same bands in both context and test-word, and in "mismatched" conditions, where the bands in which reverberation was added differed between context and test-word. Constancy effects were obtained in matched conditions, but not in mismatched conditions, indicating that this type of constancy in hearing precedes across-channel grouping.

4pSC2. Speech intelligibility in speech noise: A perceptual training study. Kristin J. Van Engen (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, k-van@northwestern.edu)

The ability to understand speech in noisy conditions is not well predicted by the traditional audiogram or by an individual's ability to understand speech in quiet. Furthermore, listeners with different language backgrounds (e.g., native versus non-native listeners) are affected differentially by the presence of noise, and listeners of a given native language experience more interference from background noise in their native language versus in a foreign language. Given these observations, it is hypothesized that the capacity to tune into target speech and tune out interfering speech noise is subject to experience-related modification and as such should be trainable. The objective of this study is to identify training parameters than can best enhance speech-in-speech intelligibility. To do this, we employ a multiple-talker training paradigm and three between-subjects training conditions: training in speech-shaped noise (a high energetic masker), in English two-talker babble (a high informational masker), and in Mandarin two-talker babble (a lower-energetic/lower-informational masker). By administering a common speech-in-speech post-test, the comparison of these training conditions allows us to begin to isolate the most trainable processes involved in speech-in-speech intelligibility. This approach also allows us to investigate whether training can mitigate the previously observed detrimental effects of native-language noise versus foreign-language noise.

4pSC3. The effect of occluded auditory feedback on articulatory learning. Silvia Lipski (Max-Planck Inst. for Neurological Res., Gleuelerstrasse 50, 50931 Cologne, Germany, lipski@nf.mpg.de), Stefanie Unger, Martine Grice (Univ. of Cologne, 50931 Cologne, Germany), and Ingo Meister (Univ. Hospital of Cologne, 50931 Cologne, Germany)

Auditory and somatosensory feedback are both claimed to be important for the acquisition and control of speech, especially during the learning of new speech sounds, where auditory feedback is said to provide a corrective signal for motor control. This study investigated adult speakers' speech motor learning capacity in the absence of auditory feedback on a contrast which has clear somatosensory feedback. Thirty German speakers were trained to produce a non-native singleton-geminate contrast on voiceless unaspirated bilabial plosives: /p/ versus /pp/, as present in Italian. For half of the speakers, auditory feedback was completely blocked during training. Results indicate that although there is no immediate effect when auditory feedback is masked, deviating productions emerge during the course of learning. By the end of training, speakers with masked feedback produce strong lengthening of segments and show more variation in their production than speakers with auditory feedback. These results provide evidence that auditory feedback is necessary for the learning of precise coordination of articulation, and that even aspects of speech which are salient in somatosensory terms are monitored auditorily during the learning of a new speech sound contrast.

4pSC4. The effect of word segmental structure on consonant length categorization by Finnish speakers. Kenji Yoshida, Kenneth de Jong (Dept. of Linguist., Indiana Univ. Bloomington, 1021 E. 3rd St., Bloomington, IN 47408, keyoshid@umail.iu.edu), and Pia-Maria Paivio (Indiana Univ. Bloomington, Bloomington, IN 47405)

While both Finnish and Japanese have a contrast in segmental length, previous research [Aoyama (2000)] has found that Finnish listeners exhibit a sharper identification for consonant length contrasts. The present study examines the effect of segmental structure on length categorization in the two languages. Twenty-two Finnish speakers were presented acoustic stimuli, where silence intervals with the duration of seven equal steps were embedded in three pairs of nonsense words spoken by a Finnish and a Japanese speaker and gave forced choice responses between single and geminate. The result shows that the slope of the identification function is steeper for the stimuli produced by the Finnish talker. Moreover, the study finds earlier identification with geminate for the words with CVC initial syllable as compared to CV. This effect of segmental structure, however, is restricted when the target consonant is located in the third syllable and for the stimuli produced by the Finnish speaker. These observations suggest that the listeners refer to their knowledge that the vowel in the second syllable is longer when

preceded by CV syllable, the phonetic fact known as "half-long vowel" in Finnish [Suomi *et al.* (2003)]. [Work supported by Finlandia Foundation National Scholarship.]

4pSC5. Listener's variation in phoneme category boundary as a source of sound change: A case of /u/-fronting. Reiko Kataoka (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720-2650)

Previous work examined listeners' recognition of a vowel in a series of [dVt] and [bVp] syllables varying perceptually from /CiC/ to /CuC/ in four different conditions (without precursor phrase, with precursor phrase, and in fast, medium, and slow speech), and found that ambiguous vowel stimuli were more often heard as /u/ in the [d_t] context as opposed to the [b_p] context (compensation for coarticulation). This paper reports the results of further analysis of the data and shows that although listeners varied in their /i/-/u/ category boundaries, their perceptual responses were systematic so that a group of listeners who had the category boundary closer to the /i/-end than the rest of the listeners in the "no-precursor" condition consistently had it this way in other conditions as well. This study also investigated the listeners' response in vowel repetition task, where the listeners were asked to listen to the same [CVC] stimuli as used in the vowel recognition task and to repeat only the vowel. Results of this part of the study will be presented and the implications for a listener-based theory of sound change will be discussed.

4pSC6. Perceptual adaptation to a spoken passage's long term spectral average. Jingyuan Huang (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, jingyuan@andrew.cmu.edu), Lori Holt (Carnegie Mellon Univ., Pittsburgh, PA 15213), and Andrew Lotto (Univ. of Arizona, Tucson, AZ 85721-0071)

The current study demonstrates that listeners adapt to a passage of speech such that subsequent speech categorizations are made relative to the passage's long-term average spectrum (LTAS). Native-English participants listened to a passage from Harry Potter for about 2 min. Next, they completed a categorization task across a series of natural speech tokens from the same talker, manipulated to vary perceptually from /ga/ to /da/. The passage was filtered to emphasize or de-emphasize regions of the LTAS without altering perceived talker identity or intelligibility. Following exposure to a passage with greater high-frequency energy, listeners more often categorized targets as /ga/ compared to target categorization following the same passage with lower high-frequency energy. Thus, listeners exhibit sensitivity to long-term spectral distributions and categorize subsequent speech relative to the LTAS of the exposure context. The spectrally contrastive directionality of the effect is consistent with earlier work demonstrating the influence of adjacent context on speech categorization, but this study extends the findings to the LTAS of a passage (across minutes) and demonstrates that context need not be adjacent to influence speech categorization. The implications for this work for talker and accent normalization will be discussed. [Work supported by NIH R01DC004674].

4pSC7. Autistic traits predict individual differences in speech categorization. Dan Hufnagle, Lori L. Holt, and Erik D. Thiessen (Dept. of Psych., Carnegie Mellon Univ., Pittsburgh, PA 15213)

Investigating individual differences in speech perception using measures of "autistic" traits in neurotypicals can gauge natural variability in speech processing [M. Stewart and M. Ota, Cognition **109**, 157–162 (2008)]. Using the autism-spectrum quotient (AQ) [Baron-Cohen *et al.*, J. Autism & Dev. Disord. **31**, 5–25 (2001)], which measures autistic traits in neurotypicals, we investigated individual differences in context-dependent speech processing. Twenty-eight neurotypicals categorized a nine-step da/ga series in the context of non-speech tone precursors [following L. Holt, Psychol. Sci. **16**, 305–312 (2005)] and completed the AQ. Context included three tone groups, including relatively high (shift toward ga), medium, and low (shift toward da) tones. Overall, the temporally adjacent tone grouping shifted perception more than distant context ($p<0.001$). Effects correlated with AQ ($r=0.53$). Lower AQ (fewer autistic traits) is associated with near-zero context dependence for endpoint categorization and large context-dependence for ambiguous speech-target categorization. Higher AQ is associated with intermediate influence of context across the series. Individual differences in context-dependent phonetic processing can be predicted from a personality trait scale, suggesting that phonetic processing is not immune from the influence of higher-order cognitive processes associated with these traits or that lower-level perceptual processing varies with these traits. [Work supported by NIH.]

4pSC8. Predicting perceptual outcomes from acoustic measures of vowels in dysarthria: A classification analysis. Kaitlin L. Lansford and Julie M. Liss (Dept. of Speech and Hearing Sci., Arizona State Univ., P.O. Box 870102, Tempe, AZ 85287, kaitlin.lansford@asu.edu)

Reductions in vowel space presumably obscure the distinctiveness of vowels produced by individuals with dysarthria. This represents a source of intelligibility decrement to the extent that more ambiguous vowels influence access of correct lexical items. The present study sought to examine (1) whether vowel formant frequencies predict perceptual decisions through classification analysis and (2) whether different forms of dysarthria are associated with different classification outcomes. Productions of phrases containing the target vowels /i/, /ɪ/, /ɛ/, /æ/, /u/, /o/, /ɑ/, and /ʌ/ in strong syllables were obtained from patients whose speech was affected by one of four neurological impairments: Parkinson disease, Huntington disease, amyotrophic lateral sclerosis, or cerebellar degeneration. The first two formant frequencies of each vowel were measured at its midpoint and then subjected to a classification analysis. Classification rules based on the acoustic measurements were used to classify each token as one of the eight target vowels. For each speaker group, the performance patterns obtained by the classification analysis were compared to those made by 15 listeners in an open transcription task. The ability of the classification analysis to predict perceptual outcomes specific to each dysarthric speaker group will be discussed. [Work supported by NIH/NIDCD.]

4pSC9. Children need coherence masking protection. Eric W. Tarr and Susan Nittrouer (Dept. of Otolaryngol.-Head & Neck Surgery, The Ohio State Univ., 915 Olentangy River Rd., Columbus, OH 43212, tarr.18@osu.edu, nittrouer.1@osu.edu)

Listeners can recognize speech targets at poorer signal-to-noise levels when more signal components are present, even if the additional components are spectrally distant and contribute no new information, but only if all components can be grouped together. [P. C. Gordon, Percept. Psychophys. **59**, 232–242 (1997)]. This phenomenon (coherence masking release) was studied in adults and children using the first formant (F1) of [ɛ] and [ɪ] as targets. Formants were presented in low-pass white noise for vowel labeling, with and without a consistent F2/F3 component above the noise cutoff. Synthetic-speech and sine-wave stimuli were used. Thresholds for accurate labeling were obtained with an adaptive procedure for F1-only and for all 3-formants. For synthetic speech, adults had 3.3 dB of masking release for 3-formants compared to F1-only. Children showed a 6.5-dB effect in the same direction, due to higher thresholds for F1-only. All listeners heard sine waves as non-speech, showed similar thresholds, and had reduced masking release (1.5 dB), but in this case F1-only had the advantage over the 3-formant condition. Conclusions were as follows: (1) children need coherent signals to recognize speech in noise and (2) when combined with Gordon's results, it seems masking release for speech signals likely has a phonetic explanation.

4pSC10. Gaze distribution patterns for audiovisual speech stimuli in preschool children with and without hearing loss. Nicholas A. Smith, Mark VanDam, and Mary Pat Moeller (Boys Town Natl. Res. Hospital, 555 North 30th St., Omaha, NE 68131, smithn@boystown.org)

Visual information plays an important role in the perception of speech. Many studies have shown that the addition of visual information increases speech intelligibility, and that the visual channel alone (as in the case of speechreading) is capable of conveying meaning. The goal of this study was to examine whether children with hearing loss use visual information in speech perception differently from children with normal hearing. An eye-tracking system (faceLAB) was used to record the looking behavior of 3- to 5-year-old listeners while they watched video samples of speech. Children with hearing loss used a variety of assistive devices, including cochlear implants and hearing aids. A group of normal-hearing peers was also tested. Preliminary results reveal that children with hearing loss distribute their gaze fixations more narrowly around the talker's mouth region, whereas

children with normal hearing scan the talker's face more broadly. These differences may reflect hearing-impaired children's ability to compensate for degraded auditory signals through increased reliance on visual cues.

4pSC11. Identification of asynchronous monaural and dichotic vowel pairs across the adult lifespan. Daniel Fogerty, Diane Kewley-Port, and Larry Humes (Speech and Hearing Sci., Indiana Univ., 200 S. Jordan, Bloomington, IN 47405, dfogerty@indiana.edu)

Temporal onset asynchrony is one cue that listeners use to identify concurrent vowels. Young (N=80; 18–31 years), middle-age (N=40; 40–55 years), and older (N=150; 60–88 years) adults identified vowel pairs in a temporal-order paradigm under monaural and dichotic stimulus presentations. Experiments used forced-choice constant-stimuli methods to determine the smallest stimulus onset asynchrony (SOA) between brief 70-ms vowels that enabled identification of the stimulus sequence. Vowels modified from four words (pit, pet, pot, and put) served as stimuli. All listeners identified the vowels in isolation with better than 90% accuracy. Results indicated that older listeners performed significantly poorer on monaural and dichotic temporal-order identification tasks than young listeners, with middle-age listeners in between. Correlations of performance with age across the full age span were moderate. For all three groups, SOAs for the dichotic task were significantly longer than those for the monaural task. A significant main effect of vowel pair was observed, indicating that not all vowel pairs were equally identifiable. Patterns of vowel pair identification were similar across all groups for both monaural and dichotic presentations; however, interactions with age group were observed. Effects of vowel order and dominance were also observed. [Work supported, in part, by NIA R01 AG022334.]

4pSC12. Effect of frequency selectivity on the perception of spectrally/temporally interrupted speech. Michelle Hsieh and Su-Hyun Jin (Dept. of Commun. Sci. and Disord., University of Texas, 1 University Station A1100, Austin, TX 78712, michellehsieh@mail.utexas.edu)

Compared to normal hearing (NH), hearing impaired (HI) listeners have shown significantly less masking release, even when the speech and noise were amplified to compensate for their hearing loss [J. Nelson (2004)]. According to Jin (2003), the reduced masking release of HI listeners was highly related to auditory filter bandwidths, which is an indicator of frequency selectivity. As a follow-up, the present study examined whether the reduced frequency selectivity might influence on spectrally-based perceptual strategies for speech recognition of NH and HI listeners in complex noise. We hypothesized that even with mild hearing loss, HI listeners would show significantly different spectral weight which would be significantly correlated with individual's hearing loss. In this study, signal processing was carried out in the digital domain and involved splitting of the signal spectra into five frequency bands for both speech and noise (band 1: 100–250 Hz, band 2: 250–750 Hz, band 3: 750–1750 Hz, band 4: 1750–3750 Hz, and band 5: 3750–7750 Hz). The subject's task was to identify the sentences heard in quiet, steady, and fluctuating noise in each band. The hypothesis will be examined and its implication on speech perception of HI listeners and signal processing strategy for the amplification system.

4pSC13. Looking for phoneme-level inhibition in spoken word recognition using auditory lexical decision. James White (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543, jameswhite@ucla.edu)

The TRACE model of spoken word recognition [McClelland & Elman (1986)] contains phoneme-level inhibition while the MERGE model [Norris *et al.* (2000)] does not include active competition between phonemes. Previous work found evidence for facilitation at the phoneme level and competition at the lexical level, but little research has been conducted looking for the existence of inhibition between phonemes. Using an auditory lexical decision task, the current study looks for phoneme-level inhibition by giving participants English target words preceded by isolated English sounds as primes. There are three conditions depending on how related the prime sound is to the final phoneme of the target word: identical (e.g., [s] before "gas"), similar (e.g., [s] before "cash"), and unrelated (e.g., [m] before gas). Facilitation, or faster average reaction times (RTs), is expected in the identical condition relative to the baseline unrelated condition. If facilitation is also found in the similar condition, the results will provide evidence against phoneme-level inhibition, supporting an approach with gradient activation of phonemes. However, if the average RT for the similar condition is greater than or equal to that of the unrelated condition, the results will support models of spoken word recognition containing phoneme-level inhibition.

4pSC14. Competition among variant word forms in spoken word recognition. Micah Geer (Dept. of Psych., Univ. at Buffalo, State Univ. of New York, Park Hall, Buffalo, NY 14260, mgeer@buffalo.edu)

Words in casual speech exhibit considerable variation in articulation. For example, alveolar stop consonants (/t/ and /d/) in certain phonetic environments may be realized as taps, glottal stops, careful /t/s and /d/s, or they may be deleted altogether. Thus, words containing non-word-initial alveolar stops may be represented in memory as multiple specific variants. Whether these multiple representations of variant forms compete for recognition, and at what level of representation such competition might occur, was investigated. Processing time was measured for monosyllabic words ending in either alveolar or non-alveolar (bilabial or velar) stops. Alveolar-ending words were responded to more slowly than carefully matched non-alveolar ending words in both lexical decision and same-different matching tasks. This result did not hold for similarly composed nonwords. In a follow-up experiment, the proportion of alveolar-ending neighbors in a word's phonological neighborhood was manipulated. Overall, the results suggest that variant word forms compete at a stage beyond sublexical processing. Implications for characterizing competition in spoken word recognition are discussed.

4pSC15. Phonological reduction in spoken word recognition. Malte C. Viebahn and Paul A. Luce (Dept. of Psych., SUNY at Buffalo, Park Hall, Buffalo, NY 14260-4110, mviebahn@buffalo.edu)

In casual speech, speakers often produce reduced phonological variants of the intended spoken words. Although these reductions are not obligatory, they occur in a regular way and only in certain phonological environments. Nasal tapping is one of these allophonic processes. During nasal tapping words such as "center" are pronounced as "cenner." Recent studies suggest that the recognition of word forms that have undergone nasal tapping is associated with substantial processing costs. The present set of experiments explores potential causes that may contribute to this effect. It is suggested that the perception of allophonic variants depends on the phonetic conditions under which they are produced and encountered.

4pSC16. At which processing level does extrinsic speaker information influence vowel perception? Matthias J. Sjerps, Holger Mitterer (Max Planck Inst. for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, The Netherlands, matthias.sjerps@mpi.nl), and James M. McQueen (Radboud Univ. Nijmegen, Nijmegen, The Netherlands)

The interpretation of vowel sounds depends on perceived characteristics of the speaker [e.g., average first formant (F1) frequency]. A vowel between /ɪ/ and /ɛ/ is more likely to be perceived as /ɪ/ if a precursor sentence indicates that the speaker has a relatively high average F1. Behavioral and electrophysiological experiments investigating the locus of this extrinsic vowel normalization are reported. The normalization effect with a categorization task was first replicated. More vowels on an /ɪ/-/ɛ/ continuum followed by a /papu/ context were categorized as /ɪ/ with a high-F1 context than with a low-F1 context. Two experiments then examined this context effect in a 4I-oddity discrimination task. Ambiguous vowels were more difficult to distinguish from the /ɪ/-endpoint if the context /papu/ had a high F1 than if it had a low F1 (and vice versa for discrimination of ambiguous vowels from the /ɛ/-endpoint). Furthermore, between-category discriminations were no easier than within-category discriminations. Together, these results suggest that the normalization mechanism operates largely at an auditory processing level. The mismatch negativity (an automatically evoked brain potential) arising from the same stimuli is being measured, to investigate whether extrinsic normalization takes place in the absence of an explicit decision task.

4pSC17. Sensitivity to input distributions and decision boundaries in auditory category learning. Sung-joo Lim and Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, sungjol@andrew.cmu.edu)

Previous research demonstrates the sensitivity of adults and infants to the statistical regularity of input distributions defining speech categories [D. L. Grieser and P. K. Kuhl, Dev. Psychol. **25**, 577–588 (1989)] and even non-

human animals exhibit such sensitivity [Kluender *et al.*, J. Acoust. Soc. Am. **104**, 3568–3580 (1998)]. Speech categories' structure also possesses information to support the use of decision boundaries in categorization. To investigate the interaction of distribution versus decision-boundary information in auditory category learning, the current research tracked listeners learning novel non-speech categories defined by two acoustic dimensions, the center frequency and modulation frequency, via explicit training with feedback. Early in learning, listeners exhibited sensitivity to distributional regularities of categories by robustly responding to more densely sampled regions of acoustic space. However, evidence of listeners' reliance on a decision boundary emerged once learning plateaued. The point in learning at which decision boundaries predicted categorization response was determined not only by the type of listeners' prior experience with the sounds but also by the perceptual salience of the acoustic dimensions. [Work supported by NSF.]

4pSC18. Integration of phonological information in obstruent consonant identification. Noah H. Silbert, Kenneth J. de Jong (Dept. of Linguist., Indiana Univ., Bloomington, IN 47405 and Dept. of Cognit. Sci., Indiana Univ., Bloomington, IN 47405), Jennifer J. Lentz, and James T. Townsend (Indiana Univ., Bloomington, IN 47405)

Speech perception requires the integration of information from multiple phonetic and phonological dimensions. Numerous studies have investigated the mapping between multiple acoustic-phonetic dimensions and single phonological dimensions (e.g., spectral and temporal properties of stop consonants in voicing contrasts). Many fewer studies have addressed relationships between phonological dimensions. Most such studies have focused on the perception of sequences of phones (e.g., bid, bed, bit, and bet), though some have focused on multiple phonological dimensions within phones (e.g., voicing and place of articulation in [p], [b], [t], and [d]). However, strong assumptions about relevant acoustic-phonetic dimensions and/or the nature of perceptual and decisional information integration limit previous findings in important ways. New methodological developments in the general recognition theory framework enable a number of these assumptions to be tested and provide a more complete model of distinct perceptual and decisional processes in speech sound identification. A non-parametric Bayesian analysis of syllable-onset consonant identification data from two experiments indicate that integration of phonological information is partially independent of both perception and decision making for most subjects, and that patterns of independence and interaction vary with the set of phonological dimensions under consideration (e.g., voicing and place of articulation versus voicing and manner of articulation).

4pSC19. Investigation of the neural bases of context-dependent speech categorization. Erika J. C. Laing, Lori L. Holt (Dept. of Psych., Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213), and Anto Bagic (Univ. of Pittsburgh Medical Ctr., Pittsburgh, PA 15213)

Previous research has demonstrated that simple sequences of preceding sine-wave tones affect speech categorization in a spectrally-contrastive manner [L. Holt, Psychol. Sci. **16**, 305–312 (2005)]. The current research explicitly links these effects to effects commonly thought to be instances of talker normalization [P. Ladefoged and D. E. Broadbent, J. Acoust. Soc. Am. **29**, 98–103 (1957)]. Synthesized sentences manipulated to sound like different talkers influence categorization of a subsequent speech target only when sentences' long-term average spectra (LTAS) predict spectral contrast. Likewise, sequences of tones modeling these LTAS differences produce parallel context-dependent speech categorization effects. The predictiveness of LTAS, rather than perceived talker, suggests that general auditory rather than speech-specific or articulatorily-driven mechanisms may play a role in effects considered to be instances of talker normalization. The behavioral measures are paired with magnetoencephalography (MEG) to investigate the neural bases of these parallel effects. Listeners categorized a /ga/-/da/ series in the context of preceding sentences and nonspeech tone sequences varying in their LTAS while MEG signals were acquired. Analyses focus on how the speech target is encoded as a function of preceding LTAS and the status of the context as speech or nonspeech. [Work supported by NIH and NOHR.]

4pSC20. Interaction of top-down and bottom-up processing on categorical perception. Ji Young Lee and Mark Hedrick (Dept. of Audiol. and Speech Pathol., Univ. of Tennessee, Knoxville, 578 South Stadium Hall, Knoxville, TN 37996-0740, jlee71@utk.edu)

To investigate the processing of acoustic and semantic information on phonemic categorical perception, two nine-step continua consisting of natural and synthetic consonant-vowel (CV) syllables were presented in four different contexts. For both natural and synthetic speech, /ti/-/pi/ continua varying in spectral tilt were constructed. The four contexts included the isolated CV syllable, a neutral context sentence with the CV at the end, and two additional sentences with varying semantic load. One semantically favored a /p/ response and one a /t/ response. Eleven adults with normal hearing (2 males, 9 females, 21 45 years old, mean age=27) labeled the consonant in the CV syllable. Results showed a pattern of statistically-significant differences between the isolated and neutral context versus the semantically-loaded sentences for both natural and synthetic stimuli, suggesting an effect of semantic context. The differences appeared in the vicinity of the boundaries of the psychometric functions, suggesting an interaction between top-down and bottom-up processing. Strictly bottom-up processing would have shown overlaid response functions, and strictly top-down processing would have shown differences at the endpoints. The pattern of results suggests interactive, parallel rather than hierarchical processing.

4pSC21. The influence of talker gender, formant-frequency scaling, and presumed sources of variance on the perception of voiceless fricatives in American English. Benjamin Munson and Alexander Coyne (Dept. of Speech-Lang.-Hearing Sci., Univ. of Minnesota, 115 Shevlin Hall, 164 Pillsbury Dr. SE, Minneapolis, MN 55455, munso005@umn.edu)

We examined listeners' identification of *sigh-shy* and *sigh-thigh* stimuli created by appending /s/-/ʃ/ and /s/-/θ/ continua to tokens of /ɑɪ/ excised from the natural productions of *sigh* spoken by one man and one woman. The values of the formants for the vocalic bases were scaled to give the illusion of differences in talkers' overall stature. Five levels of scaling were created for both the man and woman's productions. For the /s/-/ʃ/ continuum, listeners perceived more /ʃ/ with women's voices than with men's voices but, contrary to earlier research, showed gradient effects of formant scaling on phoneme boundaries for both the male and female tokens. For the /s/-/θ/ continuum, more /s/ tokens were identified for women's voices than for men's voices. Gradient effects of formant-frequency scaling were found for both men's and women's voices, and these were much smaller than those for /s/-/ʃ/. Listeners who were led to believe that the scaling reflected differences in talkers' age height showed a bigger influence of scaling on /s/-/θ/ perception than did listeners who were told they reflected differences in height. Together, these results show that perceived gender, vocal-tract normalization, and social stereotypes influence fricative identification.

4pSC22. Potential information predicts sentence intelligibility better than vowels, consonants, or duration. Christian E. Stilp and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin, 1202 W. Johnson St., Madison, WI 53706, cestilp@wisc.edu)

Speech sounds are typically divided into consonants and vowels. When either vowels or consonants, as delineated in TIMIT sentences, are replaced by noise, listeners are more accurate understanding sentences in which consonants are replaced relative to sentences in which vowels are replaced. From such data, vowels have been suggested to be more important for understanding sentences; however, interpretation of these data is unclear because consonant segments are roughly one-third shorter than vowels as marked within TIMIT. We report two experiments that demonstrate listener performance is better predicted by simple acoustic measures of spectral change, not duration or consonant/vowel distinction. First, listeners identified sentences in which consonants, vowels, C-V transitions, or V-C transitions were replaced by speech-shaped noise. Contrary to previous work, performance suffered most when consonants were replaced by noise. In a second experiment, portions of sentences were replaced based on cochlea-scaled spectral entropy (change across time). Speech segments having relatively high, medium, or low change over 80-ms (consonant duration) or 112-ms (vowel duration) intervals were replaced with noise. Intelligibility decreased linearly as amount of replaced spectral entropy increased independent of duration. Spectral change (entropy) is more important than consonant/vowel distinctions and duration. [Work supported by NIDCD.]

4pSC23. The perceptual time course of coarticulatory nasalization. Patrice S. Beddor, Julie E. Boland, Andries Coetzee, and Kevin McGowan (Dept. of Linguist., Univ. of Michigan, 611 Tappan St., Ann Arbor, MI 48109, beddor@umich.edu)

Listeners' moment-by-moment processing of anticipatory vowel nasalization and a following nasal consonant was investigated. English-speaking participants' eye movements were monitored as they heard instructions to look at one of two pictured objects on a computer screen. Trials included pictured pairs for naturally produced words of the form CVNC-CVC (e.g., bend-bed), CVNC-CVNC (bend-bent), and CVC-CVC (bed-bet). Vowels in CVNC words were coarticulatorily nasalized. Results to date show that, when participants heard a CVNC word (bend), they visually fixated the correct picture earlier when the competing picture was CVC (bed)—that is, when the vowel in the competitor would be expected to be non-nasal—than when the competitor was another CVNC word (bent). Results also suggest that participants often fixated the target CVNC picture in CVNC-CVC trials after onset of vowel nasalization but before N onset. However, although vowel nasalization facilitated early selection of CVNC over CVC, a non-nasalized vowel was not similarly helpful for selecting CVC over CVNC. When participants heard CVC (bed), they did not fixate the correct picture earlier when the competing picture was CVNC (bend) than when the competitor was CVC (bet). Findings are interpreted in light of production data for English and perceptual theories.

4pSC24. Effects of native language on perception of the relative amplitudes of the first and second harmonics. Jody Kreiman, Bruce R. Gerratt (Head and Neck Surgery, UCLA School of Medicine, 31-24 Rehab Ctr., Los Angeles, CA 90095-1794, jkreiman@ucla.edu), and Sameer ud Dowla Khan (Cornell Univ., Ithaca, NY 14850)

As an incidental result of previous studies of the perception of spectral slope, we found that Mandarin speakers were nearly twice as sensitive to changes in the amplitude of the first harmonic relative to the second (H1-H2) as are English speakers. Although two Mandarin tones are typically produced with characteristic changes in voice quality, Mandarin does not contrast phonation types directly. Thus, the observed difference between listener groups could be due to increased attention to overall source characteristics in speakers of tone languages (because F0=H1), or specifically to experience with voice quality differences apart from F0. To clarify this matter, we measured just-noticeable differences in H1-H2 for speakers of Gujarati, a non-tonal language that contrasts breathy and modal vowels in H1-H2, and for speakers of a language like Yoruba that contrasts tones but not phonation types. If increased sensitivity to H1-H2 differences is due to attention to H1, then just-noticeable differences for both groups should be smaller than those for English speakers. However, if perceptual experience specifically with the H1-H2 contrast is required, then Gujarati speakers, but not Yoruba speakers, should show increased sensitivity relative to English speakers. [Research supported by NIH and NSF.]

4pSC25. Examining the role of pitch, duration, and intensity in word segmentation. Chad Vicenik (Dept. of Linguist., UCLA, Los Angeles, CA 90095 cvicenik@humnet.ucla.edu)

English listeners can make use of stress cues in word segmentation. Stress is associated with a number of phonetic cues, including pitch movements, longer durations, and greater intensity. Studies on stress perception have shown that pitch is the most powerful cue to English stress, followed by increased duration and greater intensity [Fry (1958); Bolinger (1958)], suggesting that pitch alone might be sufficient to cue word boundary. Here, I test whether pitch alone is enough to cue word boundary for English listeners, using an artificial language paradigm. The artificial language used contains no distributional cues to word boundary, so words can only be segmented using pitch. I also pit pitch cues against intensity and durational cues in order to test whether English listeners weight cues for stress in a word segmentation task, or if they require correlates of stress to be bundled together.

4pSC26. Lexicalizing nonadjacent dependencies. Steven A. Berg (Dept. of Psych., SUNY at Buffalo, 206 Park Hall, Amherst, NY 14226, sberg @buffalo.edu)

Previous research [e.g., McLennan *et al.* (submitted)] has demonstrated that listeners are sensitive to discontinuous dependencies among consonants, vowels, and syllables in nonsense words. An extension of this work was done by attempting to lexicalize nonwords containing discontinuous dependencies through a referent matching paradigm. Our attempt at lexicalization was to determine the consequences of discontinuous dependencies at the lexical rather than sublexical levels. Our findings are discussed in terms of sublexical and lexical effects of phonotactics in spoken word recognition.

THURSDAY AFTERNOON, 29 OCTOBER 2009 RIO GRANDE CENTER, 1:30 TO 5:15 P.M.

Session 4pUW

Underwater Acoustics and Acoustical Oceanography: Progress in Modeling Complex and Dynamic Acoustic and Oceanographic Characteristics of Continental Shelves and Slopes II

Preston S. Wilson, Cochair
Univ. of Texas at Austin, Dept. of Mechanical Engineering, 1 University Station, Austin, TX 78712

David P. Knobles, Cochair
Univ. of Texas at Austin, Applied Research Labs., 10000 Burnet Rd., Austin, TX 78758

Contributed Papers

1:30

4pUW1. Maximum entropy inference method applied to shallow-water ocean waveguides. Jason D. Sagers and David P. Knobles (Appl. Res. Labs., Univ. of Texas at Austin, 10000 Burnet Rd., Austin, TX 78758)

The principle of maximum entropy (MaxEnt) is used to analyze the uncertainty that is associated with determining geo-acoustic waveguide parameters for shallow-water ocean environments. Marginal probability distributions and statistics, such as the mean and standard deviation, for each geo-acoustic parameter are obtained using MaxEnt methods. Both theoretical and numerical components of the MaxEnt method are addressed. The method is applied to experimental acoustic measurements taken on the New Jersey continental shelf on a sand ridge where *a priori* information includes sound velocity measurements in the water column and chirp seismic reflection grams that provide geophysical ground truth on the seabed layering structure. The acoustic data used in the geo-acoustic inference result from a towed multi-frequency source. The effect of discretization choices for the geo-acoustic representation on the marginal probability distributions is investigated. The effect of changing *a priori* information (such as the assumed sound velocity profile in the water column) on the marginal probability distributions of the geo-acoustic parameters is also explored.

1:45

4pUW2. Modeling intensity fluctuations of acoustic transmissions from the Research Vessel Sharp during Shallow Water 2006. Georges A. Dossot, James H. Miller, Gopu R. Potty (Dept. of Ocean Eng., Univ. of Rhode Island, Narragansett Bay Campus, Narragansett, RI 02882), Mohsen Badiey (Univ. of Delaware, Newark, DE 19716), James F. Lynch, Ying-Tsong Lin, Arthur E. Newhall (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and Kevin B. Smith (Naval Postgradraduate School, Monterey, CA 93943)

During the Shallow Water 2006 (SW06) experiment, a J-15 source deployed from the University of Delaware's R/V Sharp transmitted various acoustic signals at several different bearings and ranges to the Woods Hole Oceanographic Institution's vertical hydrophone line array. The array was intentionally positioned near the shelf-break front and in an area where internal waves are known to occur. During several of the R/V Sharp's acoustic transmissions, internal waves passed through the sound field such that the internal wave front was near parallel to the acoustic transmission path. Measured data show substantial intensity fluctuations that vary over time and space due to complex multimode and multipath (both two and three dimensional) interference patterns. This presentation provides an overview of the R/V Sharp's transmissions and the corresponding intensity fluctuations of received signals at the array. In order to model the intensity fluctuations, a sound speed field is reconstructed using data from environmental sensors such as acoustic Doppler current profilers, ship radar, conductivity-temperature-depth profiles, environmental moorings, and satellite images. Measured data are compared to a simulated sound field computed by a three dimensional acoustic propagation model. [Work sponsored by the Office of Naval Research.]

2:00

4pUW3. Observation and modeling of sound propagation through a dynamic ocean on the continental shelf of the northeastern South China Sea. Ching-Sang Chiu, Christopher W. Miller (Dept. of Oceanogr., Naval Postgrad. School, Monterey, CA 93943), Chi-Fang Chen, Linus Chiu (Natl. Taiwan Univ., Taipei, Taiwan), and Ruey-Chang Wei (Natl. Sun Yat-sen Univ., Kaohsiung, Taiwan)

In April 2005, a 3-day acoustic transmission experiment was carried out on the shelf of the northeastern South China Sea. A 400-Hz signal was transmitted every 5 min from a moored source to a moored vertical hydrophone array 17 km away. Additionally, a series of oceanographic moorings was deployed along the transmission path to sample the water-column variability that was dominated by the evolution of nonlinear internal tides and high-frequency nonlinear internal waves. Applying time-series filtering, principal component analysis, and a feature tracking technique to the oceanographic data, a continuous space-time empirical model for the sound-speed field was developed. Using a coupled-mode sound propagation model, interfacing with the sound-speed model, the temporal variations of the vertical distribution of signal intensity at the hydrophone array were computed. Analyzing the model results and comparing them to the measured signal intensities have allowed for quantification and comparison of the effects of the nonlinear internal tides, depression waves, and elevation waves on the sound transmission. The modeled sound-speed and sound-intensity fields and their comparisons to the measured data are discussed in this paper. [This research is jointly sponsored by the U.S. Office of Naval Research and the Taiwan National Research Council.]

2:15

4pUW4. Observed anisotropy in low frequency acoustic signal loss at the sea floor. Warren Wood, Dennis Lindwall (Naval Res. Lab., 1005 Balch Blvd., Stennis Space Ctr., MS 39529), and Nathan Bangs (The Universtiy of Texas at Austin, Austin, TX, nathan@utig.ig.utexas.edu)

Anisotropy of 20%–30% in the low-frequency (10–120 Hz) acoustic bottom loss of marine sediments 1200-m deep off the coast of Oregon is reported here. Two nearly perpendicular acoustic profiles were acquired with air guns towed at the sea surface, crossing at a point directly over a four-component ocean bottom seismometer (OBS) on the seafloor. The area is a convergent margin, with a regional principal stress field perpendicular to broad bathymetric ridges and the shoreline, and numerous faults running parallel to the bathymetry. Amplitudes of sediment reflections are consistently 20%–30% less on the profile parallel to the bathymetry than on the profile perpendicular to the bathymetry. This is the opposite of what would be expected if the dominant attenuation mechanism were scattering from the fault planes. The anisotropy exists at angles for which acoustic energy penetrates sediments only a few hundred meters from the crossing point, where layering in both directions is laterally consistent. The direct arrival amplitudes, from the source to the OBS, are identical on both transects, suggesting that neither the acquisition technique nor the water column is responsible for the anisotropy. Possible explanations include seafloor bed forms and wavelength-scale roughness at the buried sediment interfaces.

2:30

4pUW5. Across-shelf versus along-shelf coherence length estimates of low-frequency pulse transmissions on the New Jersey continental shelf. Jon M. Collis (Colorado School of Mines, Golden, CO 80401), Timothy F. Duda, James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), and Harry A. DeFerrari (Univ. of Miami, Miami, FL 33149)

Sound at 100–400 Hz propagating in approximately 80-m depth water from fixed sources to a joint horizontal/vertical line array (HVLA) is analyzed. The data are from an area east of Delaware Bay in the Mid-Atlantic Bight populated with tidally-generated long- and short-wavelength internal waves that propagate generally shoreward. Two paths are used: a 29-km path in the cross-shore (across internal wave crest) direction and a 19-km path in the along-shore direction. Spatial fluctuations of HLA arrivals have been computed as a function of HLA beam steering angle and time: array gain, horizontally lagged field coherence function, and coherent beam power. These quantities and fitted characteristic coherence scale lengths vary in apparent response to internal-wave variability [Collis *et al.*, J. Acoust. Soc. Am. **124**, EL97 (2008)]. In this presentation, coherence length estimates are provided for the month of August 2006. In additional to internal volume effects, consideration is given to signal duration, angle of incidence, and along-shelf versus across-shelf propagation. [Work supported by the Office of Naval Research.]

2:45

4pUW6. Impact of shelfbreak fronts on long-range underwater sound propagation in the continental shelf area. Ying-Tsong Lin, Alexey Shmelev, James F. Lynch, Arthur E. Newhall (Appl. Ocean Phys. & Eng. Dept., Woods Hole Oceanograph. Inst., Woods Hole, MA 02543, ytlin@whoi.edu), and Pierre F. J. Lermusiaux (MIT, Cambridge, MA 02139)

The dynamic shelfbreak front commonly observed in the continental shelf area has significant impact on underwater sound propagation. Analytical study with an idealized front model has shown that combining inshore refraction from the front and offshore refraction on the sloping shelf, long-range propagating sound on the continental shelf will be trapped by the front and form "whispering gallery" modes on the inshore side of the front. In this paper, numerical propagation simulations with three-dimensional normal mode and parabolic approximation methods are implemented to investigate the acoustic impact of a nearly realistic front model. This front model results from the MIT Multidisciplinary Simulation, Estimation, and Assimilation System which assimilated the field measurements collected during the New Jersey Shelf Shallow Water 2006 experiment. Since the front model captures very well the temporal and spatial variability of the shelfbreak frontal system in the field, the numerical propagation simulations presented here are fairly realistic. Horizontal sound refraction due to the shelfbreak front and (sub)-mesoscale eddies are observed, and causes part of long-range propagating sound energy to be trapped on the shelf area.

3:00—3:15 Break

3:15

4pUW7. *In situ* acoustic measurements of transmission loss for improving the effectiveness of antisubmarine warfare sonar searches in tactical scenarios. Charles H. Wiseman (Peninsula Publishing, 26666 Birch Hill Way, Los Altos Hills, CA 94022)

The execution of sonar search plans against submarines by antisubmarine warfare ships is not as effective as it could be, especially in littoral waters. The reason is that the high spatial and temporal variability of the water environment cannot be accurately characterized as inputs into models leading to the sonar equation, used for the prediction of sonar detection ranges, the primary input into search plans. The effectiveness of search can be significantly improved, however, by adding the use of *in situ* acoustic measurements of transmission loss to the mix of measurements currently used by tactical forces. This paper foresees improvements in the accuracy in use of the sonar equation, despite the likely incompleteness of *in situ* acoustic measurements of transmission loss relative to the entire field that would be desired to be measured. Acoustic measurements of transmission loss will bring into consideration a degree of ground truth into search that has not been realized tactically heretofore. The more effective search will translate to a closer to optimum use of navy ships, aircraft, and submarines, and a saving of fuel and search time. Accordingly, this concept should be analyzed and modeled to determine if it is worth investigating more thoroughly.

3:30

4pUW8. Full-field acoustic transmission variabilities including horizontal refraction and back-scattering effects in the shelf and shelfbreak environments. Jinshan Xu and Pierre F. J. Lermusiaux (Dept. of Mech. Eng., MIT, 77 Massachusetts Ave., Cambridge, MA 02139, jinshan@mit.edu)

This study is motivated by the striking difference in acoustic transmission data collected on the shelf and shelfbreak northeast of Taiwan within the QPE 2008 Experiment. On the shelf, the acoustic transmission from a moving source along a circular track showed little variation with respect to bearing angles. However, on the shelfbreak, the data for the same type of circular tracks showed no transmission when the source was in the deeper waters. Our previous studies have investigated the uncertainties in transmission loss (TL) estimation due to the geo-acoustic model, bathymetry, and ocean environment by comparisons of $N\times$two-dimensional and forward scattering acoustic modeling to *in situ* acoustic data. Here, we present a novel coupled oceanographic(four-dimensional)-acoustic[three-dimensional (3-D)] field estimation with ocean data assimilation modeling. A 3-D wave propagation model based on a parabolic approximation is used to study horizontal refraction effects on the TL. A coupled-normal-mode wave propagation model is also applied to investigate back-scattering effects on the TL. Our model estimates including uncertainties are compared to observed data, and results are utilized to study and quantify the variability of the TL due to the regionally complex seabed topography and ocean water column (tides, fronts, and eddies). Our study illustrates when processes such as sound wave horizontal refraction and back-scattering should be accounted for in modern SONAR performance prediction systems.

3:45

4pUW9. Use of acoustic frequency shifts for monitoring oceanographic processes in shallow water. Altan Turgut (Acoust. Div., Naval Res. Lab., Washington, DC 20375)

Waveguide invariant theory is used to calculate the frequency shifts of the acoustic intensity spectral levels caused by environmental variability in shallow water. In winter conditions, long-term acoustic monitoring of tide and slope water intrusions has been demonstrated using measured frequency shifts of the acoustic intensity spectral levels [Turgut *et al.*, J. Acoust. Soc. Am. **121**, 2534–2541 (2007)]. Acoustic frequency shifts caused by other oceanographic processes such as background internal waves and solitary internal waves are also investigated. Additional broadband acoustic data from several summer experiments indicate the feasibility of extracting background internal wave parameters as well as internal solitary wave parameters when they propagate almost perpendicular to the acoustic propagation path. [This work was supported by the ONR.]

4:00

4pUW10. A model for four-dimensional coastal internal waves with applications to acoustics. James F. Lynch, Timothy F. Duda, Ying-Tsong Lin, Arthur E. Newhall (Woods Hole Oceanograph. Inst., 98 Water St., Woods Hole, MA 20543, jlynch@whoi.edu), and Pierre F. J. Lermusiaux (MIT, Cambridge, MA 02139-4307)

Acousticians need a fully four-dimensional (4-D) coastal oceanography model that can provide the soundspeed field down to the internal wave scale. While mesoscale ocean models exist, given the complexities of a full primitive equation (i.e., nonlinear), non-hydrostatic, fine-scale model including internal waves, it may be a decade or more before such a model is available from the mainstream oceanographic community. We pose the possibility of creating a usable approximate 4-D ocean model based on a combination of a powerful acoustics technique, the Weinberg–Burridge vertical modes and horizontal ray solution to the wave equation, and existing primitive-equation ocean models and two-dimensional nonlinear wave equations. In our technique, the local ocean internal wave modes are found, creating a horizontal index of refraction grid for internal-tide ray tracing from identified source points. Broadband internal waves are propagated along the trajectories using the KdV or other appropriate nonlinear wave equation, with internal-tide initialization and interpolation between trajectories. An ocean numerical model with good bathymetry provides the source distribution as well as buoyancy frequency profiles and current profiles, necessary inputs for the internal wave modes and propagation. The initial design for this model as well as examples of its usages for ocean acoustics are presented.

4:15

4pUW11. Computational grid design to improve three-dimensional parabolic equation modeling efficiency. Melanie E. Austin (JASCO Appl. Sci., 2101-4464 Markham St., Victoria, BC V8Z 7X8, Canada, melanie.austin@jasco.com) and N. Ross Chapman (Univ. of Victoria, Victoria, BC, V8W 3P6, Canada)

One class of very accurate computer models that can be used to predict underwater acoustic fields is based on three-dimensional (3-D) solutions of the parabolic form of the reduced acoustic wave equation. 3-D parabolic equation (PE) models contain differential operators in both depth and azimuth. These 3-D solutions are very computationally intensive, but strategic definition of the model grid can save computation time. A 3-D PE model (MONM3D) has been developed that incorporates techniques that reduce the required number of model grid points. The concept of tessellation is used to optimize the radial grid density as a function of range, reducing the required number of grid points in the horizontal planes of the grid. The model marches the solution out in range along several radial propagation paths emanating from a source position. Tessellation, as implemented in MONM3D, allows the number of radial paths in the model grid to depend on range from the source. In addition, the model incorporates a higher-order azimuthal operator which allows a greater radial separation and reduces the required number of radial propagation paths. The efficiency and accuracy of the MONM3D model will be presented by demonstration of a test case.

4:30

4pUW12. Low-frequency sound field focusing in shallow water. Valeriy G. Petnikov, Andrey A. Lunkov (A. M. Prokhorov General Phys. Inst., RAS, 38 Vavilov Str., Moscow 119991, Russia, petniko@kapella.gpi.ru), and Alexander A. Stromkov (RAS, Nizhny Novgorod 603950, Russia)

Within the framework of numerical simulation a feasibility of low-frequency (100–300 Hz) sound field focusing was studied in the waveguide common to shallow water. Two types of focusing are discussed: focusing of a wide-band sound signal by time reversal mirror (TRM) and focusing of a monochromatic sound field by a phase conjugation mirror (PCM). It is demonstrated that if we use TRM we can focus sound field by both point source-receiver (SR) element and a corresponding SR vertical array. At the same time, the focusing quality is somewhat improved with vertical array, and TRM provides focusing with approximately the same quality as PCM operating at the frequency equal to the carrier frequency of a wide-band signal. It is shown that in a waveguide with a constant sound velocity at the range of 10 km wind waves deteriorate focusing quality at a wind speed greater than 12 m/s. [This research was supported by the Russian Foundation for Basic Research Project Nos. 08-02-00283 and 07-02-001205 and by the Civilian Research and Development Foundation Project No. REC 010.]

4:45

4pUW13. Determining of directivity index and effective aperture at the idealized model of directivity pattern of antenna with elliptical cross section of main lobe. Zvonimir Milosic (MoD of the Republic of Croatia, Inst. for Res. and Development of Defense Systems, Trg kralja Petra Kresimira IV 1, 10000 Zagreb, Croatia)

In this paper a new universal procedure of calculating directivity index (directivity factor) and effective aperture dependent on the measured directivity pattern of antenna will be presented. The procedure derives from an idealized model of directivity pattern measured characteristics of a sonar antenna in accordance with the elliptical cross section perpendicular to acoustic axes of the main lattice of directivity pattern. My new, strong, and original condition for very low level of suppression of side lobes at the contemporary sonars (LoSoSMiZ condition <-50 dB), its meaning in directivity index and effective aperture on ranges at previous congresses, and its meaning in theory of reverberation are in accordance with this theory too. In accordance with the equations in this paper, the de-embedding of directivity indices and effective aperture of the receive antennas of sonar systems is simple, fast, reliable, accurate, and widely usable. Finally, the de-embedding of directivity index and effective aperture according to the conditions in this paper makes a better evaluation of given parameters possible: for buyers, in R&D, in academies, and on the operative levels.

5:00

4pUW14. The effect of time resolution (bandwidth) on clutter statistics: A vertical echo sounder as a simple test case. Chris Harrison (NATO Undersea Res. Ctr., Viale San Bartolomeo 400, 19126 La Spezia, Italy)

The scintillation index (SI) of sonar returns from a rough seabed was recently simulated using a Kirchhoff approach both numerically and semi-analytically [C. Harrison, J. Acoust. Soc. Am. **125**, 2661 (2009); IEEE JOE. In press]. Of particular interest was the potentially high SI associated with seabed reflection focusing. In carrying out this kind of numerical investigation, one needs to be very careful of what precisely constitutes the ensemble. Here we perform the same analysis but now with a finite pulse length. At sufficient altitude, the tail of the pulse may contain contributions from many different almost specular facets. Depending on the sonar's time resolution these may, or may not, be seen as many intersecting echoes. Even with a pulse length long enough to incorporate the vertical extent of the roughness, it is possible to find sharp spikes in the echo tails. Thus a plot of Si versus travel time (the ensemble being ping-to-ping variation) has a peak more than one pulse length after the main arrival. Amongst other things, this makes it clear that SI is highly dependent on time resolution and furthermore may be more a function of the observing system than the object being observed.

THURSDAY EVENING, 29 OCTOBER 2009 7:30 TO 9:30 P.M.

Open Meetings of Technical Committees

The Technical Committees of the Acoustical Society of America will hold open meetings on Tuesday, Wednesday, and Thursday evenings. On Tuesday and Thursday the meetings will be held starting immediately after the Social Hours at 7:30 p.m. On Wednesday, one technical committee will meet at 7:30 p.m.

These are working, collegial meetings. Much of the work of the Society is accomplished by actions that originate and are taken in these meetings including proposals for special sessions, workshops and technical initiatives. All meeting participants are cordially invited to attend these meetings and to participate actively in the discussions.

Committees meeting on Thursday are as follows:

Animal Bioacoustics	Rio Grande West
Biomedical Ultrasound/Bioresponse to Vibration	Rio Grande East
Noise	Regency East 1
Speech Communication	Regency East 3
Structural Acoustics and Vibration	Live Oak
Underwater Acoustics	Rio Grande Center

Session 5aNS

Noise and ASA Committee on Standards: Advancements in Soundscape Applications and Standards

Alexander U. Case, Cochair
Fermata Audio & Acoustics, P.O. Box 1161, Portsmouth, NH 03802-1161

Gary W. Siebein, Cochair
Univ. of Florida, School of Architecture, P.O. Box 115702, Gainesville, FL 32611

Chair's Introduction—8:30

Invited Papers

8:35

5aNS1. Soundscape design applications for transportation noise. Gary W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611), Robert M. Lilkendey, Hyun Paek, and Chris Jones (Siebein Assoc., Inc., Gainesville, FL 32607)

Three case studies are presented that demonstrate soundscape methods for assessing noise impacts and proposed mitigation strategies for train, traffic, and rail noise impacts. Comparisons among time average measures short term measurements of specific acoustic events are presented. One is a large housing project located adjacent to a major rail line. The soundscape methods allowed the designers to auralize the results of proposed noise mitigation strategies for clients to make informed decisions about the acoustical feasibility of developing the site. Second is a case study of aircraft noise impacts on a glass curtain wall luxury condominium project located in the flight path of a major urban airport. Soundscape methods were used to work with the developers to assess strategies for the reduction in aircraft noise into living units associated with the decision to enhance glazing systems for the project. Third is a case study of planning for noise mitigation of road traffic noise for future developments at the time of initial permitting for projects. A framework for long term planning strategies for controlling incremental noise impacts as they occur was developed for a growing medium size city adjoined by rural and agricultural land.

8:55

5aNS2. Soundscape realization through standard surround sound. Alex U. Case (Sound Recording Technol., Univ. of Massachusetts, 35 Wilder St., Lowell, MA 01854, alex_case@uml.edu)

The need for both qualitative and quantitative analyses may be the key challenge separating soundscape efforts from traditional environmental noise practice. On-site soundscape tours and interviews are a proven but resource-intense method for soundscape study. A means to obtain a soundscape and reproduce it elsewhere for subsequent analysis could be an efficient way to collect data and increase survey sample size. Reproduction off-site also makes soundscape modifications and comparisons possible, facilitating the measurement, evaluation, design, and composition processes. Standards specifying the full record and reproduce systems and procedures may never be able to anticipate the vast range of soundscape qualities that are salient across any and all soundscapes, loud to soft, low to high, large to small, transient to steady-state, urban to wilderness. The more pressing need is a review of the merits and shortcomings of the many soundscape recording and reproduction techniques most likely to be employed, and one small contribution is offered here. Cinemas and home theaters, with a somewhat-standardized 5.1-channel surround sound reproduction capability, might serve as a useful playback venue. They are well-understood by recordists and acousticians and have been used to fabricate a range of soundscapes in film and music, sometimes with great success. This paper evaluates the feasibility of using the 5.1 surround sound format for soundscape design.

9:15

5aNS3. A soundscape *pons asinorum*: Interim adaptation of A-weighted measurement practices to foster more accurate assessments of noise impacts in protected natural areas. Kurt M. Fristrup (Natl. Park Service, Natural Sounds Program, 1201 Oakridge Dr., Ste. 100, Fort Collins, CO 80525)

A-weighted measurements of environmental sound levels in functional ecosystems are often elevated by natural sources of sound energy above 1 kHz. It is aesthetically and biologically questionable to assert that a rich chorus of bird, frog, and insect sounds renders an environment less vulnerable to the effects of noise confined to lower frequencies. Accordingly, it seems worthwhile to explore alternative approaches to computing A-weighted sound levels that would reduce or eliminate this source of inflated measurements in park and wilderness settings. A systematic evaluation of acoustical measurements in units of the U.S. National Park Service reveals the extent to which A-weighted values are influenced by environmental sounds in each 1/3 octave band. This analysis can help motivate and inform the development of revised standards for A-weighted noise measurement and evaluation in protected natural areas. This seems a plausible first step toward a more biologically rigorous framework for assessing the impacts of noise to wildlife and visitors in National Parks.

9:35

5aNS4. Visitor dose-response data acquisition protocol development at four U.S. National Parks. Richard D. Horonjeff (81 Liberty Square Rd. #20-B, Boxborough, MA 01719, rhoronjeff@comcast.net) and Grant S. Anderson (76 Brook Trail, Concord, MA 01742)

In response to a mandate set forth by The National Parks Overflight Act of 1987 (PL 100-91), visitor response to tour aircraft overflights was sought by the National Park Service and the Federal Aviation Administration during the early 1990s. At issue was the extent to which varying doses of aircraft overflight noise resulted in both visitor-reported consequential degrees of annoyance and interference with an appreciation of natural quiet. This paper first presents a brief overview of the results of these studies (previously published in the literature), then focuses on the data collection protocol used. Several alternative dose and response data collection methodologies were considered and evaluated. This paper presents the perceived strengths and weakness of the various alternatives as well as the rationale behind the final adopted protocol. Deciding factors included cost, accepted social survey practices, ability to match noise doses with individual visitors, avoiding any pre-sensitization of visitors, and the ability to separate potentially confounding variables during data analysis. [Work supported by the National Park Service and the Federal Aviation Administration.]

9:55

5aNS5. Soundscape documentation of parks and natural areas. Gary W. Siebein (School of Architecture, Univ. of Florida, P.O. Box 115702, Gainesville, FL 32611) and Reece Skelton (Siebein Assoc., Inc., Gainesville, FL 32611)

A multi-media method to document the distinctions among qualities of sounds with relatively similar levels in a natural, outdoor environment was developed. Simultaneous recordings of sound levels, calibrated aural recordings of .wav files of specific acoustic events, and video photography of the events were overlaid on a map of the site to document the itinerary of the observer. The multi-media method attempted to more completely document the nature of the soundscape with its subtle changes in sources of sound, level, frequency, and time duration than typical environmental noise metrics such as Leq's, LDN's and other long term average sound metrics. The sound levels measured near the road are similar to those taken at locations within the park; however, the quality of sounds is very different from each other. The acoustical transition from a street edge into a dense, natural park was examined. The experiment shows how the nature of sound shifts dramatically through the site while the perceived average sound level remains relatively constant. The acoustical evaluation of the site includes both quantitative and qualitative elements so the true essence of a place can be considered and the subtle but important sonic qualities of sites can be designed.

10:15—10:30 Break

10:30

5aNS6. Soundscapes and city noise ordinances. Nancy S. Timmerman (Consultant in Acoust. and Noise Control, 25 Upton St., Boston, MA 02118-1609, nancy.timmerman@alum.mit.edu)

The word "soundscape" has been used to get away from the negative connotations of the word "noise." In reality, whether or not a sound or soundscape is annoying depends on the viewpoint of the listener. Two cities in Massachusetts, Boston and Somerville, will be examined to compare and contrast what soundscapes have been deemed "acceptable." In addition, the Massachusetts state code, covering areas which do not have such ordinances, will also be examined.

Contributed Papers

10:50

5aNS7. Comparison of sound pressure levels of mid-to large size on-road motorcycles through volume modeling. William T. C. Neale and Toby Terpstra (Kineticorp LLC, 6070 Greenwood Plaza Blvd. Ste. 200, Denver, CO 80111, wneale@kienticorp.com)

Noise from motorcycles varies between make and model, style, and engine size. The motorcycle's pipe design, materials, and gearing are just a few of the factors that determine the noise being emitted. In regular traffic situations, though, an observers relative position to the motorcycle varies, and this change in position greatly affects the experience of motorcycle noise. This paper examines how noise level differs depending on ones location to a motorcycle by mapping the sound pressure level of various styles and designs of motorcycles three dimensionally. Specifically, decibel levels and frequency spectra are measured volumetrically around the motorcycle to create a three dimensional map for each motorcycle. These three dimension sound maps are then compared and evaluated for patterns of direction bias, maximum sound level, and frequency.

11:05

5aNS8. Squealing noise in light rail transport systems: Implications in noise mapping. Lucie Habaskova (Faculty of Civil Enginering, Czech Tech. Univ. in Prague, Thkurova 7, Czech Republic, lucie.habaskova@fsv.cvut.cz) and Antonio P. Carvalho (Univ. of Porto, Porto 4200-465 Portugal)

Squealing noise is generated by railway vehicles riding through curves having closed radius. It is a result of wheels sliding on the rail when they negotiate a curve. This kind of noise is very annoying for people living in the surrounding areas where this phenomenon occurs because it has a characteristic spectrum that is dominated by discrete frequency components. Usually, squeal noise is not correctly accounted in urban areas noise maps. This paper presents the results of field measurements of squealing noise in Porto (Portugal) where a new light-rail system is operating. The goal was to characterize the acoustic effect of this type of noise in making noise maps. With the results obtained from *in situ* measurements, the squealing noise was simulated using CADNAA software. This paper compares three variants of squeal noise use and simulation with the software CADNAA. Comparing the noise map generated using the field measurements, with maps taking no account for squealing noise, a large difference in noise levels can be found (up to 17 dBA). A correction method was found to be incorporated in this model to improve the noise map estimation procedure.

11:20

5aNS9. Seattle's Sound Transit Residential Sound Insulation Program reduces sound sustainably. Julie A. Wiebusch (The Greenbusch Group, Inc., 1900 West Nickerson, Ste. 201, Seattle, WA 98119)

Light Rail has arrived in Seattle, making access to downtown easier and more efficient through the new 14-mile light rail system. However, the new system brings with it an increase in noise levels for some residents. In order to minimize the annoyance for those residents along the alignment, Sound Transit has created a Residential Sound Insulation Program (RSIP). The initial program provided 137 residences with a package to reduce the intrusion

of exterior sound and to increase the ventilation and air quality within the home. Sound Transit has taken the process one step further with its commitment that RSIP will have zero impact on local landfills.

11:35

5aNS10. A method for direct impedance measurements in long range propagation of low-frequency sound. Jin H. So, Carrick L. Talmadge, Roger Waxler, and Kenneth E. Gilbert (Natl. Ctr. for Physical Acoust., Univ. of MS, Univ. MS 38677, jso@olemiss.edu)

A method for direct impedance measurements in long range propagation of sound, significantly below 200 Hz, is presented. The linear least-squares fit to the measured pressure data, as a function of the sensor height, the so called "pressure derivative method," was utilized to obtain the pressure derivatives. Several sensors were installed vertically on the array above the ground. The ground was a flat agricultural farm. The propagation distance of sound over the ground was 0.41 km. Sound sources included pulses and pure tones. The impedances for the pulses and tones showed good agreement. Overall, the imaginary components of the impedance were much larger than the corresponding real components in the frequency range studied.

11:50

5aNS11. Analysis of the environmental effects on the ringtone volume. Cheol Hong Kim (Chonnam Natl. Univ., Dept. of Electron. and Com. Eng., 300 Yongbong-Dong, Buk-Gu Gwangju 500-757 KR, chkim22@chonnam.ac.kr), Hong Jun Choi (Chonnam Natl. Univ., Korea), and Jong-Myon Kim (Univ. of Ulsan, Korea)

Due to the wide spread of mobile phones, the use of them in public space keeps increasing. Thanks to their advantageous features, mobile phones enable people to feel convenience in making a call. However, the use of mobile phones in public space is blamed for causing others' inconvenience due to reckless noisy ringtones, especially in the quiet places such as classroom and conference room. In this work, we analyze the environmental (temperature, humidity, ambient noise, and vibration) effects on the ringtone volume to reduce the noise in public space. In fact, many users put their mobile phones in silent or vibrate mode in public space. However, we can still hear ringtones occasionally in public space. Therefore, if ringtones can be automatically adjusted to the appropriate volume level by considering the environmental effects, it can be very helpful to mobile phone users. To find the appropriate volume level, we analyze the various environmental effects by conducting actual experiments. According to our experimental results, vibration and ambient noise have strong relation with the ringtone volume, whereas temperature and humidity show little correspondence on the ringtone volume. [Work supported by the MKE, Korea, under the ITRC supervised by the IITA [IITA-2009-(C1090-0903-0008)].]

FRIDAY MORNING, 30 OCTOBER 2009 PECAN, 8:30 TO 10:15 A.M.

Session 5aSA

Structural Acoustics and Vibration and Biomedical Ultrasound/Bioresponse to Vibration: Structural Acoustics in Biological Media

Karl Grosh, Chair
Univ. of Michigan, Mechanical Engineering Dept., 2350 Hayward St., Ann Arbor, MI 48109-2125

Chair's Introduction—8:30

Invited Paper

8:35

5aSA1. Estimation of the contact pressure on the medial surface of the vocal folds during phonation. Li-Jen Chen and Luc Mongeau (Dept. of Mech. Eng., McGill Univ., McDonald Eng. Bldg., 817 Sherbrooke St. West, Montreal, Qvebec H3A 2K6, Canada, ljchen@purdue.edu)

Voice production involves flow-induced self-oscillations of the vocal folds. Collision between the vocal folds is commonly observed during normal phonation. The contact pressure experienced by human vocal folds during collision is usually considered as the most likely source of phonotrauma. The goal of the present study was to quantify contact pressures in human subjects during phonation. A pressure sensor was developed for direct measurements. Verification data and preliminary data on human subjects were obtained. Subject response and other clinical challenges lead to the development of a less intrusive approach for the contact pressure estimation from high speed images based on a Hertzian impact model. A verification of the accuracy of this approach was made. Results from the nonintrusive approach were compared with results from direct measurements using a hemilaryngeal physical model of the human vocal folds. The experimental setup was designed to reduce sensor's interference with the vocal fold oscillations. The accuracy of the estimated contact pressure from the nonintrusive method was found to be within around 10%. Advantages and possible sources of error are discussed. [Work supported by NIH.]

9:00

5aSA2. Influence of viscoelasticity on the ultrasonic wave propagation in cortical bone: Application to the axial transmission technique. Guillaume Haiat (Laboratoire de Recherches Orthopediques, CNRS, Univ. Paris 7, UMR CNRS 7052 B2OA, 10 avenue de Verdun, 75010 Paris, France), Salah Naili, Mai-Ba Vu (Univ. Paris 12-Val de Marne, 94010 Creteil cedex, France), Quentin Grimal, Maryline Talmant (Univ. Paris 6, 75006 Paris, France), Christophe Desceliers, and Christian Soize (Univ. Paris-Est, 77454 Marne la Valle, France)

Cortical bone quality is assessed in clinical practice using axial transmission (AT) devices. Cortical bone and the surrounding soft tissues are attenuating media, which might affect the results obtained with AT devices. Following the work of in Haiat *et al.* [J. Acoust. Soc. Am. (2009)] which considers the elastic case, the aim of this work is to evaluate the effect of anisotropic heterogeneous dissipative phenomena occurring in bone and in soft tissues on the ultrasonic response of the bone structure. A two-dimensional finite element time-domain method is derived to model transient wave propagation in a three-layer medium composed of an inhomogeneous transverse isotropic viscoelastic (Kelvin–Voigt description) solid layer sandwiched between two dissipative acoustic fluid layers. Bone viscoelasticity is assumed to be heterogeneous and a constant spatial gradient of viscoelastic properties is considered for two values of bone thicknesses corresponding to relatively thick and thin bones. The results allow the derivation of a contributing depth (CD) for a thick bone width. For a gradient of viscoelasticity, CD≅1.6 mm, for a gradient of C11, CD≅0.6 mm, and for a gradient of mass density, CD≅1 mm. [The Agence Nationale de la Recherche (Contract No. no BLAN06-2 144779) is acknowledged.]

9:15

5aSA3. Prediction of the basilar membrane response to low-intensity acoustic stimulation using a mathematical model of the cochlea. Julien Meaud and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48108-2125 jmeaud@umich.edu)

In the mammalian cochlea, the basilar membrane separates two fluid-filled ducts and vibrates in response to acoustic stimulation. *In vivo* measurements of a viable basilar membrane show that the membrane is very sensitive to low-level acoustic stimulation. A finite element model of the cochlea has been developed that explicitly couples the fluid, mechanical, and electrical domains of the cochlea. The proper function of the cochlea relies critically on structural acoustic coupling as well as electromechanical active control. The high sensitivity of the basilar membrane to low-intensity sounds is explained by the presence of somatic motility in outer hair cells. The outer hair cells convert electrical energy to mechanical energy, amplifying basilar membrane motion. Recent experiments show that the tips of the outer hair cells, the hair bundle, can also produce an active force at acoustic frequencies. Active force production by the hair bundle is included in the model to predict the effects of hair bundle motility on the basilar membrane vibrations. Our model allows us to answer important questions about the interplay of the various active forces in the cochlea; questions that are difficult to address by experiments alone.

9:30

5aSA4. How does sound leave the cochlea? Yizeng Li and Karl Grosh (Dept. of Mech. Eng., Univ. of Michigan, Ann Arbor, MI 48109)

In normal operation, sounds enter the cochlea via the stapes and generate forward traveling waves on the fluid-loaded basilar membrane [e.g., Bekesy (McGraw-Hill, 1960)]. Kemp [J. Acoust. Soc. Am. **64** discovered that sound is emitted by the cochlea in a way that is physiologically vulnerable. These sounds are called otoacoustic emissions (OAEs). Because of their relation to the health of the cochlea, OAEs have evolved as a tool for both noninvasive diagnosis and understanding cochlear mechanics [Probst, J. Acoust. Soc. Am. **89**]. However, OAEs are not yet completely understood. OAEs are hypothesized to arise from active processes in the cochlea. The path that intracochlearly generated sound takes when emitted from the cochlea—as a slow coupled fluid-structure wave or a fast mainly fluid borne wave (Wilson, Phys. Today (2008)]—has yet to be established. This work uses animations of mechanical-electrical-acoustic finite element model transient response predictions due to internal excitation of the basilar membrane to study the path of propagation. Models of sensitive and insensitive cochlea are used to determine the effect of sound level and cochlear health.

9:45

5aSA5. Rayleigh–Lamb wave propagation on a fractional order viscoelastic plate. Faik C. Meral, Thomas J. Royston, and Richard L. Magin (Univ. of Illinois at Chicago, Chicago, IL 60607, troyston@uic.edu)

Previous studies of the authors published in JASA focused on shear and surface wave motion in and on a viscoelastic material representative of biological tissue. Specific cases considered were that of surface wave motion on a half-space caused by a finite rigid circular disk located on the surface and oscillating normal to it and compression, and shear and surface wave motion in a half-space generated by a subsurface finite dipole. In both situations, recently it has been shown that viscoelasticity models based on fractional order derivatives result in closer agreement with experiment as compared to conventional models, such as those of Voigt and Zener. In the present study, this comparative analysis is extended to another configuration and wave-type, that of out-of-plane response of a circular viscoelastic plate to harmonic anti-symmetric excitation. The Rayleigh–Lamb equation is solved using complex wave numbers based on conventional and fractional order viscoelastic models for the medium. The new theoretical solutions are compared with experimental measurements. The underlying physical motivation for the use of a fractional viscoelastic assumption and the potential practical benefits, such as improved medical diagnostics, are discussed. [Work supported by NIH Grants EB004885, EB008373, and EB007537.]

10:00

5aSA6. Characterizing Rayleigh damping parameters of post-cured, hybrid, methacrylic materials used in stereolithography systems. Thomas H. Burns and Andrew R. Beeson (Starkey Labs, Inc., 6600 Washington Ave. S., Eden Prairie, MN 55344, tburns@starkey.com)

Stereolithographic systems of methacrylic are commonly used in the large-scale production of custom earmolds for hearing aids. In hearing aids, feedback occurs when energy from the receiver (earphone) is sensed by the microphone; the feedback manifests itself as an annoying whistle. One source of feedback is the mechanical vibration transmitted by the receiver to the microphone through the structural housing or earmold. Mechanical feedback can be reduced by appropriately designing the earmold housing to have proper isolation features and damping so as to minimize mechanical transmissibility. Toward this end, a procedure is described that utilizes both experimental measurements and finite element analyses to characterize an earmold material's stiffness and damping in the frequency range of 500 Hz–6 kHz. The experimental set-up consists of an electromagnetic shaker, laser vibrometer, dual-channel analyzer, and multiple cantilever (fixed-free) sample beams fabricated from post-cured, hybrid, methacrylic polymers. The dimensions of the samples are chosen so that the resonance frequencies are sufficiently distributed over the aforementioned frequency range. Bending resonances are measured and compared to a finite element model via a half-power bandwidth technique; parameters for Young's modulus, the loss factor, and Rayleigh damping are amended in the numerical model to match the empirical results.

Session 5aSC

Speech Communication: Cross-Language Speech Perception and Production (Poster Session)

Alexei Kochetov, Chair
Univ. of Toronto, Dept. of Linguistics, 130 St. George St., Toronto, ON M5S 3H1 Canada

Contributed Papers

All posters will be on display from 8:30 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:30 a.m. to 10:15 a.m. and contributors of event-numbered papers will be at their posters from 10:15 a.m. to 12:00 noon.

5aSC1. The effect of short-term training on production and perception of s/ʃ by Korean speakers. Robert Allen Fox, Ewa Jacewicz (Speech Percept. and Acoust. Labs., Speech and Hearing Sci., Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210, fox.2@osu.edu), Fred R. Eckman (Univ. of Wisconsin-Milwaukee, Milwaukee, WI), Gregory K. Iverson (Univ. of Maryland, College Park, MD), and Soyoung Lee (Univ. of Wisconsin-Milwaukee, Milwaukee, WI)

This study examines the efficacy of short-term training on learning the English s/ʃdistinction by native speakers of Korean. The production and perception of Korean L2 learners was tested before and after a short computer-controlled training session. Of interest in this study is to determine which acoustic characteristics of these fricatives (if any) improve after a focused training session, whether English listeners perceive them more "English-like," and whether the L2 learners can more accurately identify these fricatives after training. Korean speakers were initially recorded producing common English words containing these two fricatives in initial, medial, and final positions. Each speaker then underwent focused training on the s/ʃdistinction in a small set of nonsense words. Training ended when speakers met an established performance criterion. In a post-test, speakers were again recorded producing the same words. Acoustic measurements (including duration, relative amplitude, spectral mean, variance, skewness, and kurtosis) were obtained for each fricative from both sets of recordings and from productions by native English speakers. Perception data (including both identification and discrimination) were also obtained from the English-speaking controls and from L2 speakers both before and after the training. Results will be discussed in light of universal principles of language acquisition.

5aSC2. Consequences of short-term language exposure in infancy on babbling. Nancy Ward (Dept. of Linguist., UCLA, 3125 Campbell Hall, Box 951543, Los Angeles, CA 90095-1543, nancyward@ucla.edu), Megha Sundara (UCLA, Los Angeles, CA 90095 and Univ. of Washington, Seattle, WA 98195), Barbara Conboy (Univ. of Washington, Seattle, WA 98195 and Univ. of Redlands, Redlands, CA 92374), and Patricia Kuhl (Univ. of Washington, Seattle, WA 98195)

Kuhl *et al.* [Proc. Natl. Acad. Sci. U.S.A. **100**, 9096–9101 (2003)] showed that the decline in the discrimination of non-native perceptual contrasts observed during development can be reversed with short-term exposure to a non-native language. In this poster, the question of whether short-term exposure also impacts the speech produced by infants is addressed. For this purpose, 9–10-month-old infants from monolingual English-speaking households were exposed to Spanish for a total of 5 h over 6 weeks (12 25-min sessions). At the end of this exposure, babbling data were collected from the infants in two sessions with (a) an English-speaking parent and (b) a Spanish-speaking research assistant. In this experiment, adult listeners were tested in their ability to identify the babbling produced by these infants as English or Spanish. Canonical syllables and multi-syllabic utterances from these sessions were played to adult native speakers of English and Spanish, who identified the tokens as English or Spanish using a forced-choice paradigm. Results will be discussed in the context of the literature on short-term exposure as well as the effects of social interactions on language acquisition.

5aSC3. The correlation between perceiving and producing English obstruents across Korean learners. Kenneth de Jong and Yen-chen Hao (Dept. of Linguist., Indiana Univ., 322 Memorial Hall, 1021 E. 3rd St., Bloomington, IN 47405)

Our previous work determined that the skill structure evident in cross-subject correlations differed in second language production and perception. This study examines how these production and perception skills correlate across subjects, finding pervasive differences between contrasts similar to ones in the L1 and novel contrasts. Twenty Korean learners of English listened to and identified eight anterior obstruents produced by four American English speakers before, after, and between the vowel /a/. The same learners were also presented with orthographic probes and produced the same eight obstruents, and English listeners identified the consonants in these recordings. Analyses show no significant correlations between production and perception accuracy in consonant voicing (similar to L1), indicating pervasive disjunction between perceiving and producing voicing contrasts in the learners. Voicing accuracy was systematically better in identification than production, suggesting a precedence for perceptual learning. Analyses of manner accuracy in labial consonants, however, reveal a robust correlation, indicating a strong relationship between perceiving manner contrasts (novel to L1) and producing them. In addition, several subjects exhibited better accuracy in production than perception, suggesting an interaction of perceptual and motor learning. Coronal stops and fricatives show similar patterns, however, without the robust correlation between perceptual and production accuracy.

5aSC4. Production and perception of English /l/ and /r/ by native-speaking children. Kaori Idemaru (Dept. of East Asian Lang. and Lit., Univ. of Oregon, Eugene, OR 97403-1248, idemaru@uoregon.edu) and Lori L. Holt (Carnegie Mellon Univ., Pittsburgh, PA 15213)

The English /l-r/ distinction is a difficult contrast to learn for L2 learners as well as for native speaking children. In this study, we examine the use of the second (F2) and third (F3) formants and the relative weighting of these cues in production and perception of /l/ and /r/ sounds in native-English-speaking children. 3-, 4-, 5-, and 8-yr-old children produced words including /l-r/ and they were also tested for identification of words including the contrast. The results indicated that whereas young children's productions of /l/s and /r/s were well distinguished acoustically, the children were still developing in how they integrate F3 and F2 in both production and perception. Specifically, children's production indicated an intriguing developmental change such that F2 frequency increased in older children's /r/ productions, thus diminishing the F2-F3 frequency difference across development. In perception, although a sharper pattern of perception of the contrast was found in older children, they were not using F2 (a secondary cue signaling the categories) in the same manner as adults even at age 8. These data are con-

sistent with a rather long trajectory of phonetic development whereby native categories are refined and tuned well into childhood [Work supported by NSF Grant No. BCS-0746067].

5aSC5. A longitudinal study of English and Japanese vowel production by Japanese adults and children in an English immersion setting. Grace Oh, Susan G. Guion (Dept. of Linguist., Univ. of Oregon, Eugene, OR 97478-1290, gracey1980@yahoo.com), James E. Flege (Via della Torretta 12, 01017 Tuscania (VT), Italy), Katsura Aoyama (Texas Tech Univ. Health Sci. Ctr., Lubbock, TX 79430-6073), Reiko Akahane-Yamada (ATR Human Information Processing Res. Labs., Japan), and Tsuneo Yamada (Natl. Inst. of Multimedia Education, Japan)

The effect of age of acquisition on first- and second-language vowel production was investigated. Eight English vowels were produced three times in two different words each by 16 native Japanese (NJ) adults and children as well as 16 age-matched native English (NE) speaking adults. Productions were recorded shortly after the NJ participants' arrival in the United States and then 1 year later. In agreement with previous investigations [Aoyama *et al.*, J. Phonetics **32**, 233–250 (2004)], children were able to learn faster, leading to higher accuracy than adults in a year's time. Based on acoustic measurements, NJ adults had more accurate production at Time 1 but made no changes across time. The NJ children, on the other hand, showed significant differences from NE children's productions for /ɪ/, /ɛ/, /ɑ/, /ʌ/, and /ʊ/ at Time 1, but produced all eight vowels in a native-like manner at Time 2. A follow up examination of NJ children's productions of Japanese /i/, /a/, /u/ revealed significant changes for Japanese /i/ and /a/ related to L2 learning. The results suggest that L2 vowel production is heavily affected by age of acquisition and that there is a dynamic interaction between first- and second-language vowels. [Work supported by NIH.]

5aSC6. Acoustic analysis of the English pronunciation of Japanese high school teachers and university students. Ian L. Wilson, Junichi Fujinuma, Naoya Horiguchi, and Kazuaki Yamauchi (Ctr. for Lang. Res., Univ. of Aizu, Aizuwakamatsu 965-8580, Japan, wilson@u-aizu.ac.jp)

As Japan attempts to meet the demands of the Ministry of Education by introducing English-as-a-foreign-language (EFL) classes in all elementary schools, there has been a shortage of qualified native Japanese EFL teachers. The English communicative ability (and pronunciation, in particular) of Japanese EFL teachers varies across educational levels throughout Japan and even within individual prefectures. This study presents an acoustic analysis of 77 Fukushima Prefecture junior and senior high school Japanese EFL teachers' read speech, comparing the two levels. It also presents an analysis of 133 Japanese university students' read speech, recorded both before and after 14 weeks of weekly 90-min explicit pronunciation instruction. The reading passage was the "Please call Stella" paragraph from the speech accent archive (http://accent.gmu.edu/). Detailed analyses of vowel formants, voice onset time, fricative spectral peaks, and intonation were carried out using PRAAT and will be presented. The data have implications for curriculum planning for both EFL pronunciation classes and teacher training courses in Japan.

5aSC7. Native language phonetic drift in beginning second language acquisition. Charles B. Chang (Dept. of Linguist., Univ. of California, Berkeley, 1203 Dwinelle Hall, Berkeley, CA 94720, cbchang@berkeley.edu)

Mounting evidence has shown that native language (L1) phonetic representations can be affected by extensive second language (L2) learning. In this study the nature and time course of this phonetic drift in L1 are examined in a longitudinal investigation of 20 L1 English speakers' first 5 weeks of learning Korean as L2. Acoustic analyses of these learners' L1 and L2 speech production over time show that learning L2 stops affects the production of L1 stops (in terms of VOT and/or f_0 onset) in as little as 1 week, with the L1 sounds approximating the L2 sounds to which they are most phonetically similar. English voiced stops do not change significantly in VOT over time, since they are already similar to Korean fortis stops in this respect; however, they rise in f_0 onset in approximation to the elevated f_0 onset typical of Korean fortis stops. Meanwhile, English voiceless stops become longer in VOT and higher in f_0 onset in approximation to the Korean aspirated stops. These results indicate that L1 phonological categories are affected by L2 learning on a very short timescale, suggesting that the L1-L2 equivalence classification that gives rise to this phonetic drift may be rather low-level in nature.

5aSC8. Do they read aloud in a different way? A glimpse on the prosodic features of corpus-based second language spoken English. Sally Chen, Janice Fon, and Hintat Cheung (Grad. Inst. of Linguist., Natl. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan)

This study investigated the prosodic features of L2 spoken English. A set of ten recordings was extracted from an in-progress learner corpus on an English proficiency test. Each recording consists of two passages as read aloud by a learner who had received a grade of 3, the median grade of the test, on a five-point scale. A group of ten native English speakers was recruited to serve as controls. They were given the same test materials and their readings recorded under a test scenario similar to that of the L2 learners. The labeling followed the English ToBI convention. Preliminary results showed that in general, the L2 learners addressed more tones in their production. In terms of break indices, the L2 utterances were consistently segmented into a larger number of intonational units, as compared to those of the native speakers. In addition, silence was frequently employed in the L2 word-level boundaries to facilitate enunciation, explained in that English has a more complex phonotactic structure than their native tongue, Mandarin. Finally, the L2 data exhibited more varied pitch accent types than the native-speaker data, which corresponds to the findings of our previous study regarding an L2 read-aloud task on simple English sentences.

5aSC9. Perception of English vowels by Chinese learners in Taiwan. Li-chin Yang and Yu-da Lai (Dept. of English, Natl. Taiwan Normal Univ. and Dept. of Appl. English Studies, MingDao Univ., 369 Wen-Hua Rd., Peetow, ChangHua, Taiwan 52345, Republic of China)

The aim of this study attempts to explore how the structure of the native language and the markedness affect the perception of English vowels by Chinese learners in Taiwan. This paper concerns how people of different English proficiency levels perceive English vowels in terms of the similarity degrees of the vowels in both native and target languages (3 groups $\times$ 3 vowel similarity degrees). Thirty Mandarin and Southern Min speaking subjects, equally distributed in three English proficiency levels, namely, low, mid, and high, participated in the study. The quantitative analysis regarded the similarity degrees of the vowels in both native and target languages as three: identical, similar with counterpart, and new. This study also investigated how similar or new vowels are perceived by learners of different English proficiency levels (3 groups $\times$ 3 vowel markedness degrees). The similar or new vowels, according to P. Smolensky (1993) and Lombardi (1997), were divided into three categories of markedness: the most marked (rounded and back vowels), marked (front vowels), and unmarked (low vowels). Results showed the influence of both the markedness of English vowels and the different structures between English and native languages. Finally, they were explained via optimality theory.

5aSC10. Cross language effects on vowel perception in noise: Detection of English vowels spoken by English, Chinese, and Korean talkers. Su-Hyun Jin, Chang Liu, and Sangeeta Kamdar (Dept. of Commun. Sci. and Disord., Univ. of Texas Austin, 1 Univ. Station, A 1100, Austin, TX 78712, shjin@mail.utexas.edu)

The purpose of this study was to measure detection thresholds for 12 isolated American English vowels spoken by English (E), Chinese (C), and Korean (K) talkers for young normal-hearing listeners in the presence of a long-term speech-shaped noise presented at 70-dB SPL. Two groups of talkers were selected from each Chinese and Korean talkers based on the vowel intelligibility found from the previous study [Jin *et al.*, (2009)]: high and low intelligibility groups. Each group consisted of male and female talkers. Two English talkers (a male and female) were also included as a control. Preliminary analysis indicated that there was no statistically significant difference in the detection thresholds for vowels produced by native and nonnative talkers. However, thresholds for vowels differed by 20 dB across the 120 vowels (12 vowels $\times$ 10 talkers), which is consistent with Liu and Eddins (2008). It suggested that regardless of the language background of talkers, vowel detectability might be more influenced by vowel category and

individual talker. The effect of acoustic features (both temporal and spectral cues) in each vowel on native English listeners' detection thresholds will be discussed.

5aSC11. Comparison of perceptual vowel space for native vowels for monolingual and bilingual speakers of Malayalam. Sreedivya Radhakrishnan and John W. Hawks (School of Speech Pathol. & Audiol., Kent State Univ., Kent, OH 44242, sradhakr@kent.edu)

This study examines the perceptual vowel space for Malayalam for monolingual and bilingual Malayalam speakers. Malayalam is the native language of the state of Kerala, India, which uses five long and short vowels. Hence, in addition to spectral differences, speakers of Malayalam may use duration for the identification of vowels. This is different from English, which uses predominantly spectral differences alone as the major cue for vowel identification, even though there are intrinsic differences in duration for the English tense/lax vowels. There were two groups of subjects, native speakers of Malayalam and speakers of Malayalam who have also been speaking English for at least 10 years, with ten subjects in each group. The stimuli consisted of synthetic tokens of isolated vowels, which varied in first, second, and third formant frequencies. An identification task, followed by rating task, was carried out. This study maps the perceptual vowel space of native Malayalam speakers and also addresses a less widely studied area of cross-language speech perception, i.e., the influence of second language on the perception of the native language. Thus, the results provide a basis for the examination of the influence of long-term use of English on the perception of native Malayalam vowels.

5aSC12. Cross-language perception of Russian plain/palatalized laterals and rhotics. Alexei Kochetov and James Smith (Dept. of Linguist., Univ. of Toronto, 130 St. George St., Toronto M5S 3H1, Canada, al.kochetov @utoronto.ca)

A number of studies investigating factors in non-native speech perception have focused on discrimination and identification of English /l/ and /r/ by listeners whose native languages do not have the relevant phonemic contrast. Relatively little work, however, has been done on non-native perception of lateral/rhotic contrasts in other languages and particularly on the perception of palatalized laterals or rhotics. This paper presents results of an AX discrimination experiment where 84 listeners, native speakers of Cantonese, English, Japanese, Korean, Mandarin, and Russian, were presented with stimuli containing intervocalic consonants /l/, /l^{j}/, /r/, and /r^{j}/ produced by a Russian native speaker. The results revealed significant differences in the perception of the lateral/rhotic contrasts across the listener groups, with relatively good discrimination of the contrasts by Korean and English listeners (yet less accurate compared to Russian listeners), and much poorer discrimination by the other groups of non-native listeners. All the groups, including native speakers, performed better with the more acoustically distinct plain /l/ and /r/ contrast than with the palatalized /l^{j}/ and /r^{j}/ contrast. The results suggest that presence or absence of similar native phonemic categories does not fully predict listeners' performance, underscoring the importance of sub-phonemic gestural/acoustic detail in non-native perception.

5aSC13. The effects of linguistic experience on the perception of [h]. Marc Garellek (Dept. of Linguist., UCLA, 3125 Campbell Hall, Los Angeles, CA 90095-1543, marcgarellek@ucla.edu)

Languages with onset-[h] are thought to be more likely to develop coda-[h] than languages with no [h] at all [D. Silverman, *Phasing and Recoverability* (Garland, New York, 1997)]. It is therefore expected that speakers of languages with onset-[h] should be more likely to perceive coda-[h] than speakers of a language with no [h]. The goal of this study is to test this hypothesis. Speakers of French (a language with no [h]) and English (which allows onset-[h]) were instructed to monitor for [h] in strings of nonce words with [h] in onset or coda positions. If English speakers perform significantly better than French speakers on coda-[h] perception, then the hypothesis is supported. Subjects also participated in a mispronunciation detection task, where they monitored for errors in words of their language. The stimuli were common English and French monosyllabic words ending in [s, f, ʃ], with the fricatives spliced and replaced with [h], keeping the formant transitions intact. The stimuli were randomly placed in a string of native words with no contextual cues. If the hypothesis is supported, findings should show faster and more accurate mispronunciation detection for English speakers than French speakers, suggesting that English speakers are more likely to perceive coda-[h].

5aSC14. Perception of non-native voicing contrasts by Buenos Aires Spanish listeners. Marcos Rohena-Madrazo (Dept. of Linguist., New York Univ., 726 Broadway, 7th Fl., New York, NY 10003, mrm359@nyu.edu)

Studies on cross-language speech perception have found that adults have difficulty discriminating segments that are non-contrastive in their native language; however, not all non-native contrasts are equally difficult to perceive. This study investigates whether Buenos Aires Spanish (BAS) listeners can perceive non-native voicing contrasts. Spanish has a phonemic distinction between voiceless stops /ptk/ and voiced stops /bdg/; however, it lacks a voicing contrast in fricatives, exhibiting only voiceless ones: /fsx/. BAS also has a palato-alveolar fricative, which is variably realized [ʃ-ʒ] depending on social factors, but this difference is not phonemic. To determine whether BAS listeners, possessing a native stop voicing contrast, can perceive the non-native fricative voicing contrast, they performed a categorial AX discrimination task, with Portuguese voicing pairs (p/b, t/d, k/g, f/v, s/z,; ʃ/ʒ). The results show that, unsurprisingly, listeners perceive the voicing distinction of stops at ceiling level. However, they are at chance distinguishing the fricative voicing contrast, with one exception: f/v. Listeners are equally accurate distinguishing the f/v contrast as they are with the stop contrasts and equally inaccurate distinguishing s/z,ʃ/ʒ. This suggests that the non-native f/v contrast has a different perceptual assimilation pattern than do the other non-native fricative voicing contrasts, s/z and ʃ/ʒ.

5aSC15. Consonant intelligibility of English-, Chinese-, and Korean-native talkers. Chang Liu, Su-Hyun Jin, and Sangeeta Kamdar (Dept. of Commun. Sci. and Disord., Univ. of Texas at Austin, Austin, TX 78712)

Sixteen American-English consonants were recorded in a phonetic context of /aCa/ from young English-, Chinese-, and Korean-native talkers with normal speech and hearing functions. The Chinese and Korean talkers were bilingual with their US residency up to 6 years. The intelligibility of these consonants was examined for seven American-English listeners. Results showed that average consonant intelligibility for all three groups of listeners was well above 90% with English-native speakers having slightly higher intelligibility than the other two groups. Combined with our previous study on vowel intelligibility, there results indicate that the lower-than-native speech intelligibility for Chinese- and Korean-native speakers may be mainly due to the reduced vowel intelligibility, implying that speech training for these non-native speakers may need to focus on vowel production.

5aSC16. Perception of prominence by Japanese and American listeners. Irina A. Shport and Susan G. Guion (Dept. of Linguist., Univ. of Oregon, 1290, Eugene, OR 97403, ishport@uoregon.edu)

Acoustic cues to prosodic prominence are different in Japanese (mainly F0) and English (duration, F0, and intensity), but F0 change is involved in coding prominence in both languages. Japanese native speakers perceive a syllable as prominent (accented) if there is a sharp F0 fall originating in the syllable [Vance, (1987); Pierrehumbert & Beckman, (1988); Kubozono, (1993)]. It is not clear, however, whether native speakers of English interpret the F0 fall as a cue to prominence. In this study, participants are asked to identify a prominent syllable in each of the three-syllable words varying in (a) the location of the F0 maximum (the midpoint of the first or second syllable), (b) the height of the F0 maximum (10-Hz step increments from 200 to 250 Hz), and (c) the degree of the F0 fall between the accented and the following syllables (two levels). All stimuli are nonce words phonologically plausible in both Japanese and English. In our analysis we explore the role of each of the three manipulated parameters for perception of prominence by listeners who speak Tokyo Japanese and American English natively.

5aSC17. Manipulating within-speaker variation in non-native speech perception. Marie de Marneffe and Meghan Sumner (Dept. of Linguist., Stanford Univ., Stanford, CA 94305-2150, mcdm@stanford.edu)

Research in speech perception and L2 learning has shown that exposure to multiple speakers is generally beneficial [B. Sommers (2005); B. Bent (2008); Winters *et al.* (2005)]. Exactly what is driving these effects remains

unclear. We examine whether varying a particular within-speaker cue improves perception. Specifically, we study perception by native English speakers of French-accented speech and concentrate on final obstruent devoicing: Are listeners better at perceiving final stops when a constant non-English V/C ratio is used by the speaker or does hearing variable ratios help? We used a click-on paradigm, with 20 pairs of pictures representing words featuring voiced and voiceless final stops (e.g., rope-robe) in two conditions: (1) constant V/C ratios (average) and (2) variable V/C ratios (short, average, long). We used modified natural tokens using the average variance from 11 native French speakers. Our results show that across items, accuracy rates for both conditions are the same, but variation yields slower RTs. Considering accuracy alone these are not different from each other, but listeners are either more cautious or processing is slowed with increased variation. Additionally, differences in accuracy surface by place of articulation, but only for the alveolar place of articulation.

5aSC18. Language and dialect discrimination by five-month-olds. Megha Sundara and Chad Vicenik (Dept. of Linguist., UCLA, Los Angeles, CA 90095)

To acquire a language, infants must be able to accurately discriminate speech utterances belonging to that language. This is particularly important for infants raised in bilingual environments, who must be able to tag utterances they hear as belonging to one language or another. From birth, infants are able to discriminate languages from different rhythm classes, but not languages within the same rhythm class [Nazzi *et al.* (1998)]. By 5 months, infants can discriminate two languages from the same rhythm class, as long as one language is native. Johnson [(2000)] showed that English-learning infants could discriminate English from Dutch, and even American English from British English, but not Dutch from German. We attempt to extend these results to American English and German, and American and Australian English. Preliminary results show that English-learning 5-month-olds cannot discriminate American English from German but are able to discriminate American from Australian English. These results fit with previous work with adults that show discriminating between American and Australian English is easier than discriminating between American English and German, when segmental information is removed.

5aSC19. A comparison of phonological errors made by native and non-native listeners. Lindsay E. Meyer (Dept. of Commun. Sci. and Disord., Louisiana State Univ., 38 Hatcher Hall, Baton Rouge, LA 70808, lindsaymeyer571@gmail.com) and Janet L. McDonald (Louisiana State Univ., Baton Rouge, LA 70808)

Previous studies have shown that listeners have difficulty recognizing non-native words in noise [Mayo *et al.* (1997)] and affective prosody in quiet [Dromey *et al.* (2005)]. Difficulty increases with less L2 ability. Findings from the current study support these results while indicating that non-native listeners have additional difficulty identifying prosodic patterns (i.e., those indicating sentence type) and using semantic information under high levels of background noise. Participants heard sentences presented in −2, −4, −6, and −8-dB S/N. Target words were semantically predictable or unpredictable. Participant identification of sentence type and target words was recorded and analyzed. Both groups were affected by increasing levels of noise, but non-natives more strongly. Semantics aided native speakers under high levels of noise, but non-native speakers under low levels of noise. Analysis of phonological errors shows that the most frequent phonological error was on the onset and vowel, but this was not related to age of exposure. Later age of exposure for non-native listeners was correlated to fewer total phonologically correct words and they were more likely to give phonologically related forms that matched on the onset and coda. Therefore, age of exposure may affect the vowel whether in perception or production.

5aSC20. Effects of speechreading and signal-to-noise ratio on understanding mainstream American English by American and Indian adults. Yori Kanekama and David Downs (Dept. of Commun. Sci. and Disord., Wichita State Univ., 1845 Fairmount St., Wichita, KS 67260, yxkanekama@wichita.edu)

The purpose of this study was to measure effects of speechreading and signal-to-noise ratio (SNR) on understanding mainstream American English (MAE) heard by 30 Indian adults compared to 30 American adults. Participants listened to a recording of a female speaker of MAE saying ten lists of ten different everyday speech sentences per list. Participants heard sentences from a TV loudspeaker at a 65 dB sound pressure level, while a four-talker babble played through two surrounding loudspeakers at various SNRs under auditory-visual and auditory modalities. Each participant's speechreading performance at each SNR was computed as the difference in words correctly heard through auditory-visual versus auditory modalities. Consistent with most previous research, American participants benefited significantly more from speechreading at poorer SNRs than at favorable SNRs. The novel finding of this study, however, was that Indian participants benefited less from speechreading than American participants at poorer SNRs, but benefited more from speechreading than American participants at favorable SNRs. Linguistic (and, possibly, nonlinguistic) variables may have accounted for these findings, including an increased need for Indian participants to integrate more auditory cues with visual cues to benefit from speechreading, presumably because they only spoke English as a second language.

Session 5aUW

Underwater Acoustics and Acoustical Oceanography: Progress in Modeling Complex and Dynamic Acoustic and Oceanographic Characteristics of Continental Shelves and Slopes III

David P. Knobles, Chair
Univ. of Texas at Austin, Applied Research Labs., 10000 Burnet Rd., Austin, TX 78758

Contributed Papers

9:00

5aUW1. Marginal probability distributions for seabed parameter values from simultaneous processing of vertical and horizontal apertures. D. P. Knobles, R. A. Koch, and J. D. Sagers (Appl. Res. Labs., The Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Information about the physical parameter values for the seabed in a littoral region can be extracted from acoustic measurements in the water column. While acoustic data from both horizontal line arrays (HLAs) and vertical line arrays (VLAs) have been processed successfully to infer seabed properties, the simultaneous signal processing of acoustic data recorded on both a vertical and a horizontal aperture for seabed properties is a new research area. Relative to only a bottom-mounted horizontal aperture, an L-array (bottom mounted HLA and a VLA) provides access to the sensitivity of the depth-dependence of the acoustic field to seabed parameters. However, the utility of the additional information can be negated by uncertainty in the assumed water sound speed profile. Acoustic data collected on two L-arrays deployed on the New Jersey continental shelf are examined for the additional information about the seabed that can be obtained relative to the case for a single aperture. The information from simultaneous and individual processing of the subapertures is quantified in terms of the marginal distributions and an entropy functional approach. [Work supported by Office of Naval Research Code 321 OA.]

9:15

5aUW2. Seabed reflection measurements from an autonomous undersea vehicle: Probing seabed spatial variability. Charles W. Holland (Appl. Res. Lab., The Penn State Univ., State College, PA, cwh10@psu.edu) and Peter L. Nielsen (NATO Undersea Res. Ctr., La Spezia, Italy)

Spatial variability of the seabed in littoral regions is driven by processes that span a large range of space and time scales, ranging from eustatic fluctuations which are more or less global in extent but can have a profound effect on sedimentary layering from hundreds of meters to sub-meter scales, to biologic reworking which might be very local and affect only the upper few tens of centimeters. These processes (as well as many others) conspire to create a rich variety of sediment property distributions both in vertical and horizontal dimensions. One way to probe the spatial variability is with wide-angle reflection measurements. Here we discuss results from the Clutter09 experiment in which seabed reflection measurements were conducted using a broadband source (800–3500 Hz) and a 32 element towed horizontal receiver. The results are compared with previous observations using a single hydrophone and a Uniboomer. [Work supported by the Office of Naval Research OA321 and the NATO Undersea Research Centre.]

9:30

5aUW3. Low-frequency propagation in extremely shallow water. R. Daniel Costley (Miltec Res. and Technol., a Ducommun Co., 9 Industrial Park, Oxford, MS 38655) and Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Oxford, MS 38655)

In water depths on the order of a wavelength, sound propagates with considerable involvement of the bottom, whose bulk and shear moduli as well as velocities and attenuation vary with depth into the sediment. This problem has been studied with a two dimensional finite-element code to illustrate the consequences of computational assumptions pertinent to these parameters and their gradients with depth into the sediment. Both time harmonic (steady state) and transient (animated pulse) sequences are treated in two dimensional depictions. [Work supported by the U.S. Army Space and Missile Defense Command.]

9:45

5aUW4. On the use of an autonomous underwater vehicle for seabed characterization. P. L. Nielsen (NURC, V. S. Bartolomeo 400, 19126, La Spezia, Italy), C. W. Holland (Penn-State Univ., State College, PA 16804), and R. Hollett (NURC, 19126, La Spezia, Italy)

In 2009 NURC conducted the CLUTTER'09 experiment on the Malta Plateau, south of Sicily, in collaboration with the CLUTTER JRP partners from USA and Canada. One of the main objectives of this experiment was to characterize the seabed for geoacoustic and scattering properties related to clutter, i.e., around regions on the bottom which generate target-like returns on active sonar displays. The equipment used was a newly developed sound source and 32-m horizontal line array at NURC towed behind the Ocean Explorer Autonomous Underwater Vehicle (OEX-AUV). The source transmitted signals in the frequency band 800–3500 Hz and were received on the line array. The array has a four-level aperture which allows both for utilizing individual hydrophone data and beamformed data for the seabed characterization. The advantage of using the OEX-AUV is that measurements can be performed close to the seabed which are difficult to obtain from conventional sonar systems towed from a surface vessel. Results of environmental characterization using the OEX-AUV from a selected region are presented and compared to independent findings from previous experiments. [Work supported by the NATO Undersea Research Centre and the Office of Naval Research OA321]

10:00—10:15 Break

10:15

5aUW5. A mode solution for sediment-refracted wave propagation. Cathy Ann Clark (Naval Undersea Warfare Ctr., 1176 Howell St., Newport, RI 02841, cathy.clark@navy.mil)

An normal mode solution for low-frequency bottom-sediment refracted wave propagation is formulated for use with a normal mode method of solving the Helmholtz wave equation to describe the underwater sound field for a fixed point source in a plane multilayered medium. The approach utilizes a factorization of the Wronskian of the Green's function which enables sediment reflecting and refracting modes to be summed independently. The sensitivity of propagation loss to the nature of mode interaction with the bottom is investigated for a sample environment and results are compared to measured data.

10:30

5aUW6. Spatial variability of density gradients in the transition layer from Bayesian inference of seabed reflection data. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC V8W 3P6, Canada), and Charles W. Holland (Penn State Univ., State College, PA 16804)

This paper considers Bayesian inference of seabed reflection-coefficient data for geoacoustic models of the transition layer (i.e., uppermost low-velocity, water-saturated sediments) at several sites along a track, with the goal of studying spatial variability. Geoacoustic models are parametrized in terms of nonlinear density and linear sound-velocity gradients. Rigorous uncertainty estimation is of key importance to resolve spatial variability between measurement sites from the inherent inversion uncertainties. Geoacoustic uncertainty estimation is carried out including comprehensive estimation of data error statistics. Model parametrization is addressed by choosing gradient parameters that allow for a large variety of profile shapes. Metropolis–Hastings sampling is used to compute posterior probability densities. Several experimental sites are considered along a track located on the Malta Plateau, Mediteranean Sea, where the transition layer is expected to change with increasing distance from shore due to changes in the sedimentation processes. Differences between sites that exceed the estimated geoacoustic uncertainties are interpreted as spatial variability of the seabed. Density and sound-velocity gradients are clearly resolved by the reflectivity data and agree well with core measurements within the credibility bounds.

10:45

5aUW7. Estimating geoacoustic properties of marine sediment on the New Jersey Continental Shelf from broadband signals. Yong-Min Jiang and N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, BC V8W 3P6, Canada, minj@uvic.ca)

This paper presents geoacoustic inversions of broadband signals collected in the Shallow Water 2006 Experiments off the coast of New Jersey. An L-shaped array was deployed on the top of a sand ridge, in 70 m of water. The acoustic source was maintained at a distance of 190 m from the vertical leg of the L-shape array, and lowered from 10 to 60 m in 10-m intervals in the experiment. Two sets of chirps, low frequency (100–900 Hz) and midfrequency (1100–2900 Hz), were transmitted with approximately the same experimental geometry. The water column sound speed profiles were measured at the source position. This study examines the sediment information content from different frequency bands recorded on the vertical leg of the L-shape array. Because the chirps have different temporal resolutions and energy penetrabilities in the sediment, the received signals exhibit different bottom structures at different frequency bands. Travel time geoacoustic inversions are carried out at both frequency bands using signals from the resolvable reflections from the sediment layers. The variation in the oceanic sound speed profile is parametrized in terms of its characteristics and included in the inversion. [Work supported by ONR Ocean Acoustics.]